THIS WEEK

EDITORIALS

GARAGE FLOWERS Why the boom in DIY biology matters to scientists **p.634**

POSTDOCS Please release me, let me go to start my own lab **p.635**

BROUGHT TO BOOK Marine census reveals its results **p.638**

Stem-cell laws in China fall short

The Chinese government's regulations of stem-cell treatments are admirable in principle, but tougher enforcement measures are needed to protect patients.

China does not want to be known as the Wild West of unproven medical technologies. Last year, the government took an important step when it announced regulations requiring, among other things, that anyone who offers stem-cell procedures should present clinical data supporting their efficacy, and secure approval from the health ministry (see *Nature* **459,** 146; 2009).

Such regulations are sorely needed. A leading bioethicist in China last year estimated that more than 100 laboratories there offer stem-cell procedures, many of them unproven, although some clinics reportedly stopped offering the treatments after the regulations took effect. But the government needs to do more than simply announce rules; it needs to give companies clear instructions for complying with them.

The regulations have made little difference so far to Beike Biotechnology in Shenzhen, China's — and perhaps the world's — most prolific purveyor of stem-cell treatments. Beike develops therapies for disorders ranging from multiple sclerosis to lupus, based on adult and umbilical-cord stem cells. Its treatments, offered by more than 30 hospitals throughout China, have been injected into about 9,300 patients, who pay as much as US$26,000 for the procedure. Roughly half have muscular dystrophy or spinal-cord injuries, but many experts say that stem-cell treatments for those conditions are not ready for clinical use.

Beike has not tested the efficacy of its treatments in formal clinical trials, says Alex Moffett, chief executive of Bangkok-based Beike Holdings and a spokesman for Beike Biotechnology, although he does say that some phase I safety trials are taking place. The company offers numerous testimonials on its website as evidence that its treatments work. But some media accounts report scepticism, and the mother of one Beike patient complained directly to *Nature* that her son's condition did not improve at all.

Moffett says that he believes in the need for evidence based on clinical trials. As the company plans an expansion into Malaysia, the Philippines and Thailand, he says, it will complete clinical trials of every treatment's safety and efficacy, at no cost to patients, before offering the procedures commercially. He says that Beike "probably should have" taken the same approach in China.

Yet the company has passed muster with the government, says Moffett. Officials have visited Beike's facilities without closing them down, which he interprets as tacit approval for the treatments. He says he knows of no application steps for formal approval from the health ministry — and requested that *Nature* forward him any information about such procedures.

The problem, it seems, is that the regulations do not include enough details for implementation and enforcement. The health ministry is now considering proposed guidelines, created by a group of scientists and ethicists, that set out clear criteria for preclinical and clinical studies, and clinical applications. One of the committee members said that the guidelines call for an approval process that is easily accessible, includes an ethical review and is based on solid scientific data.

The committee member says that the government will decide "soon" whether and how to implement the guidelines. (A Chinese academic who has been following the issue says that implementation has been held up by disagreements between different government agencies.)

Soon cannot be soon enough. The guidelines need to be approved and put in place as fast as possible, and enforced swiftly and effectively. The longer that unproven therapies stay on the market, the greater is the risk that a history of use can be framed as evidence of safety and efficacy.

Since an investigation of China's food and drug agency a few years ago found it to be rife with corruption (see *Nature* **446,** 598–599; 2007 and *Nature Med.* **13,** 889; 2007), the government has worked hard to put together a system that balances drug-company profit, the requirements of drug innovation, and patients' health and welfare. Those efforts offer a model for regulation of stem-cell therapies. Anything less than a clear, detailed set of rules is a disservice to scientists who are working hard to understand stem cells and their clinical promise, to companies that are taking big financial risks and doing proper clinical trials, and most of all to patients. ■

Transgenic harvest

African nations are laying foundations to extend the use of GM technology on the continent.

The use of genetically modified (GM) crops for food divides opinion, especially when it comes to Africa. Sharp views on the technology in the developed world, honed by more than a decade of arguments in Europe and elsewhere, are too easily projected onto Africa, with the continent portrayed as a passive participant in the global melodrama over GM food. So it is heartening to see a group of 19 African nations working to develop policies that should make it clear to all sides in the debate that Africa can make up its own mind.

After more than nine years, talks between member states of the Common Market for Eastern and Southern Africa (COMESA) have produced a draft policy on GM technology, which was sent for national consultation last month. COMESA is a trade bloc, and its proposals aim to develop research and trade in GM crops. But they also state that decisions should be based on sound science and evidence.

Under the proposals, a nation that wants to grow a GM crop commercially would inform COMESA, which would then carry out a science-based risk assessment — COMESA seems to have sufficient access to scientific expertise to fulfil this role. The body would judge

whether the crop is safe for the environment and human consumption. If the assessment proves positive, blanket approval would probably be given for the crop to be grown commercially in all COMESA countries. National governments would retain the power to decide whether or not to proceed.

Risk assessments are currently left to individual countries, but this requires scientific expertise, money and a well-established regulatory system. That combination is rare in Africa, and only four countries — Malawi, Zambia, Zimbabwe and Kenya — have passed laws specifically to govern GM organisms. This helps to explain why there are so few GM crops grown commercially across Africa. Even field trials of GM crops are scarce, although tests of a banana engineered to resist bacterial disease will begin in Uganda this week (see *Nature* doi:10.1038/news.2010.509; 2010).

Under the COMESA plan, the African nations are consulting on a biosafety road map to guide the development of national regulations on transgenic organisms, and on regimes and mechanisms for monitoring and inspection. A communication strategy to provide countries with the latest scientific information on GM organisms is also under discussion.

The consultation is expected to continue until March, with a decision coming from the relevant ministers soon after. If agreed, the proposals will help many more African nations to explore agricultural biotechnology should they wish to, and perhaps to profit from the increased food security that the technology has the potential to provide. By working together, nations will also benefit from greater access to the experience of commercial issues relating to GM technology that is currently the preserve of just a few African countries.

For their efforts so far, these nations should be applauded, as should the African scientists who have managed to get their voices heard in a difficult and contentious debate. The moves signal a shift towards evidence-based assessments of technologies that could hold much promise for the continent.

African countries have been wise to draw from the speed and enthusiasm with which nations such as Brazil have exploited GM technology, rather than the confused and fearful stance of European countries such as France. The few GM crop initiatives across Africa are already dispelling some myths peddled by the anti-GM lobby, such as the image of poor African farmers being exploited by profiteering multinational companies. In fact, many of the existing projects involving GM organisms in Africa are public–private partnerships through which companies donate their best technologies royalty-free.

It is by no means certain that the COMESA proposals will get through the consultation unscathed. A key sticking point is concern in some countries that regional guidelines would usurp national sovereignty. And although Zambia is the only country in the bloc to take an explicit anti-GM stance, others are pushing for tougher rules that could restrict the adoption of the technology.

African countries should not let ideological opposition to GM technology cloud the admirably clear view that they have taken on the issue so far. Food and water shortages that already ravage the continent will only get worse, and GM technology offers a promising way to tackle poverty and poor agricultural productivity. The question is not whether countries there should adopt GM crops, but how quickly. ■

Garage biology

Amateur scientists who experiment at home should be welcomed by the professionals.

For the past two years, a group of molecular-biology enthusiasts has met regularly in Cambridge, Massachusetts, to discuss science. Their conservation is not entirely theoretical: they swap stories about the experiments they perform in rudimentary labs built in their kitchens, basements and garages. These meetings are not unique: similar gatherings are cropping up across the United States and Europe, as amateur scientists get together to compare protocols and results from experiments they design and conduct at home.

Do-it-yourself biologists emerged into the spotlight after the first meeting, in a Cambridge pub, in 2008. Their exploits have since earned them a moniker fit for the headlines of the twenty-first century: biohackers (see page 650). Media coverage has taken its toll on the public's perception of 'DIYbio'. Stories in the press are often peppered with sweeping claims of the monumental advances to be made by unleashing the talents of the public at large on important biological questions. Equally common are breathless warnings that a bioterrorist is busy crafting the next plague in a garage, safe from the watchful eye of the authorities.

Neither image rings true. Most biohackers are hobbyists who delight in crafting their own equipment and who tackle projects no more sophisticated than those found in an advanced high-school biology lab. This is not to belittle their achievements — the most basic lab experiments can be a challenge without the institutional infrastructure professional scientists take for granted. And it is not necessarily the sophistication of the techniques, but the questions to which they are applied, that makes for compelling science. Nevertheless, the high financial and educational barriers to cutting-edge molecular biology means that garage labs are unlikely to solve the world's energy or health problems any time soon. As for that imagined bioterrorist, US experts at the FBI's Weapons of Mass Destruction Directorate have investigated and found no sign of a biohacker who intends harm.

Nevertheless, the bureau is wise to plan ahead. The FBI has embarked on a laudable and proactive programme to establish ties with the amateur biology community. FBI agents attend DIYbio meetings and invite DIYbio leaders to conferences on bioterrorism. This has yielded some practical plans, such as notifying police and fire stations about local garage labs, to avoid unpleasant surprises or false alarms in the event of an emergency. But some in the biohacking community worry that the constant focus on bioterrorism has taken attention and resources away from a more pressing issue: basic biosafety. How should a biohacker dispose of unwanted genetically engineered bacteria? How does an amateur biologist avoid exposure to fumes from the chemicals used to isolate and manipulate DNA? What is a safe bacterium for a hobbyist to play with?

These are questions that crop up daily in a garage lab, and amateur biologists have struggled to find answers. Although institutions such as the US National Institutes of Health (NIH) and the Centers for Disease Control and Prevention have established biosafety guidelines, these are aimed at institutional biosafety officers with training in the field. Laden with jargon and focused on advanced work with dangerous chemicals and pathogens that hobbyists are unlikely to encounter, the guidelines are little help in the garage. Does this knowledge gap provide an opportunity for professional scientists to engage and support the DIYbio community? Some researchers argue it does, with professionals helping garage biologists craft safety guidelines and standards that could be understood by the enthusiast. Biohackers could also be brought onto biosafety committees at their local university or medical centre. These committees are required by the NIH to include at least one member who is not a professional scientist. Serving on a such committee would expose the hobbyist to the regulations and protocols that research institutions use to protect workers and the environment.

Biohackers are an example of the growing 'citizen science' movement, in which the public takes an active role in scientific experiments. Citizen science can help stimulate public support for science, and can introduce fresh ideas from novel disciplines. Science is a professional business but it would be a shame if the only interested knock on the hobbyists' doors came from those in law enforcement. ■

K. WOLF/NATURE

Scientists need a shorter path to research freedom

Francis Collins *explains why the NIH is launching a bid to help some doctoral students dramatically reduce the time required to start an independent career.*

Over the past half-century, a great many things have changed in biomedical research. Along the way, postdoctoral training has become an established step in a research career. But this development has proved a double-edged sword for some — and possibly for the whole field.

Without question, postdoctoral training has enriched the experience of many by allowing protected time for full immersion in research. Postdocs provide essential skills and serve as first authors on many important papers, thus boosting research productivity. But these gains must be set against the significantly longer time it now takes for most young scientists to launch independent research careers. The average age of PhD scientists awarded their first research grant from the US National Institutes of Health (NIH) last year was 42. In 1981, the average was 36. As director of the NIH, I believe this is a problem that should be addressed. We must develop ways to liberate our brightest minds to pursue high-risk, high-reward ideas during their most creative years.

There are many complex reasons for the increased training periods, including an academic culture that emphasizes the need for longer, sometimes multiple, postdoc positions to build a stellar CV. There is a shortage of faculty vacancies, and institutions often insist that recruits win independent funding before appointing them to tenure-track posts. And there is too little emphasis on alternative scientific careers, such as industry, law, teaching and policy.

Many young researchers baulk at the prospect of such an extended period of limited intellectual autonomy. It is also a concern to veterans such as myself. I fear that science may be suffering because of a failure to encourage the independence of the next generation of great minds.

My own pathway to independence involved a three-year postdoctoral fellowship in human genetics in the lab of Sherman Weissman at Yale School of Medicine in New Haven, Connecticut. I was fortunate to be mentored by an adviser who encouraged autonomy and creativity. I used the opportunity to develop an innovative approach, called chromosome jumping, for crossing large strands of DNA to identify genes responsible for inherited disorders. It was a good launching pad; I received my first R01 grant from the NIH at age 34, the same year I began a faculty position at the University of Michigan in Ann Arbor.

In my lab at the NIH, I strive to cultivate the independence of young scientists as early as possible. One of my strategies is to assign new recruits a 'thinking period' devoted to formulating project ideas. Through an iterative process involving myself and the recruit, we refine the research direction until we have settled on a good fit. I think this strategy has worked well in encouraging forward thinking, but it may still be a halfway solution. For the most creative of young scientists, nothing can equal the chance to have a lab of one's own.

WE MUST LIBERATE OUR BRIGHTEST MINDS TO PURSUE HIGH-RISK, HIGH-REWARD IDEAS.

To provide such opportunities, several programmes aimed at promoting greater independence at earlier career stages have sprung up over the years, producing some spectacular investigators. And so, after much consultation with outside advisers, the NIH this week launched its own effort, the Early Independence Award Program (see go.nature.com/nFqYE5), which will initially support ten creative young scientists to pass almost immediately from completing a PhD to running their own laboratories. The awards will be paid by the NIH Director's Common Fund and administered through a peer-reviewed application process, supporting an investigator at a level of US$250,000 in direct costs per year for five years — the equivalent of a standard NIH R01 grant.

Unlike many similar programmes, the awards will give students flexibility to seek a position at any suitable institution. Applicants will need to work with the institution's academic leaders to negotiate an independent position that would be activated if they win an award. We hope that department heads will find this an attractive tool for recruiting talent to invigorate their institution's research environment. For its part, the institution must provide the young investigator with space and resources, and a level of mentoring equivalent to that provided to assistant professors.

I am aware that many speed bumps may lie on this expressway to independence. The programme requires highly motivated and mature applicants who are talented and confident enough to launch their own research programme and negotiate support from a department chair. And it requires institutions willing to support an award winner who will be unusually young in their career. The pilot programme, which we expect to be highly competitive, will issue its first awards next year. Although not intended to replace traditional postdoctoral training, the pilot can be scaled up if successful.

This programme is not for everyone, and postdoctoral positions will continue to expand the skills and experience of most young scientists. But for exceptional individuals with the intellectual and experimental sophistication to initiate an independent career at the end of doctoral training, this programme will provide the opportunity. I have been involved in the launch of many pilots, including that of the Human Genome Project, but I have a special affinity for this one: the future of biomedical research relies on the creativity and energy of its investigators. Unleashing that capability at all stages of a scientist's career should be a priority for us all. ■

➲ NATURE.COM
Discuss this article online at
go.nature.com/QGk2LR

Francis Collins *is director of the US National Institutes of Health.*
e-mail: francis.collins@nih.gov

RESEARCH HIGHLIGHTS

Selections from the scientific literature

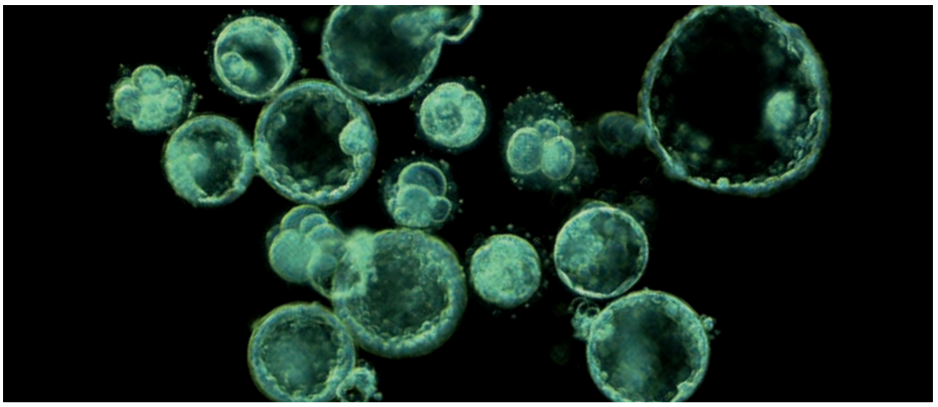
COGNITIVE NEUROSCIENCE

Vicious cycle of overeating

Obese people are known to have a less sensitive reward centre in the brain, which drives them to overeat. This may, in turn, further dampen their reward circuitry for food.

Eric Stice at the University of Texas at Austin and his co-workers used functional magnetic resonance imaging to scan the brains of 20 overweight female volunteers of similar body mass index (BMI) as they sipped either a chocolate milkshake or a tasteless solution. The researchers repeated the tests six months later. They found that women who had a greater than 2.5% increase in their BMI over the interim period showed a reduced response in the brain's striatum to the milkshake relative to their baseline response, as well as to women whose weight had remained steady over the six-month period.

J. Neurosci. 30, **13105–13109 (2010)**



K. LOEWKE/AUXOGYN

BIOTECHNOLOGY

Pictures predict embryos' fate

The likelihood that a human embryo cultured during *in vitro* fertilization (IVF) will develop successfully to the five-day mark can be predicted with about 93% sensitivity and specificity from three early developmental events.

In IVF, 50–70% of embryos never make it to the blastocyst stage, which begins five or six days after fertilization. Renee Reijo Pera at Stanford University in California and her group analysed images of 242 IVF embryos (pictured) taken with microscopic time-lapse photography. They found that those that would go on to form blastocysts showed specific developmental patterns, such as the first cytokinesis — cleavage that results in two separate cells — lasting less than 33 minutes. The team devised an algorithm to automatically screen embryos for this and two other parameters, and found that it could predict which embryos would reach the blastocyst stage.

Nature Biotechnol. **doi:10.1038/nbt.1686 (2010)**

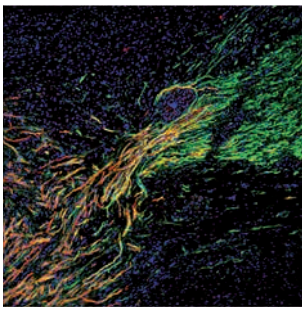
NEUROSCIENCE

Guide the way to nerve repair

If severed, nerves outside the brain and spinal cord can reconnect and resume functioning. Unexpectedly, the molecular mechanism behind this remarkable ability turns out to involve fibroblasts — a type of cell that helps with wound repair.

Alison Lloyd at University College London and her colleagues used fluorescence microscopy to observe how cut nerves mend themselves in rats. They observed fibroblasts at the site of injury make contact with Schwann cells, which surround and protect neurons. Signalling between the two cell types prompted the Schwann cells to clump into tiny cords that guide the regrowth of neurons across the wound (pictured). This response, the authors found, is mediated by a protein called SOX2, which is also involved in reprogramming cells to a stem-cell-like state.

Cell 143, **145-155 (2010)**



ELSEVIER

HYDROLOGY

Groundwater stores running dry

Groundwater supplies are shrinking at an increasing rate — according to new estimates, annual depletion more than doubled from 126 cubic kilometres in 1960 to 283 cubic kilometres in 2000.

About one-third of the world's inhabitants have limited access to fresh water, and many must draw from underground aquifers — often more rapidly than natural processes can refill them. Marc Bierkens at Utrecht University in the Netherlands and his colleagues combined a groundwater database with a global hydrological model. They found depletion rates to be highest in some of the world's major agricultural regions, including northwest India, northeast China, and the central United States.

The authors also found that a significant amount of the extracted groundwater evaporates and precipitates over the ocean, accounting for about 25% of the annual rise in sea level.

Geophys. Res. Lett. **doi:10.1029/2010GL044571 (2010)**

CANCER BIOLOGY

Tumours pave their own path

For tumour cells to infiltrate new tissues, they must first clear an escape route from their

primary site by remodelling their environment. Michael Olson of the Beatson Institute for Cancer Research in Glasgow, UK, and his colleagues have identified two proteins, LIMK1 and LIMK2, that are active in cells on a tumour's leading edge and seem to pave the way.

The LIMK proteins regulate actin, a key protein in the cellular skeleton. The researchers found that if these proteins were inhibited in cultured breast and skin cancer cells, the cells were still motile on a two-dimensional surface, but became less invasive in three-dimensional assays. Two processes normally associated with remodelling — degradation of nearby proteins and deformation of the matrix surrounding the cancer cells — were also impaired, suggesting that the cells failed to reshape their environment and forge a way out.

J. Cell Biol. **doi:10.1083/jcb.201002041 (2010)**

INFECTIOUS DISEASE

Battling bacterial blood infection

The tissue damage seen in sepsis — which is triggered most often by a microbial infection and can cause organ failure and death — is caused by a component of red blood cells. Help is at hand, however, as the blood also contains a protein that combats these effects.

Bacterial infection causes red blood cells to rupture, releasing the oxygen-transporting molecule haemoglobin. As this oxidizes, it releases free haeme, which can trigger programmed cell death. Miguel Soares at the Gulbenkian Institute of Science in Oeiras, Portugal, and his team found that mice lacking a protein that breaks down haeme had higher levels of haeme in their blood and an increased susceptibility to sepsis. In addition, administering extra haeme to normal mice pushed low-grade infections to become septic. But giving the animals a protein called haemopexin neutralized haeme's toxic effects.

Sci. Transl. Med. 2, **51ra71 (2010)**

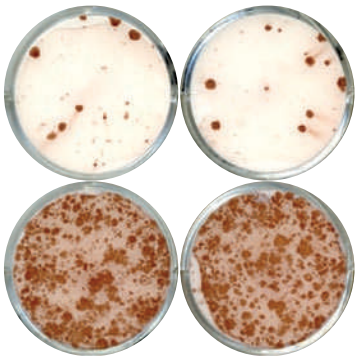
STEM CELLS

Reprogramming cells with RNA

Careful reprogramming yields specialized cells able to develop into any tissue type. Such cells, known as induced pluripotent stem (iPS) cells, have been made by using viruses to insert four key genes into their genome, but this carries the risk of turning the cells cancerous. A new method not only does away with genes, it also seems to be more efficient.

Derrick Rossi at Harvard Medical School in Boston, Massachusetts, and his colleagues chemically modified RNAs transcribed from the four genes — *KLF4*, c-*MYC*, *OCT4* and *SOX2* — and introduced these into human fibroblast cells. This method (pictured bottom, in Petri dishes) proved more efficient at generating iPS cells than the virus method (pictured top). Furthermore, treating the iPS cells with an additional RNA transcript turned them into muscle cells.

Cell Stem Cell **doi:10.1016/j.stem.2010.08.012 (2010)**



ELSEVIER

ECOLOGY

Not-so-extinct animals

Of all the mammalian species thought to have become extinct since the year 1500, about one-third have at some stage been rediscovered.

Diana Fisher and Simon Blomberg at the University of Queensland in Brisbane, Australia, examined the scientific literature and compared past and present Red Lists of threatened species compiled by the International Union for Conservation of Nature. The authors' analysis revealed that animal species that have suffered habitat loss are more likely to be rediscovered than those that have become extinct owing to over-hunting, or introduced predators or disease.

Furthermore, iconic species such as the Tasmanian tiger (*Thylacinus cynocephalus*) are searched for more frequently — and with less success — than less-iconic animals, such as the Australian lesser stick-nest rat (*Leporillus apicalis*), which may still exist.

Proc. R. Soc. B **doi:10.1098/rspb.2010.1579 (2010)**

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

CELL BIOLOGY

Thriving with genomic errors

Organisms with the wrong number of chromosomes often die or have growth abnormalities, yet most cancers have the same error and thrive. To figure out how cancer cells overcome this growth disadvantage, Angelika Amon at the Massachusetts Institute of Technology in Cambridge and her colleagues analysed the genomes of 14 yeast strains with extra chromosomes and higher growth rates. The researchers teased out a mutation in a gene coding for an enzyme known as Ubp6. This enzyme normally works to remove a molecule called ubiquitin from proteins, preventing the proteins from being degraded in the cell.

The researchers found that some strains with the Ubp6 mutation proliferated faster than similar ones without the mutation. These strains also had protein compositions that were closer to those of normal yeast cells than were those of strains without the mutation, suggesting that the Ubp6 mutants degrade the excess proteins generated by their extra chromosomes.

Cell 143, **71–83 (2010)**

COMMUNITY CHOICE

The most viewed papers in science

ENVIRONMENTAL SCIENCE

Where greenhouse gases start

HIGHLY READ on pubs.acs.org in the last month

Household consumption is responsible for 72% of global greenhouse-gas emissions, according to work by Edgar Hertwich and Glen Peters at the Norwegian University of Science and Technology in Trondheim. Another 10% is a result of government consumption, with the remainder due to activities such as building construction.

The authors analysed emissions using a model of goods and services consumption and trade across 73 nations. The analysis suggests that food, including agricultural production, is the largest component of consumption, accounting for 20% of all emissions worldwide. Residential energy consumption and building maintenance comes in second at 19%, with private household transportation contributing 17%.

Environ. Sci. Technol. 43, **6414–6420 (2009)**

NEW ONLINE

Papers published this week at nature.com PAGE 663

SEVEN DAYS *The news in brief*

POLICY

Stem-cell stay

US federal funding for human embryonic stem-cell (hESC) research can continue while a lawsuit seeking to block it works its way through the courts, a District of Columbia appeals court ruled on 28 September. The ruling puts a long-term stay on a preliminary injunction, which froze new and unpaid grant monies, that a lower court issued on 23 August. It had agreed with plaintiffs arguing that hESC research violates federal law by destroying embryos. The government is appealing that decision. Oral arguments in the case are to be scheduled shortly after 4 November, the date by which appeal briefs must be filed. For updates, see www.nature.com/stemcellfunding.

NASA budget

NASA's future priorities were firmed up on 29 September when a three-year, US$58-billion outline budget for the agency was passed by Congress. The act instructs NASA to use resources from a cancelled Moon-rocket programme to develop a heavy-lift vehicle for missions to deep space, and provides $1.6 billion to commercial companies building rockets to reach the International Space Station. It also calls for an additional Space Shuttle flight before the fleet retires in 2011, at a cost of $500 million, which may divert resources from research. See go.nature.com/ngtqnb for more.

French budget

France's ministry of science and higher education was spared deep cuts in the country's deficit-reducing budget for 2011, announced last week. Research minister Valérie Pécresse said the ministry would get a €4.7-billion (US$6.4-billion) increase over last year, although €3.5 billion of this comes from last year's research stimulus package, much of which cannot be spent immediately. Cash directly for research would rise by €468 million, or 1.9% — inflation is currently 1.4%. French unions note that in real terms some core-research funding agencies — such as the marine research agency (IFREMER) and the national research centre (CNRS) will see budgets drop. See go.nature.com/sj17nq for more.

Ocean census

The completion of a 10-year, US$650-million project to record ocean biodiversity was announced at a symposium in London on 4 October. Begun in 2000, the Census of Marine Life employed more than 2,700 scientists to catalogue the diversity, distribution and abundance of life in the world's seas. It has logged nearly 30 million observations and has described more than 1,000 new species — such as the 'Yeti crab' *Kiwa hirsuta* (pictured, right), discovered in 2005. Despite such successes, the organization that got the project started — the Sloan Foundation in New York — is not funding a second census, and no other major sponsors have stepped up (see *Nature* **467,** 514–515; 2010).

IFREMER/A. FIFIS

RESEARCH

Earth-like planet

Astronomers have found the first potentially habitable planet outside our Solar System. Steve Vogt of the University of California, Santa Cruz, Paul Butler of the Carnegie Institution of Washington in Washington DC and their colleagues declared the finding on 29 September. A paper is in press at *The Astrophysical Journal*. The planet, the sixth to be detected orbiting Gliese 581, a cool red dwarf star 20.3 light years from Earth, is 3–5 times Earth's mass; its average surface temperature is 228 K, or higher if the planet has an atmosphere. Astrophysicist Sara Seager of the Massachusetts Institute of Technology in Cambridge says the discovery is both "incremental and monumental". See go.nature.com/n9eg4y for more.

China's Moon shot

China's second unmanned lunar probe, Chang'e 2, was due to be orbiting the Moon five days after its launch on 1 October. The probe is carrying a laser altimeter and a camera to survey prospective sites for a Moon lander, itself expected to launch in 2013.

Oil-spill research

Details of BP's ten-year US$500-million research programme to monitor environmental damage from the oil spill in the Gulf of Mexico were belatedly announced on 29 September. BP had pledged the money in May, but an edict from the White House in June stalled much of its distribution, by directing that the company coordinate its efforts with state authorities. The programme will now be administered by BP and a partnership of five Gulf Coast states; BP and the partnership will appoint an equal number of scientists to a peer-review research grants board.

Unethical study

The US government on 1 October issued a formal apology for a 1946–48 study in Guatemala on potential treatments for syphilis, in which as many as 696 prisoners, soldiers and mental-health patients were infected with or exposed to the disease. The study, conducted by US-funded scientists, was unearthed by Susan Reverby, a professor of women's studies at Wellesley College in

Massachusetts. She alerted the government to her findings before publishing them. See p.645 for more.

Plants under threat

More than 20% of the world's 380,000 plant species are at risk of extinction, making plants more threatened than birds, says the first global analysis of plant biodiversity. The study, called the Sampled Red List Index for Plants, was conducted by researchers at the Royal Botanic Gardens, Kew in the UK, and was published on 28 September. Seed-bearing plants, or gymnosperms, are most at risk; habitat loss is the biggest threat to survival. See nature.com/7gf5lu for more.

BUSINESS

Patent pool nudge

The US National Institutes of Health has given a small but symbolic boost to a fledgling effort to speed access to affordable HIV/AIDS medicines in developing countries. The agency said it would become the first to share a drug patent with the three-month-old Medicines Patent Pool (MPP), an initiative backed by the international drug-purchasing facility UNITAID. The health agency will license to the MPP its patent on the anti-retroviral drug darunavir. That won't immediately allow a cheaper generic version, as additional darunavir patents are held by drug maker Tibotec, a subsidiary of US firm Johnson & Johnson. Tibotec is in discussions with the MPP, a spokeswoman said.

Herpes vaccine fail

An experimental vaccine against herpes simplex virus (HSV) failed to prevent infection in a trial of more than 8,000 women, its manufacturer GlaxoSmithKline announced last week. The drug firm, based in London, plans to scupper the Simplirix vaccine. Despite a huge potential market, investors didn't have high expectations for Simplirix, says Hedwig Kresse, at Datamonitor in London. There were already questions over its effectiveness.

PEOPLE

Nobel physicist dies

Georges Charpak, who won the 1992 Nobel Prize in Physics for inventing the principles behind modern particle detectors, died on 29 September aged 86. From 1959, he worked at CERN, Europe's premier particle-physics laboratory near Geneva, Switzerland. His designs enabled researchers to track many particles electronically in real time. The

A. BRUCELLE/SYGMA/CORBIS

technique involved an array of parallel wires suspended in a gas. A particle passing through the gas would create ions, which were attracted to the wires, producing a current. That design "wasn't the most elegant", he said, "but it was useful".

NSF head approved

The US Senate on 30 September confirmed Subra Suresh as director of the National Science Foundation (NSF) for a six-year term. Nominated for the post by President Barack Obama's administration in June, Suresh was previously dean of engineering at the Massachusetts Institute of Technology in Cambridge (see *Nature* **465,** 673; 2010). He replaces nuclear engineer Arden Bement.

Schön keeps PhD

A court in Freiburg, Germany, on 27 September upheld the right of disgraced physicist Jan Hendrik Schön to keep his doctoral degree. The judgement reverses a 2004 decision by Schön's alma mater, the University of Konstanz in Germany, to withdraw his 1997 PhD because his later behaviour showed him "unworthy" to hold it. The university is considering an appeal. Schön, who was a staff physicist at Lucent Technologies' Bell Labs in New Jersey, is notorious for perpetrating a remarkable string of fabrications in high-profile papers published between 2000 and 2002 in the fields of organic and molecular electronics.

COMING UP

10–13 OCTOBER
Evidence on what drove the evolution of marine animals that once lived on land, such as penguins and sea turtles, is presented at the Society of Vertebrate Paleontology's 70th anniversary meeting, in Pittsburgh, Pennsylvania.
go.nature.com/tdkswu

11–13 OCTOBER
Cancer genome sequencing and the Human Microbiome Project are discussed at a Beyond the Genome meeting, at Harvard Medical School in Boston, Massachusetts.
go.nature.com/9avsal

AWARDS

Nobel winners

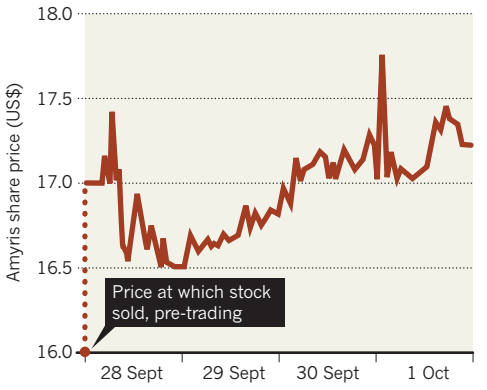
Robert Edwards won the 2010 Nobel Prize in Physiology or Medicine for his development of *in vitro* fertilization. The Nobel Prize in Physics went to Andre Geim and Konstantin Novoselov for their experiments on graphene. See pages 641–642 for more. *Nature* went to press before the chemistry prize was awarded, but full details will be available at go.nature.com/4bmndd.

BUSINESS WATCH

Amyris Biotechnologies, a synthetic-biology start-up co-founded by engineer Jay Keasling at the University of California, Berkeley, has had a modest stock-market debut. The firm, of Emeryville, California, engineers yeast to make hydrocarbon fuels and other chemicals from sugarcane. On 28 September, Amyris raised US$85 million, selling 5.3 million shares at $16 apiece; the price climbed 8% by the end of that week. It had hoped for $100 million. The launch can still be considered a success, as the biofuels market is weak and many firms are seeing low valuations, says Mark Bunger, research director of business consultancy Lux Research in San Francisco, California.

GREENTECH IPO

Investors cautiously welcomed the initial public offering of synthetic biology start-up Amyris.



NEWS IN FOCUS

CORBIN O'GRADY STUDIO/SPL

Robert Edwards (pictured in 1989) overcame a series of obstacles to make *in vitro* fertilization a reality.

AWARDS

Baby boom bags Nobel prize

UK pioneer of in vitro *fertilization wins medicine honour.*

BY ALISON ABBOTT

Very few scientists can say that four million people are alive because of their work, but Robert Edwards is one of those few. His development of the technique at the heart of that claim — *in vitro* fertilization (IVF) — has won him this year's Nobel Prize in Physiology or Medicine.

To make IVF possible, Edwards had to solve numerous problems in basic biology — some of which opened the door for embryonic stem-cell research — while facing bitter opposition from churches, politicians and even some of his eminent colleagues at the University of Cambridge, UK. An outgoing yet thoughtful personality who eagerly engaged in public debate, Edwards was hurt by charges that his work was unethical.

But thanks also to his collaboration with another outsider, Patrick Steptoe, an obstetrician at the Oldham and District General Hospital, the world's first test-tube baby, Louise Brown, was born in 1978. Within five years, 150 test-tube babies had been born worldwide. Since then, IVF has become mainstream, and Edwards and Steptoe have been lauded for helping give life to millions. Had he not died in 1988, Steptoe would probably have shared the prize.

In 2001, Edwards won a Lasker award, which often presages the Nobel. Two years ago he celebrated the 30th anniversary of IVF at a symposium where the impact of this work on many levels of society — biology and medicine, but also law, ethics, the arts and social anthropology — was discussed. At 85, Edwards is now too frail to give interviews, but his wife told the Nobel Foundation of his happiness at receiving the prize. "No other scientist could have transformed so many aspects of our society," says Martin Johnson, one of Edwards's first graduate students and now professor of reproductive sciences at the University of Cambridge.

Edwards began his research career in the early 1950s working on the reproductive biology of mice. After harvesting eggs from female mice, he learned how to coax them, and eggs from other species, to mature and be fertilized in a test tube. He also worked out how to control the timing of the rodents' ovulation — which annoyingly tended to happen at night — by administering certain hormones.

Soon after he joined the National Institute for Medical Research in London in 1958, Edwards began applying his findings, and those of other groups working on reproductive biology, to humans. He acquired slices of human ovaries from surgeons, and from these he isolated immature eggs. He spent two disappointing years failing to coax them to mature *in vitro*, until he realized that the process required at least 24 hours of incubation, not the 12 hours that rodent eggs required. "It is these empirical observations that move science forward," says Ian Wilmut of the MRC Centre for Regenerative Medicine at the University of Edinburgh, UK, who also had to modify the conventional timing of cell incubation to create the first cloned mammal, Dolly the sheep. "These things seem very small in retrospect, but they are critical."

By 1968, Edwards had fine-tuned the maturation of human eggs, learning how to fertilize them with the potential father's sperm and to prod them into forming embryos that could be implanted. Having moved to the University of Cambridge he needed a collaborator to help him apply these techniques in human ▶

patients. Having read about Steptoe's pioneering work on laparoscopy — the placement of a fibre-optic endoscope into the abdomen to view internal organs — in his small hospital in northern England, Edwards picked up the phone. Steptoe was already using the method to withdraw fluid from the reproductive tract and agreed that he could also use it to extract eggs. Working as equal partners, the pair set out their own ethical guidelines, agreeing to stop if patients or children were endangered, but not in deference to what Edwards called "vague religious or political reasons".

The UK Medical Research Council refused to fund their work, believing it could lead to babies with severe abnormalities, and disapproving of the pair's high profile in the media (M. H. Johnson *et al. Hum. Reprod.* **25,** 2157–2174; 2010).

Johnson recalls the "strange atmosphere" in the 1960s and 1970s, when the prospect of overpopulation seemed to be a bigger societal concern than infertility. "There was no awareness then of the personal pain of infertility," he says. "I remember eminent Cambridge scientists would tell us that our PhD supervisor was off his rocker." He also recalls Max Perutz and James Watson, both Nobel laureates at Cambridge, telling him it was irresponsible to interfere with the beginning of life. "Often people refused to speak to us in the tea room because they disapproved of what we were doing." Johnson stuck by Edwards though, finding him "inspirational and visionary".

The technique has not only benefitted infertile couples — it can also help parents to avoid passing on serious inherited diseases such as cystic fibrosis or Huntington's disease to their children, by selecting embryos that are free of dangerous mutations for implantation.

In addition, it has enabled the field of human embryonic stem-cell research. Reproductive biologist Outi Hovatta of the Karolinska Institute's IVF clinic in Stockholm, where new human embryonic stem-cell lines are derived from spare embryos, says that Edwards was the first, in 1984, to publicly discuss the benefits of such cells to medical research, and the ethical dilemmas that would inevitably accompany them. He was equally prescient on the need for oversight of his powerful technique, advocating in 1971 that a legal authority should be established to control IVF. The UK Human Fertilisation and Embryology Authority was founded 20 years later. ■

AWARDS

Graphene speeds pair to Stockholm win

Research on carbon sheets scores Nobel Prize in Physics.

BY GEOFF BRUMFIEL

Sheets of carbon with the potential to revolutionize electronics and materials science have bagged this year's Nobel Prize in Physics. Andre Geim and Konstantin Novoselov at the University of Manchester, UK, have been awarded the prize for their work on graphene, a one-atom-thick hexagonal mesh of carbon atoms that has become physicists' material *du jour*.

Geim and Novoselov reported the first free-standing graphene samples in 2004, having used little more than adhesive tape to create the material. Their team stuck the tape to a piece of graphite, peeled off flakes of carbon and then separated graphene from the rest of the flakes. Placing the graphene onto a silicon substrate, the researchers showed that it is a good electrical conductor[1].

Graphene's relatively recent rise to prominence makes it an unusual candidate for a Nobel, and marks the shortest lag-time between experiment and award since Johannes Georg Bednorz and Karl Alexander Müller won the physics prize for their discovery of high-temperature superconductivity in 1987, 18 months after their findings were published.

"I think very few doubted that there would be a Nobel prize," says Andrea Ferrari, an electrical engineer at the University of Cambridge, UK, who researches graphene applications. But, he adds, "I was surprised that it came so early." The award caught even Geim off guard. "When I got the telephone call, I thought, 'oh shit!'", he told reporters at a press conference shortly after the announcement. "The second thought that came to my mind was, 'Oh dear, I will not win many more prizes'."

Graphene's win may be down to the astonishing speed at which the field has developed. Almost immediately after its discovery, researchers realized that graphene was no ordinary material. Electrons travelling through the sheets display unusual quantum behaviours that can be easily studied[2,3]. Graphene's two-dimensional nature, and its atomic structure, also causes electrons to move through it much faster than they do through materials such as silicon.

Konstantin Novoselov (left) and Andre Geim: from sticky tape to Nobel prize in just six years.

UNIV. MANCHESTER (K. N.); J. KING-HOLMES/SPL (A. G.)

These properties make graphene a hot prospect for constructing computer chips. Although the sheets themselves do not behave as semiconductors, thin ribbons of graphene do. The ribbons' properties are not ideal for electronics, but advocates believe that graphene's speedy electrons and potential affordability could allow it to one day supplant silicon. A nearer-term use may be as a transparent, conducting layer in touch screens[4], or as flexible displays.

Graphene has also been teamed with DNA to create a chemical sensor, and can even act as a sponge to clean polluted water. Geim thinks that the material has the potential to be as revolutionary as plastics. "My hope is that graphene and other two-dimensional crystals will change our everyday lives," he says. ■

1. Novoselov, K. S. *et al. Science* **306,** 666–669 (2004).
2. Zhang, Y. *et al. Nature* **438,** 201–204 (2005).
3. Novoselov, K. S. *et al. Nature* **438,** 197–200 (2005).
4. Kim, K. S. *et al. Nature* **457,** 706–710 (2009).

MORE NOBELS ONLINE

PODCAST

● Join *Nature*'s podcast team for more discussion of this year's Nobel prizes
www.nature.com/nature/podcast
● Listen to a 2007 interview with the 'godfather of graphene', Andre Geim
go.nature.com/EWfJzA

LATEST NEWS

Catch *Nature*'s coverage of the Nobel Prize in Chemistry at go.nature.com/4BmnDD

FROM THE ARCHIVE

● Robert Edwards's Commentary on IVF and the history of stem cells
go.nature.com/XzoH76
● The future of reproductive medicine
go.nature.com/z8VL6p
● Graphene gets ready for the big time go.nature.com/spRvF5

NIDA

Some researchers contend that all addiction-related work should be overseen by one entity.

POLICY

Plan for addiction institute splits NIH

Researchers are divided after a vote to combine work on drugs and alcohol in one unified body.

BY MEREDITH WADMAN

Is alcohol a drug? Biochemically, the answer seems clear: the health effects of ethanol are well documented, as is its addictive nature. Yet culturally, politically and economically alcohol stands apart from other drugs, as does the research that has built up around it.

This conundrum has now acquired a practical urgency, thanks to a debate within the US National Institutes of Health (NIH). Should separate research institutes for drugs and alcohol be dissolved and replaced by a unified addictions institute?

"We've been talking about this for a dozen years. It's time to just do the right thing," said Harold Varmus, director of the National Cancer Institute and a proponent of the plan, at a meeting of the NIH's Scientific Management Review Board in Bethesda, Maryland, on 15 September. The board voted 12–3 to recommend dismantling the National Institute on Drug Abuse (NIDA) and the National Institute on Alcohol Abuse and Alcoholism (NIAAA) in favour of a new entity that would house the addiction work of both, along with that of other NIH institutes. Non-addiction programmes would be moved elsewhere within the NIH.

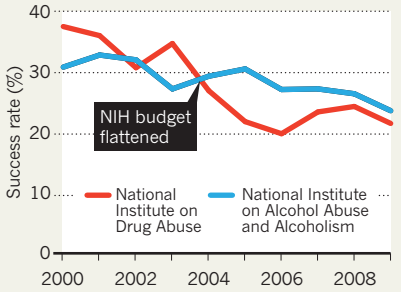
Rarely has an issue caused such division in the NIH: earlier this year, advisers to NIAAA voted unanimously against the plan, whereas advisers to NIDA voted unanimously for it.

"It makes infinite scientific sense," says Alan Leshner, chief executive of the American Association for the Advancement of Science in Washington DC, who directed NIDA from 1994 to 2001. "There's no question that these are all drugs of abuse and addiction and they should be treated together, dealt with together."

"It's a terrible idea," counters Enoch Gordis, director of the alcohol institute for 15 years, until 2001. "This is not a meeting of equals. This is one institute trying to take over another."

GRANT DATA DISTILLED

Since 2004, alcohol-institute grant applicants have enjoyed discernibly better success rates than drug-institute grant applicants.



SOURCE: NIH

The angst at NIAAA stems in part from a fear of being swallowed: with a 2010 budget of US$462 million, the alcohol institute is less than half the size of the $1-billion NIDA. Since 2004, external researchers funded by NIAAA have also enjoyed better grant-application success rates than have those seeking funding from the drug institute (see graph).

Alcohol-institute defenders argue that the loss of an independent NIAAA will discourage scientists from entering the field, obscure a social problem far more costly than drug abuse, rob the smaller institute of agility in responding to opportunities and imperil its systems approach to the study of alcohol toxicity. (Under the proposal, studies of end-organ damage such as liver cirrhosis, and of related conditions such as fetal alcohol syndrome, would move to other institutes.) Groups including the American Association for the Study of Liver Diseases in Alexandria, Virginia, and the National Organization on Fetal Alcohol Syndrome in Washington DC have vocally opposed the merger. So has Mothers Against Drunk Driving, a non-profit organization based in Irving, Texas.

Although the alcohol industry is unlikely to relish its legal product being lumped in for study with street drugs such as cocaine and heroin, it has so far remained silent. US Trade groups including the Beer Institute, the Wine Institute, the American Beverage Institute and the Distilled Spirits Council of the United States all declined to comment for this article.

Proponents of consolidation argue that because illegal-drug abusers and addicts often also abuse alcohol, their maladies are best understood and treated under the auspices of one institute. "The idea of a single strong institute that powerfully pushes forward the frontier in addiction research is certainly exciting," says Daniele Piomelli, a pharmacologist at the University of California, Irvine, who is funded by both institutes. Still, he adds, "the big risk is that in the end the overall funding will decrease, because we have one institute instead of two".

To move forward, the proposal must be approved by NIH director Francis Collins, be signed off by the Department of Health and Human Services and elicit no objections from Congress. Three weeks ago, Collins promised a decision "in the fairly near future" but didn't tip his hand. He is also considering an alternative proposal that would "functionally" integrate addiction research at the NIH through coordination and collaboration between institutes.

Smaller NIH institutes are eyeing the proceedings, wondering whether they will be next.

If there is an argument for consolidating institutes, "it is much more far-reaching than just merging two", says William Roper, a member of the NIH review board and dean of the University of North Carolina School of Medicine in Chapel Hill. Roper says that, if consolidation went ahead, he would choose to reduce the 20 current institutes to as few as 10, and "really have at it and make some significant changes". ■

CIVIL RIGHTS

NASA privacy case goes to highest court

The US Supreme Court will rule on sweeping background checks on scientists at the Jet Propulsion Laboratory.

BY EUGENIE SAMUEL REICH

For Robert Nelson, a planetary scientist who led the science team for NASA's Deep Space 1 flyby of Comet Borelly in 2001, the happiest day of the past three years came when a respected civil-rights law firm agreed to take on his lawsuit against the US space agency.

This week, *Nelson v. NASA* reaches the US Supreme Court, which will have to decide whether to uphold a lower court's preliminary injunction halting extensive investigations into the personal lives of employees at the Jet Propulsion Laboratory (JPL) in Pasadena, California. The case may have far-reaching implications for the privacy rights of scientists receiving US government funds.

"Yes, the investigators even want to ask about who we've slept with," Nelson told a press conference convened last week by the Union of Concerned Scientists (UCS), an advocacy group in Cambridge, Massachusetts, which has filed a brief with the court in support of Nelson and 27 co-plaintiffs.

The Jet Propulsion Laboratory in California.

NASA

The background investigations stem from Homeland Security Presidential Directive 12, introduced in 2004 during the administration of George W. Bush. Its purpose was to institute a standard identification badge to gain access to federal facilities, although the procedure needed to obtain the badge was left up to individual agencies. NASA required its contractors' employees to sign a waiver permitting investigators to collect "any adverse information" from anyone they liked, including information related to "abuse of alcohol and/or drugs", "mental or emotional stability", "general behavior or conduct", or "other matters".

Most of the JPL's 5,000 or so workers are employed by the Pasadena-based California Institute of Technology — which runs the JPL under contract with NASA — and are therefore subject to the wide-ranging inquiries required by NASA's reading of the presidential directive. In August 2007, under threat of losing their jobs, Nelson and the 27 other JPL employees sued rather than submit to the extra investigations.

On 5 October, the day of the JPL's deadline requiring them to sign the waiver or be dismissed, an appeals court issued a preliminary injunction stating that the investigations infringed a constitutional right to privacy because they were not narrowly tailored to national security interests. Two subsequent rulings by the same court set the injunction on course to become permanent, but in November 2009 the US Department of Justice appealed to the Supreme Court, which will hear the case on Tuesday and could issue an opinion any time after that. The current Supreme Court has heard few cases balancing privacy rights with national security, making it unclear which way it will rule.

The US Department of Justice did not respond to a request for comment. In its brief to the Supreme Court, it describes the investigations as "basic background checks" that are needed to ensure the safety and security of federal facilities. It also notes that the US Privacy Act should prevent the personal information that is collected from ever becoming public.

In a brief supporting the scientists, the American Astronomical Society (AAS) in Washington DC argues that the case could set a precedent that would allow similarly open-ended investigations of any scientist who needs access to federal facilities or who is applying for federal grant money. "We were concerned that this was impacting some people's ability to do research," says AAS president Debra Elmegreen, an astronomer at Vassar College in Poughkeepsie, New York. Kurt Gottfried, an emeritus professor of physics at Cornell University and a UCS board member, told the press conference that the Obama administration was taking essentially the same approach to national security as that of Bush. "This policy will set a harmful precedent. It will make it much more difficult to retain and attract top scientific staff," he said. ■

➲ NATURE.COM
Why are Saturn's rings so icy?
go.nature.com/qwpkxw

M. HALEVI

Q&A Susan Reverby
A shocking discovery

Susan Reverby is a historian and professor of women's and gender studies at Wellesley College near Boston in Massachusetts. She is an authority on the notorious Tuskegee experiments, during which treatment was withheld from more than 600 African American men with syphilis. Her recent discovery that the US Public Health Service exposed several hundred Guatemalans to the disease in an undocumented research project in 1946–48 led last week to an official apology from the United States to the Guatemalan government and the promise of a full investigation.

Why did the US government do this research?
By 1946 it was known that syphilis could be cured with penicillin. The primary aim of the study was to look at whether penicillin could also be used as a prophylaxis to get rid of the disease before it established itself in the body. They were looking for something that would be more successful than the 'pro-kits' soldiers had been given during the Second World War, which required them to apply a mercury-based ointment. As you can imagine, there was not a lot of compliance with that approach. They were also interested in whether or not someone could be re-infected with syphilis once cured by penicillin.

The arrangements were set up through Juan Funes, a Guatemalan doctor who had previously trained with the US Public Health Service (PHS) and was then the chief of the venereal disease control division of the Guatemalan Sanidad Publica. Funes and [PHS researcher] John Cutler, who later participated in the Tuskegee experiments, essentially did the study together.

How was the study conducted?
Prostitution was legal in Guatemala, as was bringing in a prostitute for sexual servicing of prisoners in the central penitentiary. They plied some of the prisoners with alcohol and sent in prostitutes whom they knew were infected with syphilis. When not enough infection appeared and they couldn't get enough cases, they made an inoculum. In later tests, they abraded people's bodies — their forearms, cheeks or penises — and applied the inoculum to a piece of cotton or gauze that was held to the abrasion for an hour and a half to two hours. But they had trouble transferring the infection this way, and eventually interest waned. By 1948, the studies were called off. [The US Centers for Disease Control and Prevention (CDC) subsequently found that 427 of 696 subjects were judged to be infected; 369 of those subsequently received "adequate treatment" with penicillin.]

Were these experiments approved?
You can't just walk into the Guatemalan central penitentiary and start doing something like this without someone in charge saying yes. Likewise for the National Mental Health Hospital, where later studies were done. At the mental hospital, Cutler and Funes were able to arrange access by giving the institution things like a projector, and metal plates and utensils. They also provided the anti-epileptic drug Dilantin [phenytoin], because a lot of people in the hospital were actually epileptics and had no medicine to help them control their seizures. That was the trade-off with the asylum keepers.

Was any of this work ever published?
No. In the early 1950s, Cutler was part of a syphilis inoculation project in Sing Sing Prison in New York. That study is published and they make it clear that they asked the permission of the prisoners. They don't reference the Guatemala work. I think they knew it was on an ethical edge. And there were internal questions in the PHS about what they were doing.

But at the time, syphilis was an enormously debilitating disease that health officials were very worried about. They thought they were at war with the disease and they were trying to find things that would help. In a war there are soldiers. I think they saw their subjects as soldiers in that war.

How did you come across all this?
Two years ago, while I was doing research at the University of Pittsburgh, I learned that Cutler, who had taught there, had left papers behind. I asked to see the papers in case there were any documents about Tuskegee I had not already seen. What I found was the Guatemalan material.

I was shocked. I could see the papers were talking about inoculations. I've been working on Tuskegee for two decades now and I've spent a lot of time explaining to people that no one was given syphilis in Tuskegee. So you can imagine what it was like to be sitting there reading this. It was just unbelievable.

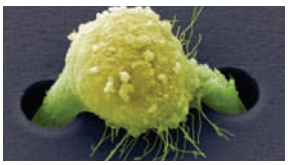
What happened next?
After completing my book *Examining Tuskegee* I went back to the University of Pittsburgh in June 2009 and re-did my research on the Guatemalan study to make sure I had it right. I wrote about it in an article that will be published [in the *Journal of Policy History*] this January. I gave a copy of it to David Sencer, who was the CDC director when Tuskegee broke in 1972 and who I've been in contact with since I interviewed him for the book. He felt this was important and asked if he could take it to people at the CDC. They, in turn, were shocked enough to send a syphilis expert back to Pittsburgh to look at the data. He confirmed what I had found and then it went up the chain of command.

What can today's researchers learn from this?
Most US drug trials are now conducted internationally. We have controls in the United States, but what's being done elsewhere in the world? If the Guatemalan study had been done by a private drug company, I never would have known about it. The lesson for today is the importance of institutional review boards, and of making sure that informed consent is really understood and applied in international trials. ■

INTERVIEW BY IVAN SEMENIUK

EPIGENETIC EFFECTS
A few disease studies in the NIH Roadmap Epigenomics Project.

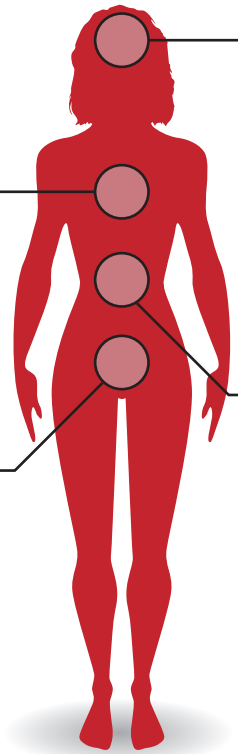
CANCER



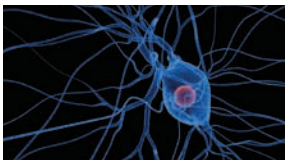
Control of gene expression by epigenetic modification could have a role in tumour formation, and could explain how environmental factors trigger cancer.

PRENATAL CHANGES

Molecular modifications to fetal and maternal DNA before birth could later make people susceptible to type 2 diabetes or cardiovascular disease.

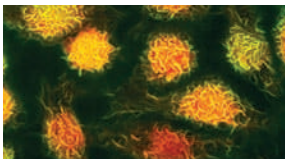


BRAIN DISORDERS



Epigenetic changes have been implicated in brain health, from cognitive decline in normal ageing to conditions such as Alzheimer's disease, schizophrenia, bipolar disorder and autism.

CHRONIC DISEASES



Complex chronic conditions such as systemic lupus erythematosus, asthma and insulin resistance in obesity and diabetes are thought to have an environmental component. Studies aim to identify how this can cause epigenetic changes that might affect disease progression.

CANCER: S. GSCHMEISSNER/SPL. BRAIN: INGRAM PUBLISHING/GETTY. CHRONIC: G. MURTI/SPL. PRENATAL: S. GSCHMEISSNER/SPL

GENE EXPRESSION

Epigenome effort makes its mark

Major release of maps charting non-genetic modifications goes beyond DNA in a bid to beat complex human disease.

BY ALLA KATSNELSON

When the human genome was first fully sequenced, it was often described as the recipe for making a person. In reality, the genome is more like an entire cookbook that can produce hundreds of different cell types and a staggering range of cell functions depending on which genes are switched on and off. That switching is accomplished using a vast suite of epigenetic marks — molecular and structural modifications to DNA that do not change the underlying sequence but ensure that the right genes are expressed at the right time.

This week, the Roadmap Epigenomics Project, a US$170-million effort to identify and map those marks — known collectively as the human epigenome — begins its first comprehensive data release. Although it is not the only such effort worldwide (see *Nature* **463**, 596–597; 2010), the US National Institutes of Health (NIH) epigenomics project is one of the most ambitious. The newly released data include more than 300 maps of epigenetic changes in 56 cell and tissue types, and represent a significant step towards the complete epigenome — the full picture of all the ways in which DNA can be modified, thus revealing the influence of epigenetics on cell development and its role in complex diseases (see graphic).

Various epigenetic mechanisms regulate gene expression. These include different types of modification on the histone proteins around which genomic DNA winds; attachment of methyl groups to the nucleotide cytosine in DNA, an alteration that is thought to switch off genes; sites of high sensitivity to an enzyme called DNase I, which cleaves accessible DNA and marks the location of gene regulatory regions; and RNA transcription, which, although not a DNA mark, is one measure of the global epigenetic state, revealing how much protein a particular gene makes in different cells. The NIH project developed a standardized protocol for measuring these four factors, and four designated centres around the United States have been charged with making reference maps of each type of modification in embryonic stem cells, induced pluripotent cells and in hundreds of primary adult and fetal tissues.

➲ NATURE.COM
Global consortium sets target of 1,000 epigenomes.
go.nature.com/FaEkfi

The project, slated to run for another five years, aims to produce maps for "a broad swathe of cell types that would be useful to disease research, fund work on specific diseases and develop novel technologies," says John Satterlee, a behavioural geneticist at the National Institute on Drug Abuse in Rockville, Maryland, and one of the coordinators of the project.

A VARIABLE RESPONSE

Some scientists have been wary of the mapping component's 'big science' approach, fearing it will churn out data without ties to the biological questions it is meant to address. Others have questioned the idea that reference maps can be useful to scientists who study specific diseases. Researchers would still have to make their own maps using cells from people without disease, because most studies compare patients to healthy controls who are matched for factors such as age or sex, says John Greally at the Albert Einstein College of Medicine in New York. His projects on epigenomic factors that affect the developing fetus and that cause kidney disease receive funding from the NIH initiative.

Moreover, adds Greally, whereas the four US mapping centres use highly sophisticated techniques to produce their reference maps, individual labs mostly use simpler, cheaper methods to determine epigenetic marks, so comparing their data to the reference maps may be tricky. "Having the [mapping] information is valuable in itself," he says, "but the focus has got to be on how you use this to understand disease."

So far, the wider community of researchers has largely been unaware of the effort's existence. "We've sort of been in stealth mode," says Joseph Ecker, a plant and molecular biologist at the Salk Institute for Biological Studies in La Jolla, California, whose lab is working with a mapping centre to produce reference maps of DNA methylation. The newly released data should push the project to a point where the information can be widely used by researchers in different fields, he says.

Satterlee notes that the mapping component is just one arm of the project, making up about $57 million of its total budget. The rest of the money is going to individual investigator grants, 53 of which have already been awarded. Indeed, says cell biologist Benjamin Tycko of Columbia University in New York, whose work on the role of DNA methylation in Alzheimer's disease is funded by the project, "they've actually adopted a small lab approach" by funding research in labs with expertise on different diseases. ■

REGULATION

Safety authority rebuts conflict claim

Greens highlight industry links in bid to discredit European food agency.

BY DECLAN BUTLER

Green groups have long alleged that the European Food Safety Authority's (EFSA's) scientific risk assessments of food, animal feed and agrochemicals tend to be too industry friendly. But the criticisms intensified after last week's accusation of a conflict of interest at the very top of the organization.

On 29 September, José Bové, a French farmer and a representative of the Europe Ecology party in the European Parliament, alleged that Diána Bánáti — chair of the EFSA management board from October 2008 to 30 June this year — had failed to declare that in April 2010 she had joined the European board of directors of the International Life Sciences Institute (ILSI), a non-governmental organization that is funded by food companies and seeks to coordinate and fund research and risk assessment.

➲ NATURE.COM
For a longer version of this story, see go.nature.com/echhvp

Bové provided no evidence that Bánáti's ILSI links had influenced her actions at EFSA, and the board has no direct role in the agency's risk assessments. EFSA points out that it has strict safeguards on conflicts of interest, but admits that the publicity over the affair risks harming public perception of its impartiality. Catherine Geslain-Lanéelle, EFSA's executive director, says that the agency has "learned lessons" from the recent attacks, and is considering introducing tougher rules for the chairs and vice-chairs of panels. "Public perception is important; we need trust," she says.

Bánáti is director general of Hungary's Central Food Research Institute in Budapest, and she remains on the EFSA board — a part-time post that is unpaid. EFSA says that the omission in Bánáti's declaration of interest — which was corrected last week — was an oversight on its part, and notes that ever since Bánáti joined the EFSA board in 2006, her membership of ILSI's scientific and advisory committee has been declared on EFSA's website.

Several other EFSA officials have links with ILSI. For example, Milan Kováč, a prominent food scientist and government official in the Czech Republic, who is on the boards of EFSA and ILSI, is also a member of the scientific advisory board of the European Food Information Council, which is largely funded by food companies. Many EFSA scientific opinions also cite ILSI-sponsored studies.

ILSI insists that it is not a lobbying group. In a statement last week, the institute said that it seeks out top talent, and that it was "not surprising" that the same individuals are sought by other public-health bodies. ■

CORRECTIONS

In 'Science safe in Brazil elections' (*Nature* **467,** 511–512; 2010) the graph should have read US$ billions not US$ millions.

In 'Brawl in Beijing' (*Nature* **467,** 511; 2010), the correct link to Fang Shimin's website is www.xys.org.

ESA/AOES MEDIALAB

WEIGHING THE WORLD

After a near-death crisis, the best gravity sensor in space is back to full strength, providing data that will keep scientists on the level.

BY QUIRIN SCHIERMEIER

On 18 July, geophysicist Reiner Rummel received a phone call that made his heart sink. The European Space Agency (ESA) had stopped receiving data from a €350-million (US$471-million) satellite that Rummel had spent nearly 20 years designing, building, testing and shepherding into orbit. The craft, burdened with the unwieldy name of the Gravity Field and Steady-State Ocean Circulation Explorer (GOCE), had been in space for little more than a year. Now it looked as if it might be lost for good, undone by a glitch in the communications system that sends data back and forth from the satellite to the ground.

Rummel, who works at the Technical University of Munich in Germany and is joint principal investigator of the GOCE project, got little sleep throughout the summer as he and his colleagues tried various strategies to restore the on-board computer system, which had failed once before. Mission scientists developed software to combine the system with a back-up computer so that they could fix the apparent problem, but still the craft failed to respond. Then, in early September, ground controllers tried something new. They sent a signal to raise the temperature of GOCE's computer compartment.

At stake was the most finely tuned gravity sensor ever to fly in space. GOCE was designed to map the subtle gravitational differences that arise across the globe because Earth's mass of roughly six sextillion (10^{21}) tonnes is not distributed evenly. Using data collected by the small satellite, researchers planned to construct a sophisticated gravitational map called the geoid, accurate to the nearest centimetre: a fivefold improvement over previous efforts to map gravity from space. Such data would provide geoscientists with a global reference for precisely measuring the heights of continents, mountain peaks and the ocean surface, which is rising because of global warming.

GOCE data could also reveal the scars from a giant extraterrestrial impact 250 million years ago, keep tabs on the shifting tectonic plates that cause huge earthquakes, and help to measure the strength of ocean currents such as the Gulf Stream — critical information that will improve climate forecasts.

All this explains why researchers were mightily relieved when, as the GOCE computer warmed up, it sputtered back to life.

A MASSIVE MISSION

GOCE races across the skies at 30,000 kilometres per hour, skimming the edge of Earth's atmosphere at an altitude of about 250 kilometres. The craft has to maintain such a low orbit to make its measurements because gravity falls off rapidly with distance. But flying through the atmosphere creates drag on the satellite, so project engineers have given it an unusually aerodynamic design and an ion engine to counteract the air friction. From its low orbit, GOCE's main on-board instrument, a gravity gradiometer, is sensitive enough to measure the gravitational tug of giant reservoirs on Earth's surface.

The heart of the sensor consists of three pairs of cubes, arranged perpendicular to each other to form a three-dimensional cross. As the satellite passes overhead, the cubes each feel a different pull from the mountains and other masses on and below the planet's surface. An electrostatic control system measures those differences with such precision that GOCE can detect changes on the order of one-millionth of the average strength of Earth's gravity at the surface (see 'The pull of the planet').

Because the gradiometer is so sensitive, engineers could not truly test it before launch. They simulated zero gravity by dropping the sensor within a tall tower, but this was not an ideal test. "We're basically flying a prototype in orbit," says Rummel.

After GOCE was launched in March 2009, it took several months to fix problems such as difficulties in orienting the satellite with the required accuracy. But the team eventually achieved the planned sensitivity for the gravity measurements, and released its first big batch of data at an ESA symposium in Norway in June. If no further glitches occur, GOCE will complete its planned mission in April 2011.

"It's a tremendous achievement," says Philip Woodworth, a sea-level expert at the National Oceanography Centre in Liverpool, UK. "GOCE-derived gravity data will be usable

for decades."

Researchers are already turning the initial harvest of data into a highly accurate geoid, a mathematically derived surface resembling a hypothetical mean global sea level. Real sea levels respond to winds, currents and other dynamic features of the planet. The geoid, by contrast, reflects what the ocean surface would look like if the world were covered by a static skin of water whose height is influenced only by gravity. It is a bulbous affair, with bumps and dips matching mountains, ocean trenches and density variations deep within the planet. At every point, the force of gravity operates perpendicular to the surface of the geoid.

The geoid was first conceived of in 1828 by German mathematician Carl Friedrich Gauss, who recognized that surveyors would need such a reference surface to determine the precise elevation above sea level for any point on Earth. To this day, uncertainties in the height of the geoid make it difficult to compare altitude measurements for different parts of Earth's surface.

GOCE will finally remedy this. The global geoid derived from the gravity measurements taken since April shows, among other things, a pronounced 'depression' in the Indian Ocean and 'plateaus' in the North Atlantic and western Pacific — mirroring convective activity and density anomalies in Earth's mantle. More of these delicate features will be added to the map in the coming months.

CURRENT EVENTS

Geodesists — scientists who study Earth's shape — have for the past few years been using a different space mission to map the planet's gravity field. The Gravity Recovery and Climate Experiment (GRACE), a joint mission of NASA and the German space agency, was launched in 2002 and uses a pair of satellites to measure relatively large-scale gravity variations, such as the loss of mass from melting sections of the Greenland ice cap. GOCE adds detail to the picture, with gravity measurements five times more precise than those of GRACE and with almost three times higher spatial resolution. Armed with its data, researchers will be able to distinguish between the ocean topography shaped by gravity and that of the 'hills' and 'valleys' of water created by wind, pressure gradients and Earth's rotation.

That will help scientists to measure ocean circulation, because large currents are slightly higher than the surrounding ocean, with the topographic height of the current directly related to its strength. "You can think of the Gulf Stream as a gentle hill that you'd need to climb if you could walk from Boston to the Bermudas," explains Woodworth. "If you're able to accurately measure the shape of the hill you can calculate the mean strength of the flow."

GOCE-derived data should therefore allow researchers for the first time to calculate the mean strength of the Gulf Stream, which carries warm water north from the Gulf of Mexico towards the Arctic. Such data on this and other currents are essential for improving computer models of the ocean and the atmosphere.

Geoid information should also help to determine the mean sea level to within a few millimetres, something not possible before. Local sea-level changes, such as those from one year to the next, can be measured by radar altimetry and tide gauges. But geodesists need an accurate global reference to compare measurements on different continents.

Even regional comparisons are difficult without an accurate geoid. For example, the UK National Tide Gauge Network suggests that the sea level off the north coast of Scotland is 30–40 centimetres lower than that off the coast of Cornwall, a difference that is almost certainly an artefact caused by levelling errors, says Woodworth.

The high-resolution gravity data from GOCE will also help to measure ice loss from Antarctica and Greenland, and can aid geologists in assessing the potential for large earthquakes. The biggest shocks, such as the magnitude-8.8 quake that struck Chile in February, rearrange Earth's tectonic plates, causing a change in the gravity field that can be sensed from space. GOCE data will give researchers new clues as to how masses of different densities readjust before and after large seismic events — and whether it is possible to measure gravity effects before the ground ruptures, says Roberto Sabadini, a geophysicist at the University of Milan in Italy, and a participant in GOCE.

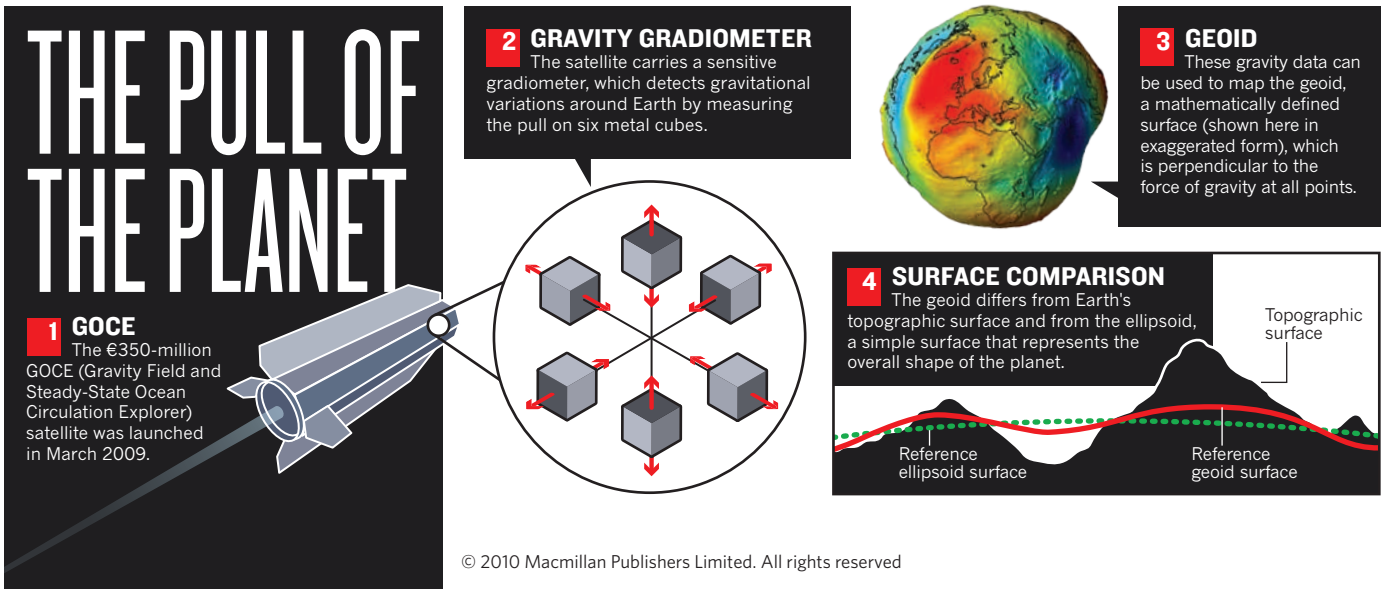
The gravity data may even help to shed light on a planetary disaster: the 'Great Dying' that wiped out a large fraction of existing species 250 million years ago, at the end of the Permian period. Ralph von Frese, a geophysicist at Ohio State University in Columbus, belongs to a school of scientists who suspect that this mass extinction was triggered by the impact of a giant meteorite, more than twice the size of the one thought to have killed the dinosaurs 65 million years ago. Their search for the crater left by this enormous impact focuses on eastern Antarctica, where GRACE data and airborne-radar imagery indicate there might be a depression, 500 kilometres in diameter, under the 2–3-kilometre-thick ice in Wilkes Land (R. R. B. von Frese *et al. Geochem. Geophys. Geosyst.* **10,** Q02014; 2009).

THE GRAVITY DATA MAY EVEN HELP TO SHED LIGHT ON A PLANETARY DISASTER

The impact would probably have lifted high-density mantle material into Earth's crust, causing a lasting gravity anomaly. But because the GRACE signal is blurry, von Frese is eagerly awaiting GOCE data that could back his theory.

He and others are crossing their fingers that the fragile satellite will stay healthy for the next six months and complete its mission without any more drama. ■

Quirin Schiermeier *is a senior reporter with* Nature *in Munich, Germany.*



GLOBE: ESA

Amateur hobbyists are creating home-brew molecular-biology labs, but can they ferment a revolution?

Life hackers

BY HEIDI LEDFORD

Rob Carlson's path to becoming a biohacker began with a chance encounter on the train in 1996. Carlson, a physics PhD student at the time, was travelling to New York to find a journal article that wasn't available at his home institution, Princeton University in New Jersey. He found himself sitting next to an inquisitive elderly gentlemen. Carlson told him about his thesis research on the effects of physical forces on blood cells, and at the end of the journey, the stranger made him an offer. "You should come work for me," said the man, "I'm Dr Sydney Brenner." The name meant little to Carlson, who says he thought: "Yeah, OK. Whatever, 'Dr Sydney Brenner.'"

It wasn't until Carlson got back to Princeton and asked a friend that he realized that "Dr Sydney Brenner" was a famed biologist with a knack for transforming the field. He took the job.

Within a year, Carlson was working with a motley crew of biologists, physicists and engineers at Brenner's Molecular Sciences Institute (MSI) in Berkeley, California, learning molecular biology techniques as he went along. The institute was a hotbed of creativity, and reminded Carlson of the scruffy hacker ethos that had spurred the personal-computing revolution just 25 years earlier. He began to wonder if the same thing could happen for biotechnology. What if a new industry, even a new culture, could be created by giving everyone access to the high-tech tools that he had at his fingertips? Most equipment was already for sale on websites such as eBay.

Carlson penned essays and articles that fanned the embers of the idea. "The era of garage biology is upon us," he wrote in a 2005 article in the technology magazine *Wired*. "Want to participate?" The democratization of science, he reasoned, would bring in new talent to build and improve scientific instrumentation, and maybe help to uncover new industrial applications for biotechnology. Eventually, he decided to follow his own advice, setting up a garage lab in 2005. "I made the prediction," he says, "so I figured maybe I should do the experiment."

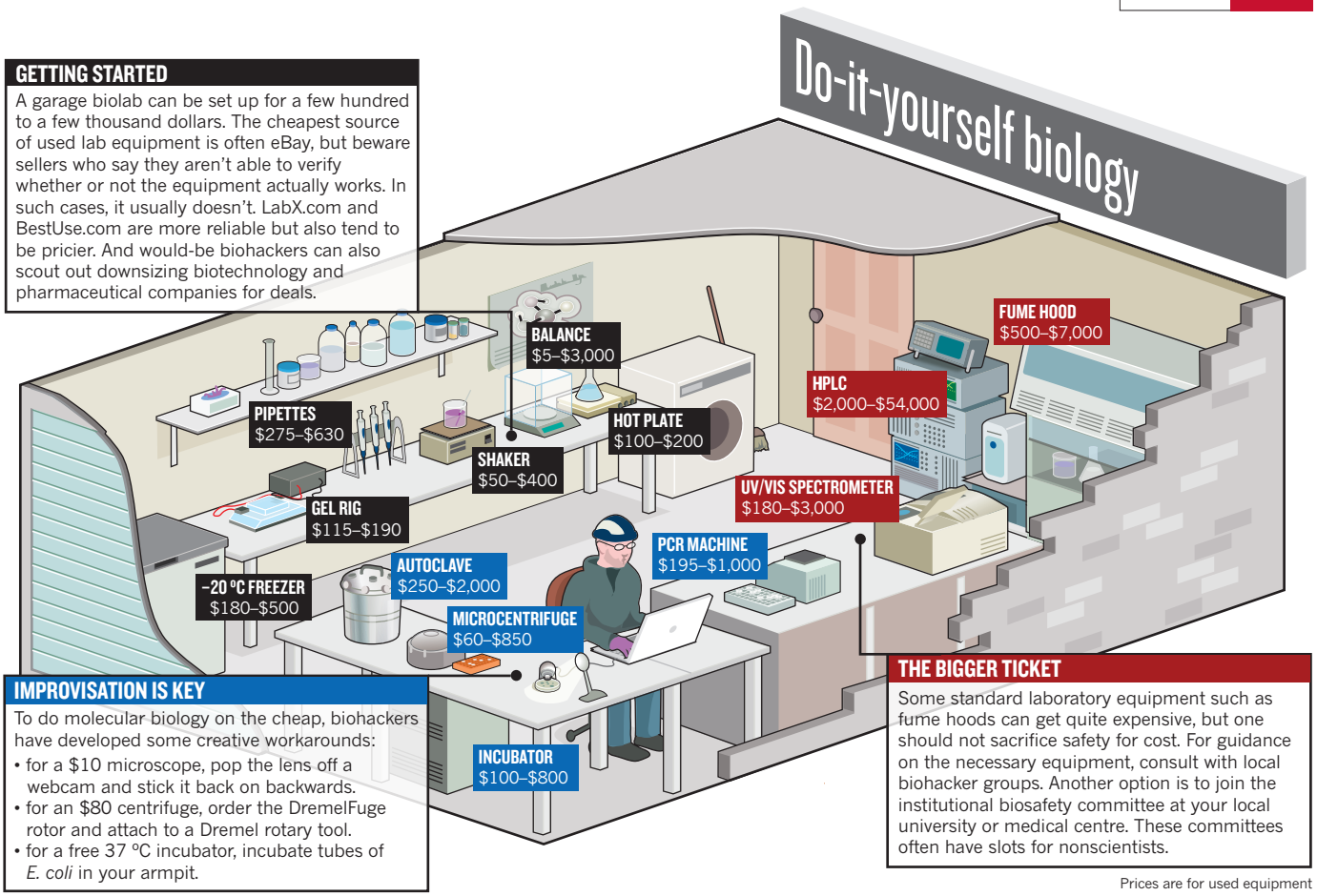
Carlson is not alone. Would-be 'biohackers' around the world are setting up labs in their garages, closets and kitchens — from professional scientists keeping a side project at home to individuals who have never used a pipette before. They buy used lab equipment online, convert webcams into US$10 microscopes and incubate tubes of genetically engineered *Escherichia coli* in their armpits. (It's cheaper than shelling out $100 or more on a 37 °C incubator.) Some share protocols and ideas in open forums. Others prefer to keep their labs under wraps, concerned that authorities will take one look at the gear in their garages and label them as bioterrorists.

For now, most members of the do-it-yourself, or DIY, biology community are hobbyists, rigging up cheap equipment and tackling projects that — although not exactly pushing the boundaries of molecular biology — are creative proof of the hacker principle. Meredith Patterson, a computer programmer based in San Francisco, California, whom some call the 'doyenne of DIYbio', made glow-in-the-dark yogurt by engineering the bacteria within to produce a fluorescent protein. Others hope to learn more about themselves: a group called DIYgenomics has banded together to analyse their genomes, and even conduct and participate in small clinical trials. For those who aspire to change the world, improving biofuel development is a popular draw. And several groups are focused on making standard instruments — such as PCR machines, which amplify segments of DNA — cheaper and easier to use outside the confines of a laboratory, ultimately promising to make DIYbio more accessible.

> "We're making $10 microscopes and the discussion around us is about weaponized anthrax."

Many traditional scientists are circumspect. "I think there's been a lot of overhyped and enthusiastic writing about this," says Christopher Kelty, an anthropologist at the University of California, Los Angeles, who has followed the field. "Things are very much at the beginning stages." Critics of DIY biology are also

Do-it-yourself biology



GETTING STARTED

A garage biolab can be set up for a few hundred to a few thousand dollars. The cheapest source of used lab equipment is often eBay, but beware sellers who say they aren't able to verify whether or not the equipment actually works. In such cases, it usually doesn't. LabX.com and BestUse.com are more reliable but also tend to be pricier. And would-be biohackers can also scout out downsizing biotechnology and pharmaceutical companies for deals.

IMPROVISATION IS KEY

To do molecular biology on the cheap, biohackers have developed some creative workarounds:

- for a $10 microscope, pop the lens off a webcam and stick it back on backwards.
- for an $80 centrifuge, order the DremelFuge rotor and attach to a Dremel rotary tool.
- for a free 37 °C incubator, incubate tubes of *E. coli* in your armpit.

THE BIGGER TICKET

Some standard laboratory equipment such as fume hoods can get quite expensive, but one should not sacrifice safety for cost. For guidance on the necessary equipment, consult with local biohacker groups. Another option is to join the institutional biosafety committee at your local university or medical centre. These committees often have slots for nonscientists.

Prices are for used equipment

dubious about whether there is an extensive market for garage molecular biology. No one needs a PCR machine at home, and the accoutrements to biological research are expensive, even if their prices fall daily (see graphic). Then again, the same was said about personal computers, says George Church, a geneticist at Harvard Medical School in Boston, Massachusetts. As a schoolboy, he says, he saw his first computer and fell in love. "Everybody looked at me like, 'Why on earth would you even want to have one of those?'"

Carlson started his garage lab as something of a hobby, but he needed to do it without sapping resources from his lab at the University of Washington in Seattle. He bought equipment such as refurbished micropipettes — a staple in any molecular biology lab — and a used centrifuge on eBay. In 2007, fed up with grant applications and eager to spend more time working in his garage lab, he gave up his position at the university altogether.

Carlson decided to follow up on work at the MSI. There, he had been part of a team developing a way to quantify small amounts of proteins in single cells using 'tadpoles', in which a protein 'head' is attached to a DNA 'tail'. The head was designed to bind to a protein of interest, and the DNA tail could be amplified and quantified by PCR, allowing researchers to calculate the number of proteins present (see *Nature Meth.* **2,** 31–37; 2005). The tadpoles have economic potential, providing an alternative to the standard approach of using fluorescently tagged antibodies, which provide at best only rough estimates of protein levels. But the original formulation was too expensive to commercialize, says Carlson. "If I could use this protein in the garage in a simple way to show that it would work, then hopefully it would be a product that would be useful in a low-tech setting, out in the field or in a doctor's office," he says.

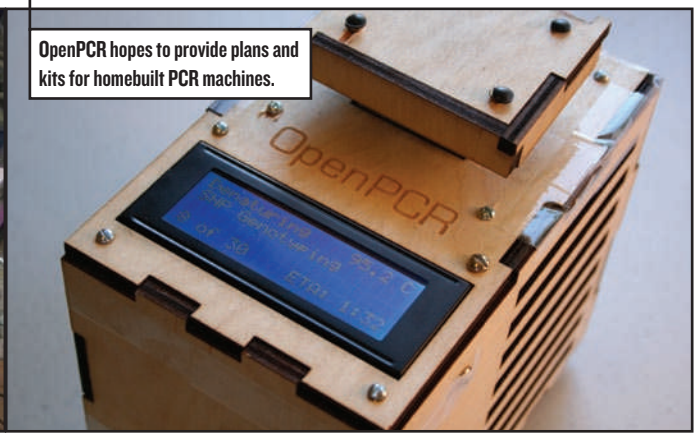
As Carlson worked, the idea of garage biohacking was taking off. In May 2008, Jason Bobe, director of community outreach for the Personal Genome Project at Harvard Medical School, and Mackenzie Cowell, a web developer in Cambridge, Massachusetts, organized the first meeting of DIYbio at the Asgard Irish pub, up the road from the Massachusetts Institute of Technology. About 25 people turned up. Two years later, there are more than 2,000 subscribers on the DIYbio e-mail list.

No one knows how many of those 2,000 are serious practitioners — Bobe jokes that 30% are spammers and the other 70% are law-enforcement officials keeping tabs on the community. But many DIY communities are coalescing: not only in Cambridge, but also in New York, San Francisco, London, Paris and the Netherlands. Some of these aim to develop community lab spaces with equipment that users could share for a monthly fee. And several are already affiliated with local 'hacker spaces', which provide such services to electronics enthusiasts. For example, the New York DIYbio group meets every week at the work-space of an electronics-hacker collective called NYC Resistor, which now has a few pieces of basic molecular biology equipment, including a PCR machine.

DIYbio is an offshoot of the open-science movement, which encourages an open exchange of materials, data and publications and has its origins in the push for open-source software in the 1990s, says Kelty. Many biohackers are also keen to tackle projects that involve engineering cells by piecing together new genetic circuits, an approach often called 'synthetic biology'. DIYbio has picked up both momentum and stigma from this field, which has been alternately hyped and decried as the solution to society's ills or the nursery for a bioterrorist scourge. The thought of hundreds of biohackers creating pathogens in unmonitored garage biology labs set off alarm bells, and in 2009, the Federal Bureau of Investigation (FBI) began sending representatives from its directorate for weapons of mass destruction to DIYbio conferences.

Biohackers are wary. They recall what happened to Steve Kurtz, an artist who was using bacteria shipped to him by a Pittsburgh geneticist. In 2004, federal agents stormed his house in hazmat suits with guns drawn. Kurtz was arrested and saddled with mail-fraud charges

Rob Carlson in his chilly garage biolab in 2005.



OpenPCR hopes to provide plans and kits for homebuilt PCR machines.

S. KELLER (LEFT); J. PERFETTO (RIGHT)

that took him four years to clear. Bobe has interacted with and advised the FBI, but says he finds many of the biosecurity fears of the FBI and the public to be unfounded. "The amateur activity right now is at the seventh- or eighth-grade level," he says. "We're making $10 microscopes and all of the discussion around us is about weaponized anthrax. Sure we're concerned about that just like everybody else, but I don't know what to say except 'Yeah, that sounds scary as hell. Let's be sure nobody does that.'"

The FBI seems to have taken that message on board, and has adopted what some call a 'neighbourhood watch' stance. The approach relies on biohackers monitoring their own community and reporting behaviour they find threatening, says Edward You, a special agent in the FBI's bioterrorism unit.

Carlson's projects are more advanced than those of the average DIYbio hobbyist, and he has found that the garage-hacker ethos eventually suffered. He says he sometimes found it hard to persuade companies to deliver lab supplies to a residential address. Carlson also wanted his garage back to restore a boat.

So, Carlson and his business partner, engineer Rik Wehbring, moved their lab out of the garage and into a small commercial space in 2009. The two fund the space and their experiments through a small consulting firm called Biodesic. Through the firm, they have advised companies on a range of technology issues from biosecurity to designing brainwave-based game controllers.

Other biohackers have also come up with creative ways to fund their projects. Several have used websites such as Kickstarter, which allows inventors to post their projects and funding targets online. Visitors to the site make donations, usually small ones, but the hope is that enough visitors making tiny contributions will add up. Two California garage biohackers, Tito Jankowski and Josh Perfetto, used Kickstarter to fund the development of a small, low-cost PCR machine known as OpenPCR. They reached their fundraising goal of $6,000 in ten days. By the time their Kickstarter listing closed 20 days later, they had doubled that figure. Another group of biohackers used Kickstarter to raise funds for a hackerspace called BioCurious, based in Silicon Valley, California. They raised more than $35,000.

But all of this is tiny compared to the cost of launching an actual business. Joseph Jackson, a self-proclaimed "professional entrepreneur-slash-activist" from Mountain View, California, and Guido Nunez-Mujica, a computational biologist from Venezuela, have teamed up with other hackers to build a portable PCR machine known as LavaAmp, which can be run from a computer's USB port. The team has poured tens of thousands of dollars into the project, says Jackson, but will need closer to $100,000 to achieve its goal of producing PCR machines that could be used by hobbyists, teachers and by researchers in developing countries.

Jim Collins, a synthetic biologist at Boston University, says that the costs of doing molecular-biology research make the comparison between amateur biologists and the hackers who drove the personal-computer revolution inappropriate. There's a vast chasm between these tinkerers and those with access to a traditional lab. Faculty members, Collins says, typically ask for hundreds of thousands of dollars from a university to start a molecular-biology lab. Smart amateurs might be able to bring fresh perspective, he says, but they face an uphill battle. "I'm not saying you need to be appropriately pedigreed. I'm saying you need to be appropriately resourced."

Carlson says that the cost of biological research is decreasing. "The predominant thought about biology used to be that it was expensive and hard," he says. "And it's still hard. It's just not so expensive." In 2003, he projected the falling costs of sequencing and synthesis of DNA and proteins, and the accelerating pace of research into areas such as protein structure determination (R. Carlson *Biosecur. Bioterror.* **1,** 1–12; 2003). His predictions echo Moore's law of computing, and some have dubbed them the 'Carlson curves'.

But the curve trajectory isn't as steep as Carlson might like. He has redesigned the protein heads of his tadpoles, and decided early on that instead of producing the protein himself — an expensive and arduous process — he would pay a company to make it for him. He could either buy cheap protein that was contaminated with other proteins, for about $3,000, or buy clean protein for about $50,000. "There was nothing in between," he said. He took the cheap route, but found that the batches he received weren't clean enough to publish his results or start selling the finished tadpoles. The project stalled.

A few months ago, Carlson realized that more protein-synthesis companies had entered the scene, including several that filled the middle range pricing gap. He ordered a fresh batch of protein that was supposed to arrive more than a month ago, but still hasn't been delivered. "If we had a million dollars in the bank, this problem would have been solved a long time ago," he says. "And if I had an experienced biochemist or molecular biologist at the bench for a year or two it probably would have cost the same and would have been done faster."

> "I'm not saying you need to be appropriately pedigreed, just appropriately resourced."

Still, five years after taking science into his garage, Carlson says he's convinced that biohacking has the potential to trigger a technological revolution. "We're going to see a lot more at the garage level that will produce a variety of products in the marketplace, one way or another," he says.

Once his tadpoles have been optimized, Carlson hopes that publishing his work will attract further investors. Meanwhile, he feels his experiment in garage-based innovation has so far been a success, despite the delays and personal sacrifices. "Part of the exercise was to determine whether or not we could bootstrap this thing," he says. "The answer appears to be 'yes'. As long as you are willing to be patient and to eat nothing but rice for dinner occasionally." ■ SEE EDITORIAL P.634

Heidi Ledford *writes for* Nature *from Cambridge, Massachusetts.*

COMMENT

D. PARKINS

Altruism researchers must cooperate

Biologists studying the evolution of social behaviour are at loggerheads. The disputes — mainly over methods — are holding back the field, says **Samir Okasha**.

Last month, 30 leading evolutionary biologists met in Amsterdam to discuss a burgeoning controversy. The question of how altruistic behaviour can arise through natural selection, once regarded as settled, is again the subject of heated debate.

The question dividing biologists is the degree to which inclusive fitness theory, or kin selection, explains the evolution of altruism — in which an animal provides a benefit to another at a cost to itself. This theory, that natural selection can sometimes favour animals that behave altruistically towards relatives, has dominated empirical work on social behaviour since it was devised by W. D. Hamilton in the 1960s and 1970s[1]. Yet some biologists are now calling for a radical rethink, arguing that kin selection is theoretically problematic, and has insufficient empirical support, and that alternative models better account for the evolution of social behaviour[2]. Others regard kin selection as solid, and the rethink as unnecessary and potentially retrograde.

Rival camps have emerged, each endorsing a different approach to social evolution. Heated exchanges have occurred at conferences, on blogs and in journals, and have even been reported in *The New York Times*. Biologists have accused each other of misunderstanding, of failing to cite previous studies appropriately, of making unwarranted claims to novelty and of perpetuating confusions. Yet I contend that there is little to argue about.

Much of the current antagonism stems from the fact that different researchers are focusing on different aspects of the same phenomenon, and are using different methods. In allowing a plurality of approaches — a healthy thing in science — to descend into tribalism, biologists risk causing serious damage to the field of social evolution, and potentially to evolutionary biology in general.

DARWIN'S PUZZLE

Charles Darwin realised that altruism poses a special problem for his theory of evolution. He was particularly troubled by the sterile workers in colonies of social ▶

Examples of favouring kin are widespread: Japanese macaques spend more time grooming their closer relatives and meerkats share feeding and guard duties.

I. ARNDT/NATUREPL.COM

T. DRESSLER/ARDEA.COM

insects, which devote their lives to helping a queen reproduce at the expense of having offspring themselves. One possible explanation, hinted at by Darwin, is that groups containing many altruists might out-compete groups containing fewer. This idea of 'group selection' fell out of favour in the 1960s when George Williams argued that it was unlikely to be a powerful evolutionary force compared to individual selection, and was not needed to explain empirical observations[3].

Inclusive fitness theory, most biologists now believe, provides the solution to Darwin's puzzle. Hamilton realised that a gene that causes an animal to behave altruistically can spread by natural selection as long as the beneficiaries are relatives, and so have a chance of carrying the same gene. In short, altruism can evolve if the cost to the actor is offset by sufficient benefit to sufficiently closely related recipients. This means that animals should behave in ways that maximize not their personal fitness (or number of surviving offspring), but rather their inclusive fitness — a measure that also takes into account the offspring of their relatives.

"Many biologists regard kin selection theory as a resounding empirical success."

Inclusive fitness theory predicts that animals should behave more altruistically towards kin than non-kin. This has been amply confirmed in diverse species, from microbes to primates[4], leading many biologists to regard kin selection theory as a resounding empirical success. In many bird species, such as scrub jays and dunnocks, for example, breeding pairs receive aid from a non-breeding 'helper' bird, typically a relative[5]. Similarly, rhesus and Japanese macaques are more likely to groom relatives than non-relatives, and to help them in disputes[6].

Several biologists, however, have recently questioned the importance of kin selection in explaining social behaviour. Edward O. Wilson, famous for his empirical work on insect societies and once a forceful advocate of kin selection, now argues that kinship plays a minor part in the evolution of ant, bee, termite and other social insect colonies[7–9]. More important, he says, are the ecological factors that make social living so successful. An easy-to-defend nest and a nearby food supply, for instance, may make it beneficial for animals to live in groups. Recently, Wilson, along with theoretical biologists Martin Nowak and Corina Tarnita, have argued that inclusive fitness theory rests on a number of assumptions that greatly limit its applicability — such as that natural selection is relatively weak[2]. Still others argue that multi-level selection — a modern-day version of group selection — best explains the evolution of altruism (although many biologists remain suspicious of appeals to group, rather than individual, advantage).

The root of the problem is the existence of several different frameworks for modelling the evolution of social behaviour. These include numerous variants of kin selection theory; multi-level selection; evolutionary game theory; and an approach from quantitative genetics based on the notion of 'indirect genetic effects'. The relationships between these frameworks are sometimes ambiguous, and biologists disagree about which is most fundamental and which most useful empirically (see 'Coming to terms with social evolution').

All this disagreement creates the impression of a field in massive disarray. In reality, many of the players involved are arguing at

GLOSSARY

Coming to terms with social evolution

Altruism
Behaviour that is costly for an animal to perform but benefits others. Costs and benefits are measured in terms of reproductive success.

Inclusive fitness
A generalization of Darwinian fitness, which takes into account the effect of an individual's actions on the reproductive success of their relatives, as well as on their own.

Kin selection theory
The idea that natural selection shapes individuals' behaviour according to the effect it has on relatives.

Group selection
The idea that natural selection favours traits because they benefit whole groups, rather than individuals.

Multi-level selection theory
The idea, closely related to group selection, that natural selection can operate on more than one hierarchical level, for instance at the level of the individual, group or species.

Social evolution
Darwinian theory as applied to an animal's social behaviour, that is behaviour affecting other individuals in the population.

Eciton burchelli **army ant workers can form living bridges for other colony members to cross. Older moorhen chicks sometimes help feed their young siblings.**

M. MOFFETT/MINDEN PICTURES/FLPA

NATURE PRODUCTION/NATUREPL.COM

cross purposes. Nowak and his colleagues, for instance, have developed a mathematical model that they claim provides a more direct way to calculate the evolutionary dynamics of a social trait such as altruism[2]. However, they overlook the fact that inclusive fitness theory explains what organisms are trying to maximize. It is not just a tool for calculating when a social trait will evolve.

Likewise, in arguing that ecological factors, rather than kinship, are key to the evolution of social-insect colonies, Wilson is imposing a false dichotomy[4]. To fully understand how these colonies evolve, researchers need to consider ecological factors and relatedness. Whether they stress the importance of one over the other will depend on the question they are asking. For example, relatedness has proved crucial to understanding conflicts between the queen and her workers over the production of male versus female offspring in ants, bees and wasps. For questions about how tasks are allocated to the workers in an ant colony or why the size of colonies differs across species, ecological factors are probably more relevant.

Lastly, kin and multi-level selection are not alternative theories; they simply offer different takes on the question of how social behaviour evolved. Proponents of kin selection, for example, explain sterile workers in insect colonies by saying that the workers are helping the queen to reproduce, and thus boosting their own inclusive fitness. Proponents of multi-level selection argue that the workers are providing a benefit to the colony as a whole, thus making the colony fitter than other colonies. These explanations may seem different, but mathematical models show that they are in fact equivalent[10–12].

At the Amsterdam meeting, certain real disagreements did surface, but they were mostly over technical points and pitted against a background of broad agreement over fundamentals. Most agreed that inclusive fitness theory has been extremely valuable for empirical biologists, but that it is not the only way to model social evolution.

BUILDING BRIDGES

Much of the current antagonism could easily be resolved — for example, by researchers situating their work clearly in relation to existing literature; using existing terminology, conceptual frameworks and taxonomic schemes unless there is good reason to invent new ones; and avoiding unjustified claims of novelty or of the superiority of one perspective over another.

It is strange that such basic good practice is being flouted. The existence of equivalent formulations of a theory, or of alternative modelling approaches, does not usually lead to rival camps in science. The Lagrangian and Hamiltonian formulations of classical mechanics, for example, or the wave and matrix formulations of quantum mechanics, tend to be useful for tackling different problems, and physicists switch freely between them.

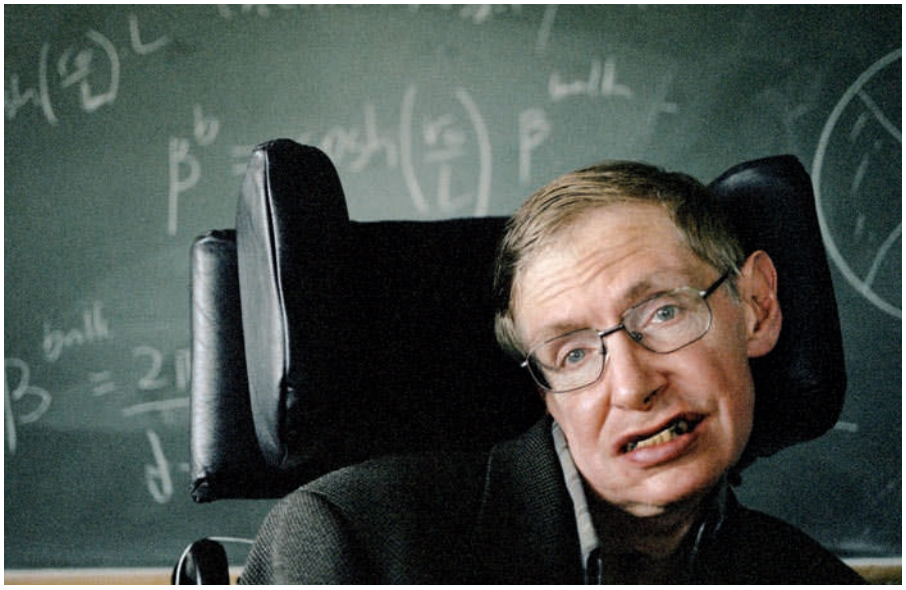
"Evolutionary biology is peculiarly susceptible to controversy and infighting."

History shows that, despite its enormous empirical success, evolutionary biology is peculiarly susceptible to controversy and infighting. This is particularly true of social evolution theory, in part because of its potential applications to human behaviour. In the 1970s and 1980s, for instance, left-wing scholars bitterly rejected biological explanations for phenomena such as religion and homosexuality, because they feared such explanations would be used to justify a continuation of existing inequalities.

Researchers should take stock before another overblown dispute does serious damage to the field. Up-and-coming researchers are unlikely to be attracted to a discipline plagued by controversy. Moreover, if the experts cannot agree about what theoretical framework works best, the supply of research funding may eventually be threatened. Also worrying is the possibility that onlookers perceive the central question of social evolution theory — how altruism can evolve — as unresolved, even though it was answered decades ago. During the 'sociobiology wars' of the 1970s and 1980s, creationists proved adept at seizing on and exaggerating the differences in opinion between biologists for their own ends. It would be a disaster if the same were to happen again. ■ SEE CORRESPONDENCE P.661

Samir Okasha *is in the Department of Philosophy, University of Bristol, Bristol BS8 1TB, UK. His book* Evolution and the Levels of Selection *was published by Oxford University Press in 2006.*
e-mail: Samir.Okasha@bristol.ac.uk

1. Hamilton, W. D. *J. Theor. Biol.* **7,** 1–16 (1964).
2. Nowak, M. A, Tarnita, C. E. & Wilson, E. O. *Nature* **466,** 1057–1062 (2010).
3. Williams, G. C. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* (Princeton Univ. Press, 1966).
4. Foster, K. R. *Cold Spring Harb. Symp. Quant. Biol.* doi:10.1101/sqb.2009.74.041 (2010).
5. Emlen, S. T. in *Behavioural Ecology: an Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 228–253 4th edn (Blackwell, 1997).
6. Silk, J. in *Moral Sentiments and Material Interests: the Foundations of Cooperation in Economic Life* (eds Gintis, H., Bowles, S., Boyd, R. T. & Fehr, E.) 43–74 (MIT Press, 2005).
7. Wilson, E. O. *Soc. Res.* **72,** 159–168 (2005).
8. Wilson, D. S. & Wilson, E. O. *Q. Rev. Biol.* **82,** 327–348 (2007).
9. Wilson, E. O. & Hölldobler, B. *Proc. Natl Acad. Sci. USA* **102,** 13367–13371 (2005).
10. Lehmann, L., Keller, L., West, S. & Roze, D. *Proc. Natl Acad. Sci. USA* **104,** 6736–6739 (2007).
11. Queller, D. *Am. Nat.* **139,** 540–558 (1992).
12. Dugatkin, L. A. & Reeve, H. K. *Adv. Stud. Behav.* **23,** 101–133 (1994).



A. BUURMAN/EYEVINE

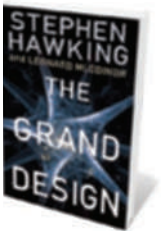
COSMOLOGY

No miracle in the multiverse

Stephen Hawking and Leonard Mlodinow suggest our Universe isn't all that special, finds **Michael Turner**.

Despite publicity to the contrary, *The Grand Design* does not disprove the existence of God. Science has not had much new to say about God since mathematician Pierre-Simon Laplace remarked to Napoleon that he had no need for "that hypothesis" when asked why he had neglected the deity in his treatise *Mécanique céleste* (*Celestial Mechanics*, 1799–1825). Rather, theoretical physicists Stephen Hawking and Leonard Mlodinow offer a brief but thrilling account of some of the boldest ideas in physics — including M-theory and the multiverse — and what these have to say about our existence and the nature of the Universe.

The Grand Design traces the history of science from the sixth-century Greek philosopher Thales of Miletus to the present, with six crucial touch points: the assertion by the Ionians around 600 BC that the world is governed by laws; the discovery of the first simple laws by Archimedes around 200 BC; and Isaac Newton's mathematical expression of his laws of motion and gravity in the 1680s. Then follows Laplace's assertion in the nineteenth century that the world is deterministic and does not need God to run it; Albert Einstein's question in the early twentieth century of whether a creator would have a choice about the laws of nature; and today's expression of those laws in the standard model of particle physics and the theory of general relativity.



The Grand Design: New Answers to the Ultimate Questions of Life
STEPHEN HAWKING AND LEONARD MLODINOW
Bantam Press: 2010. 208 pp. $28, £18.99

Many think that the standard model and general relativity together come close to encapsulating the full set of rules that the Ionians hoped to find. These theories jointly describe everything from biochemistry to the large-scale structure of the Universe. Yet the mismatch between the deterministic nature of general relativity and the probabilistic quantum approach of particle physics points to a grander theory. Finding this theory — and unifying all the forces and particles — has been the holy grail of modern theoretical physics. Hawking and others held out hope that the ultimate theory's uniqueness would answer Einstein's question and reveal that no, the creator didn't have a choice.

Meanwhile, physicists of a philosophical bent concern themselves with another puzzle of the fundamental laws: their apparent specialness. Hawking and Mlodinow describe the "miracle" that the laws of physics allow for a hospitable Universe — one in which there is an excess of matter over antimatter, where galaxies host stars that last billions of years and harbour planets, and in which carbon-based organisms evolved. Such a miracle would not have occurred if the constants of nature had been slightly different. This has led some (myself not included) to promote the anthropic approach to the Universe: the laws of physics are what they are because if they were not, life would not have evolved to discover them. In a theory of everything, the fact of our existence should fall right out.

In searching for the holy grail, Hawking and others pinned their hopes first on supergravity and then on string theory. Both are now seen as different regimes of a grander mathematical framework called M-theory, where M is yet to be determined — is it master, miracle or mirage? M-theory unifies gravity with the other fundamental forces (weak and strong nuclear and electromagnetism), predicts seven additional dimensions of space and suggests that space and time might be emergent phenomena rather than fundamental. It is exciting and important, but much of it remains to be explored.

Besides the absence of any compelling experimental evidence for M-theory, there is another difficulty — its predictions are far from unique. There are 10^{500} different ways to curl up the extra seven dimensions and hide them, and how they curl up determines the fundamental constants and what we four-dimensional creatures see as the laws of physics. So even if M-theory is the only theory of everything available, there remain 10^{500} possibilities for the laws of physics we observe.

As Hawking and Mlodinow explain, inflationary cosmology turns this embarrassment into a virtue, partially answers Einstein's question and eliminates the need for a miracle. Cosmic inflation is the process by which a small part of the very young Universe blows up into a vast, geometrically flat and almost-smooth patch large enough to encompass all we can see and more, thereby accounting for the Universe around us today. Inflationary theory is on firmer ground than M-theory — it makes a number of predictions that have been verified. Yet because of quantum mechanics, inflation is not a one-time event but occurs continuously. Enormous bubbles of space-time are constantly being spawned, each one causally disconnected from the others and harbouring its own laws of physics.

Thus, say Hawking and Mlodinow, there is no miracle — inflation plus M-theory equals multiverse. Our special Universe is a selection effect: all possibilities have been tried and we find ourselves in the ▶

only kind of inflationary patch that can support our existence. The grand design is unnecessary. One is reminded of Winston Churchill damning the United States with faint praise — they get it right after they have exhausted all the alternatives.

The multiverse is possibly the most important idea of our time, and may even be right, but it gives me a headache. Is it science if we cannot test it? The different patches are incommunicado, so we will never be able to observe them. The multiverse displaces rather than answers the question about choice and who chooses, and does not explain why there is something rather than nothing. Hawking and Mlodinow argue that negative gravitational potential energies allow something to arise from nothing — but that still begs the question of why there is space, time and M-theory at all.

Hawking has not ruled out the existence of God, or even the odd possibility that our creator is a physics student in an advanced civilization carrying out a routine lab experiment. He has strengthened Laplace's argument that, although some assembly process is required, no creator is necessary. It is well known that Hawking is no fan of religion, but it was the media who took "no necessity for God" to mean "no God".

Hawking and Mlodinow's book is one of many works by big thinkers on the multiverse concept — including Leonard Susskind's *The Cosmic Landscape* (Little, Brown, 2005), Alex Vilenkin's *Many Worlds in One* (Hill and Wang, 2006) and Martin Rees's *Our Cosmic Habitat* (Princeton University Press, 2001). But when Hawking speaks, people listen. His clear, direct approach and his willingness to be provocative are enjoyable whether or not you agree with the details of his argument. With strong statements such as "philosophy is dead", he implies that it is now the duty of physicists to take up the big metaphysical questions.

Yet *The Grand Design* reminds me, as I tell my students, that science doesn't do 'why' — it does 'how'. Physicist Richard Feynman discussed the dangers of 'doing why' in his 1964 Messenger Lectures. He warned that should we achieve the Ionian goal of finding all the laws, then "the philosophers who are always on the outside making stupid remarks will be able to close in", trying to explain why those laws hold; and we won't be able "to push them away" by asking for testable predictions of those ideas. Time will tell if we are on to something big with the multiverse, or if we are becoming the philosophers that Feynman warned about. ■

Michael Turner *is director of the Kavli Institute for Cosmological Physics at the University of Chicago, Illinois, USA.*
e-mail: mturner@uchicago.edu
Competing financial interests declared; see http://dx.doi.org/10.1038/467657a

ANIMAL BEHAVIOUR

The wisdom of the bees

Swarms teach us that leaders should create conditions for collective decisions, learns **John Whitfield**.

You can never tell when apparently blue-sky science will be useful, as biologist Thomas Seeley's career shows. His knowledge of honeybees, for example, helped to defuse a cold-war confrontation in the 1980s, when he showed that yellow dots on Thai jungle foliage were not residues of Soviet chemical weapons but bee shit. And he has run his own department by the rules that swarms use to select a new home. *Honeybee Democracy* describes Seeley's quest to understand collective decision-making in social insects and humans.

Bee swarming is impressive and mysterious. Early in summer, a queen honeybee flies from her hive with a retinue of about 10,000 workers, leaving the home of her birth to be inherited by a daughter. The swarm might bivouac on a handy surface for several days before invading a new nest site in a tree hollow or building cavity. The collective must quickly decide where to settle, because it is risky to hang around in the open as food reserves dwindle. And it is important to pick the right spot — a colony that chooses poorly is unlikely to survive the winter.

Bees communicate through dancing. In the 1940s, German biologist Karl von Frisch decoded the waggle that worker bees perform to recruit foragers to food sources — the dance shows the direction, distance and quality of the food. His student, Martin Lindauer, noticed that during swarming some dancing honeybees were not covered in pollen, as were returning foragers, but in brick dust. He suspected that they had returned from potential nest sites, and were advertising them to their swarm-mates. By reading that dance, he worked out the site's probable location, and confirmed his hunch by following the swarm through the streets of Munich to its new home.

Honeybee Democracy
THOMAS D. SEELEY
Princeton University Press: 2010. 280 pp. $29.95

Seeley picked up the baton in the 1970s. *Honeybee Democracy* describes how, in a series of ingenious experiments, he deduced what kind of site bees prefer — a cavity of about 40 litres with a small entrance that faces south — and how a swarm homes in on the best of many possible nest sites. His story's heroines are the scout bees, a few hundred workers who trigger the swarm's departure, seek out nest sites, debate their merits, come to a decision, rouse the swarm and guide it to the new home.

A scout converts knowledge of a particular nest site into a waggle dance. The better the site, the longer and harder she dances. If another scout bumps into a dancing bee, she goes off to inspect the site. If she likes it, she too will dance. But any bee only


T. SEELEY

Tracking individuals in a swarm reveals how they turn house-hunting into a democratic process.

advertises a site for a few hours, even if she has found a dream home.

This stops the swarm jumping to a premature conclusion — a vital delay, as the best site is rarely found first. Eventually, the dynamics of dancing cause about 20–30 scouts to arrive at a single, high-quality nest site. Once this quorum is reached, the scouts stop the debate and communicate their decision to the swarm with high-pitched piping sounds and by running amid the other bees buzzing their wings, a preflight routine. The swarm then warms up and moves off, the scouts pointing the way.

This form of decision-making is extremely robust. Each bee's job is simple. Even if one makes a mistake, the rules that transform individual deeds into collective behaviour set the swarm back on course. Other systems have independently evolved the same tricks. A neuron, for example, carries little information. But by using similar rules to bees, cells combine to enable our brains to do clever things, such as tracking a moving object.

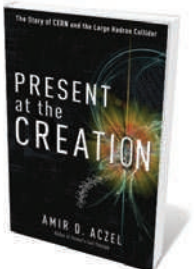
In the final chapter, Seeley lists his bee-derived rules for good human decision-making, and describes how he applied them as head of Cornell University's neurobiology and behaviour department in Ithaca, New York. He points out that groups make the best decisions when leaders interfere as little as possible. Individuals are then free to explore and debate options, and are most likely to arrive at the best decision. The wise leader, he advises, manages the process of decision-making and lets the product take care of itself.

In his own community, Seeley ensured that all possibilities were considered and that everyone had his or her say. He then stepped back to let the group make up its own mind by secret ballot. Such a process (discovered by several human societies independently) ought to work well in situations where a group with a common interest chooses between many options, from friends choosing a holiday destination to a government poised to invest billions in a defence system.

However, this rule about leaders facilitating decisions rather than making them is also the one humans find hardest to apply. Why struggle to the top if you can't push your own agenda? Or why pick leaders if they don't make their presence felt? In a crowd-sourcing exercise this year by the new UK government, for example, the public was asked to propose policy ideas and money-saving tips. Thousands of suggestions came in. But people were not asked to choose between the proposals. The decisions remained with those at the top. Humans prize their power and expertise — and that, Seeley's splendid book suggests, may be a cause of many of our problems. ■

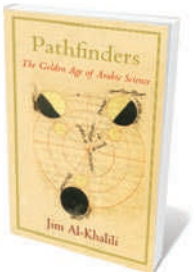
John Whitfield *is a science writer based in London. His book about reputation will be published in 2011.*
e-mail: j.a.whitfield@gmail.com

Books in brief

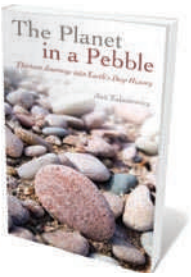


Present at the Creation: The Story of CERN and the Large Hadron Collider
Amir D. Aczel CROWN *288 pp. $25.99 (2010)*
Mathematician and author Amir Aczel describes the origins and science of the Large Hadron Collider at CERN, Europe's particle-physics lab near Geneva. Conveying his excitement at visiting the game-changing machine, Aczel's odyssey includes the voices of key scientists. After explaining the history of the standard model of particle physics, he looks ahead to string theory, the identity of dark matter and tests of the Higgs mechanism for conferring mass. But the real gems, he believes, will be beyond our imagination.

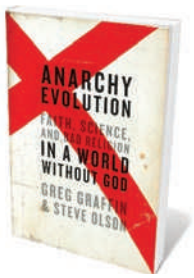
Good Faith Collaboration: The Culture of Wikipedia
Joseph Michael Reagle Jr MIT PRESS *256 pp. £20.95 (2010)*
Joseph Reagle, a computer-science historian, looks at the collaborative culture behind online encyclopaedia Wikipedia. He charts the technology-driven attempts in the 1930s to collect the world's knowledge and bypass elite publishers, such as Paul Otlet's information indexing system and H. G. Wells's proposal for a World Brain stored on microfilm. Wikipedia, Reagle argues, comes close to the goal of a universal encyclopaedia owing to the openness of its users in assessing knowledge. But there are downsides to inclusivity — for example, censorship, lawsuits and bureaucratization.



Pathfinders: The Golden Age of Arabic Science
Jim Al-Khalili ALLEN LANE *336 pp. £25 (2010)*
Physicist, author and broadcaster Jim Al-Khalili celebrates the forgotten pioneers of early Arabic science. His focus is the House of Wisdom, a great centre of learning established in the ninth century by the caliph of Baghdad, Abu Ja'far Abdullah al-Ma'mun. Among its wise alumni are Syrian astronomer Ibn al-Shatir, whose work inspired Nicolaus Copernicus's heliocentric model of the Solar System; Andalucian physician Ibn al-Nafees, who described blood circulation 400 years before William Harvey; and zoologist al-Jahith, who proposed natural selection 1,000 years before Charles Darwin.



The Planet in a Pebble: A Journey into Earth's Deep History
Jan Zalasiewicz OXFORD UNIVERSITY PRESS *256 pp. £16.99 (2010)*
Every pebble holds the story of Earth, shows geologist Jan Zalasiewicz. He extracts from a humble stone evidence of the violent formation of the Solar System, in which our embryonic planet was dusted with the detritus of supernova explosions and the elemental litter of the Big Bang. Trapped, too, is the tale of Earth's evolution — the lives and deaths of disappeared plants and animals, volcanic eruptions and long-vanished oceans. Zalasiewicz explains how oil and minerals form and how geologists use clever chemistry to sniff out the resources on which we depend.



Anarchy Evolution: Faith, Science and Bad Religion in a World Without God
Greg Graffin and Steve Olson IT BOOKS *304 pp. $22.99 (2010)*
Greg Graffin's memoir, co-authored with science writer Steve Olson, offers an unusual perspective on evolution. As an evolutionary biologist and lead singer of punk band Bad Religion, Graffin argues that research and punk rock have much in common: both require an open mind and look to evidence and rationality. Bucking authority and the religious views of his family, Graffin explains how he has developed a personal philosophy that celebrates the power of nature.

R. MÄRZ

Mark Dion's wall of curiosities bears witness to Berlin's place in science history.

HISTORY

The light and shade of German science

From physiology to physics, a stirring exhibition reflects 300 years of science in Berlin, discovers **Alison Abbott**.

The first image to confront visitors to the *Weltwissen* exhibition in Berlin's Martin Gropius Bau is formed of shadows. Silhouettes of 250 historical objects in 96 giant shelving cubes are projected using bright lighting onto a vast canvas that backs the two-sided display, filling the atrium. The items — ranging from statues of Greek philosophers to skeletons and an iron lung — were selected from local museums by New York artist Mark Dion as 'witnesses' of Berlin's scientific past.

The display is a neat metaphor for the light and shadow of Berlin's 300-year scientific history, reflected in this ambitious show. The city was home to some of Europe's most important science before the rise of Nazism in the 1930s, the physical destruction of the city in the Second World War and the 1949 rise of the Berlin Wall. *Weltwissen* opened on 24 September, 10 days before the 20th anniversary of the reunification of Germany.

The exhibition celebrates scientific rather than political anniversaries — 300 years since the founding of the Berlin Academy of Sciences and the Charité, now Berlin's university hospital; and 200 years since the Humboldt University was established. The Kaiser Wilhelm Society, now the Max Planck Society, runs 80 research institutes throughout Germany and was founded in 1911; running until early next year, the show also covers that centenary.

Weltwissen (World Knowledge): 300 Years of Science in Berlin
Martin Gropius Bau, Berlin. Until 9 January 2011.

The exhibition's powerful fascination lies largely in the placing of Berlin science in the political and social contexts of its times, good and evil. It also lies in the curators' reliance on real objects, rather than multimedia, to tell the stories. To stand close to the 1880 full-body cast of a naked tribesman — made while he was alive for the then-fashionable science of anthropometrics — and to see how he squeezed his eyes shut against the wet plaster, is to experience a raw emotional force that would be hard to create by digital means.

Dion spent two months in the city sifting through tens of millions of historic objects for his installation; the curators spent even longer, and to good effect. The exhibition's rooms are organized into historical eras or are dedicated to eternal themes in scientific culture, such as quarrels over data interpretation or experimental methodology. The first few rooms depict the city's mad rush in the eighteenth century to catch up with established centres of science such as Paris and London, when its first observatory and anatomical theatre were built. Subsequent rooms focus on the next two centuries, when the city grew to be a major force in European research.

Towering scientific figures are introduced: naturalist and explorer Alexander von Humboldt (1769–1859); physiologists Hermann von Helmholtz (1821–94) and Emil du Bois-Reymond (1818–96), who did much to link animal physiology with the laws of chemistry and physics; Werner von Siemens (1816–92), who founded the famous electrical and telecommunications company; microbiologist Robert Koch (1843–1910), who identified the bacterium that causes tuberculosis; and chemist Fritz Haber (1868–1934), who won the 1918 Nobel Prize in Chemistry for his synthesis of ammonia.

Ethics is a running theme, and the dark side of science is prominently displayed. The ambitious Koch, for example, dodged local restrictions on human experimentation by going to East Africa in 1906 to test potential medicines for sleeping sickness. During the First World War, Haber developed poison gases for use in the trenches. The unspeakable Nazi abuses of science and medicine are laid out soberly. Personal letters and diaries of Jewish scientists who fled or were expelled from Nazi Berlin — such as Albert Einstein and Haber — are deeply moving. The exhibition tells us that, after the war, in East Berlin the communist state highlighted these abuses as evidence of the necessity of its regime, while West Berlin closed its collective mind to the issue until the late 1980s.

Notable films include historic footage of the 1933 Nazi book burning, and a new film made for an installation on experimental methods in Alzheimer's disease, a focus of research activity in Berlin. It shows a classic test of rodent memory: a mouse is dropped into a water maze and swims to find the submerged platform. Simple animated graphics show how a memory-impaired mouse must search much longer for the out-of-sight platform. But it is the mouse's perspective of its forced activity that will captivate biologists.

Of note also are the recorded reflections of 16 scientists involved in decisions about which East German institutes and individuals were worth retaining in the science system of reunified Germany. It was a cruel time for many; others, such as chemist Joachim Sauer, husband of Chancellor Angela Merkel, survived the cuts. Now a Humboldt University professor, he remembers the unfair handling of some older colleagues.

Berlin is still struggling with the expense of reunification. But the quality of the exhibition — the €5.5-million (US$7.4-million) cost of which was met by the Berlin lottery — demonstrates the current intellectual wealth of the city. *Weltwissen* is gorgeous to look at, yet visitors will find their preconceptions challenged, and will leave better educated than when they entered, having faced the shadows. ■

Alison Abbott *is* Nature*'s senior European correspondent.*

CORRESPONDENCE

Replies to 'The evolution of eusociality' by Nowak et al. (Nature **466,** *1057–1062; 2010*). SEE COMMENT P.653

Better living through physics

Nowak and colleagues' explanation of the evolution of altruism in terms of individual-level selection might be reconciled with the views of their kin-selection opponents by striking an analogy with statistical mechanical and thermodynamic treatments in physics.

Statistical mechanics provides the microscopic basis for the macroscopic variables in thermodynamics, which is an equilibrium theory treating aggregate variables. As with thermodynamics, traditional multilevel selection theory is based on equilibrium solutions operating on nominal, aggregate variables. In the Hamilton kin-selection framework, variables correspond to the terms benefit, cost and relatedness. But because that treatment is not fundamentally mechanistic, it is often unclear what the units of these variables are, and how best to measure them.

Population genetics presents an evolutionary analogue of statistical mechanics that complements Hamilton's evolutionary thermodynamics. Hamilton's rule — which expresses relatedness between the helped and the helper in terms of cost and benefit to the fitness of both — and its related inequalities all express dependencies among macroscopic variables of state in structured populations.

The greater complexity of biological systems over physical ones, and their strong interdependency, make for a zoo of biological macroscopic laws with many multilevel selection principles, each with its adherents and disciples.

The great promise of evolutionary statistical mechanics is that it should allow us to enumerate the full space of possible fundamental evolutionary inequalities and the mechanistic conditions under which they apply, thence identifying those with the greatest empirical generality.

David C. Krakauer, Jessica C. Flack *Santa Fe Institute, USA, krakauer@santafe.edu*

Inclusive fitness is just bookkeeping

Using a very general modelling approach, Nowak and colleagues make two important and valid points. First, when standard natural selection theory and inclusive fitness theory can both be applied, they must yield the same answer. Whether the realm of applicability of inclusive fitness models is as restricted as claimed remains to be seen, but it is in any case contained in the realm of standard models. Second, the paper makes it clear what inclusive fitness theory really is: an accounting method, not a biological mechanism.

Champions of inclusive fitness often refer to the underlying mechanism as kin selection, but this just restates the fact that the benefit a particular gene generates at a cost to its carrier must preferentially go to the gene's other carriers (kin). The real biological problem is to understand mechanisms that lead to such assortment between helper and help. For eusocial insects, Nowak *et al.* convincingly argue that the basic mechanism of assortment is the formation of groups owing to ecological pressures, such as the need for nest defence.

Despite the indignant response of the inclusive-fitness crowd, there can be no doubt about the fundamental tenet that, with or without the concept of inclusive fitness, in principle we have access to exactly the same amount of evolutionary knowledge. Personal modelling preferences may vary, but there is nothing magic about bookkeeping techniques.

Michael Doebeli *University of British Columbia, Canada, doebeli@zoology.ubc.ca*

Ground truth is the test that counts

A strength of kin-selection theory for explaining the evolution of altruism is that it generates multiple hypotheses that can fail in resulting tests, forcing principles to be re-examined. Lessons can be learned from one such failure discussed by Nowak *et al.*: the haplodiploid hypothesis.

With haplodiploidy, fertilized eggs become females and unfertilized ones become males. Thus, females can have higher fitness if they raise their full sisters instead of their own offspring.

This prediction was not wrong in the sense that it arose from flawed mathematics. Indeed, practitioners of the natural-selection approach championed by Nowak *et al.* could have made the exact same prediction. It fails experimentally because a basic assumption is not met in nature: females rarely get the opportunity preferentially to raise full sisters.

Theoretical papers that developed more viable alternative scenarios also used inclusive fitness methods (for example, see D. C. Queller *Proc. Natl Acad. Sci. USA* **86,** 3224–3226; 1989). This robust predict–fail–reevaluate triumvirate is why kin selection has been the main informative model of choice in the past and why it should continue to help us understand the evolution of cooperation and conflict.

Peter Nonacs *University of California, Los Angeles, USA, pnonacs@biology.ucla.edu*

Call for a return to rigour in models

Why are numerous reactions to the Nowak *et al.* paper so ferocious? And how is it possible that theorists even seem to disagree about mathematics? An important key to the heatedness of the debate is that many theory papers on the evolution of cooperation use the Price equation. This is regularly treated as if its generality makes it the $E = mc^2$ of population genetics. Those who use it, however, tend to forget that it is only general because it contains no modelling assumptions. And without these, one cannot derive theoretical predictions.

The Price equation inspires theoreticians to confuse identity with causality, and probability theory with statistics (M. van Veelen *J. Theor. Biol.* **237,** 412–426; 2005). It is this theoretical blur that obfuscates whether or not claims are theoretical results that follow from actual model assumptions.

References to 'results' derived with the Price equation are a recurring element in discussions of inclusive fitness. As long as the Price equation is thought useful for finding theoretical predictions, these 'results' will collide with results from actual models, which is a recipe for hot-headed debates.

Nowak and colleagues' paper is exciting because it goes back to basics: it builds proper models. Rather than saying the paper is wrong, it would be more fruitful if critics also went back to basics: state model assumptions, derive predictions, test empirically. Such a return to rigour would help the field advance to the next level.

Matthijs van Veelen, Julián García, Maurice W. Sabelis, Martijn Egas *University of Amsterdam, the Netherlands, c.m.vanveelen@uva.nl*

SURFACE SCIENCE

Seeing the spin through

Interfaces can have quite different properties from those of their constituent materials. But it's surprising that the adsorption of a single organic molecule onto a magnetic surface can drastically modify that surface's magnetism.

STEFANO SANVITO

Traditionally, magnetism and organic chemistry have not shared the same laboratory. In recent years, however, the two disciplines have become closer to each other as interest has blossomed in studying organic molecules as a transport medium for electronic spins. This has raised the expectation that memory and logic devices one day will be built from cheap plastic materials. So far, experiments on spin transport across organic molecules have been based on large devices, in which the microscopic details of the interfaces between the organic and the magnetic materials from which the spins are injected are averaged out. Writing in *Physical Review Letters*, Atodiresei *et al.*[1] and Brede *et al.*[2] now investigate how spins are transferred from a magnetic surface to a single organic molecule.

The results are surprising. The authors[1,2] demonstrate that it takes only a small, non-magnetic organic molecule to filter and reverse the spin orientation of the conducting electrons of an inorganic, magnetic material. The implication is that organic–inorganic interfaces can be tailored to specific applications in spin-transport electronics (spintronics).

The spin, as well as the electrical charge, is an intrinsic property of all electrons. If an electron is pictured as a tiny planet, its spin will represent the planet's direction of rotation, with spin 'up' denoting clockwise rotation and spin 'down' anticlockwise. However, in contrast to planets, all electrons 'rotate' at the same speed, and it is only their direction of rotation that can change. The ability to create, manipulate and detect electrical currents made by electrons with the same spin orientation is the goal of spintronics.

In a magnet, the electron spins have a preferential orientation, which is that of the material's magnetization. Thus, when an electron current flows from a magnetic to a non-magnetic material, the spins will maintain their preferential orientation for some distance, beyond which the number of spin-up and spin-down electrons will eventually become equal. In their studies, however, Atodiresei *et al.*[1] and Brede *et al.*[2] show that a single organic molecule is capable of filtering a particular spin direction. The authors find that, whereas the current flowing across a thin film of magnetic iron is composed mainly of spin-down electrons, the current that tunnels through an organic molecule (either metal-free phthalocyanine or cobalt phthalocyanine) adsorbed on the same iron film comprises mostly spin-up electrons.

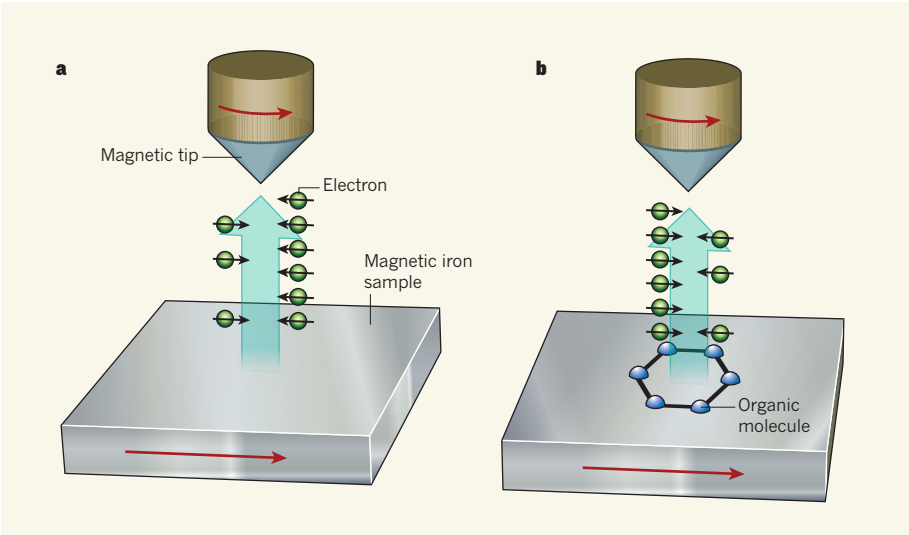


Figure 1 | Studying spin transport with SP-STM. In this technique, an atomically sharp magnetic tip is scanned across a magnetic sample, and electrons tunnel through the vacuum between the sample and the tip with a preferential spin orientation. The electron current intensity depends on the mutual orientation of the surface and tip magnetization (red arrows). **a**, In the case of an iron surface and a chromium-coated tungsten tip, the preferential spin orientation of the electrons leaving the surface is opposite to that of the surface magnetization. **b**, Atodiresei *et al.*[1] and Brede *et al.*[2] find that, when an organic molecule is adsorbed onto the iron surface, the electrons' spin orientation changes: most of the electrons have their spins oriented along the direction of the surface magnetization.

The fact that organic molecules can drastically modify the spin nature of an electrical current was predicted some time ago[3] and recently demonstrated experimentally[4]. However, this is the first time that the phenomenon has been observed at the atomic level. The authors[1,2] achieved this by using spin-polarized scanning tunnelling microscopy (SP-STM), a technique that combines sensitivity to spin orientation with atomic-scale image resolution (Fig. 1). In brief, STM involves scanning an atomically sharp conducting tip across a sample and measuring the electrical current that tunnels through the vacuum between the sample and the tip when a voltage difference is applied between the two. The intensity of the current as a function of the tip's position allows information about the shape and the electronic structure of the sample to be extracted. In the specialized SP-STM form of the technique, the tip is made of a magnetic material, meaning that the intensity of the current is also sensitive to the imbalance between the two spin orientations (the spin polarization) of the electrons moving from the sample to the tip. SP-STM therefore allows the spin polarization of an electrical current to be mapped.

Atodiresei *et al.*[1] and Brede *et al.*[2] find not only that the spin polarization of the current emerging from an organic molecule can be opposite to that emerging from the substrate in the absence of the molecule, but also that different regions of the molecule can sustain different current spin polarizations. The reason for both observations is rooted in the chemical bond formed between the molecule and the substrate, which is dominated by those orbitals in the molecule and in the magnetic surface that have the largest overlap with one another. In the case of iron, the orbitals involved in the bonding are shaped like a dumbbell with a ring around it (put more technically, they have d_{z^2}

symmetry); these same orbitals are responsible for the spin-down polarization of the conducting electrons in the substrate. In the molecule, the orbitals that drive the molecule–substrate interaction are dumbbell-shaped (p_z symmetry), just like those that contribute to conductivity in graphene. The formation of the bond creates new (hybrid) orbitals, which are an admixture of both the substrate's d_{z^2} orbitals and the molecule's p_z orbitals. These have energy levels that are different from those of the constituent orbitals, leading to the reversal of the spin orientation of the substrate's conducting electrons — they become mostly spin-up. The authors' experiments[1,2] thus indicate that engineering the chemical bond of organic–inorganic interfaces translates into engineering their ability to filter electrons of a single spin[5].

But the experiments are also remarkable for another reason. They firmly establish SP-STM, in combination with first-principles theoretical modelling, as a primary tool for studying spin phenomena in organic materials at the atomic scale. In general, this is a challenging task because of the peculiar nature of organic molecules: they lack a significant spin–orbit coupling; that is, the electrons' spins are weakly affected by their motion. Although this is good news for the prospect of using them to build organic spintronic devices, it makes the characterization of organic media complicated; the standard optical techniques used to characterize inorganic semiconductors rely on a significant spin–orbit interaction to work. As such, SP-STM joins the approaches of two-photon photoemission[6] and muon spin-rotation[7] as the techniques of choice for organic spintronics, and is a welcome addition to a field in which accurate metrology has been a pressing issue[8].

The experiments of Atodiresei *et al.*[1] and Brede *et al.*[2] are a step forwards in understanding the subtle interplay between chemical bonds, magnetism and spin transport. They also open up exciting avenues of research. For instance, the interaction between magnetic molecules and magnetic substrates, or between magnetic molecules themselves, could be exploited to lead to fully functional magnetic–magnetic interfaces at which magnetism can be finely and dynamically tuned (for instance, by manipulating the quantum-mechanical process of exchange coupling). Furthermore, SP-STM could be used to investigate at the atomic level the mechanism with which spins in molecular systems respond to external stimuli. This is a tool that has already helped enormously in the development of magnetism at the nanoscale, but that we physicists believe has much more to give. ■

Stefano Sanvito *is at the School of Physics and the Centre for Research on Adaptive Nanostructures and Nanodevices (CRANN), Trinity College Dublin, Dublin 2, Ireland.*
e-mail: sanvitos@tcd.ie

1. Atodiresei, N. *et al. Phys. Rev. Lett.* **105,** 066601 (2010).
2. Brede, J. *et al. Phys. Rev. Lett.* **105,** 047204 (2010).
3. Rocha, A. R. & Sanvito, S. *J. Appl. Phys.* **101,** 09B102 (2007).
4. Barraud, C. *et al. Nature Phys.* **6,** 615–620 (2010).
5. Sanvito, S. *Nature Phys.* **6,** 562–564 (2010).
6. Cinchetti, M. *et al. Nature Mater.* **8,** 115–119 (2009).
7. Drew, A. J. *et al. Nature Mater.* **8,** 109–114 (2009).
8. Szulczewski, G., Sanvito, S. & Coey, J. M. D. *Nature Mater.* **8,** 693–695 (2009).

MEDICAL MICROBIOLOGY

A toxin contest

The bacterium *Clostridium difficile* can cause life-threatening human disease. The question is which of the organism's two toxins is the more crucial to its pathogenicity. The answer is one or the other, or both. SEE LETTER P.711

JIMMY D. BALLARD

In this issue, Kuehne and colleagues[1] show that the pathogenic bacterium *Clostridium difficile* can cause disease using either of its two main toxins — toxin A and toxin B. These findings may come as a surprise, because a similar study[2] last year reported that only toxin B is required for the pathogen's virulence. This new paper will undoubtedly reignite debate over the matter.

Clostridium difficile affects hospital patients receiving antibiotic treatment, and can lead to diarrhoea, intestinal inflammation and severe systemic complications[3,4]. The disease is difficult to manage with traditional antibiotics, because treatment with these drugs can often trigger the illness[5]. Most concerning has been the emergence of hyper-virulent, drug-resistant strains of the bacterium, which have resulted in an increased death rate among patients. It is therefore crucial to better understand the pathogen's main mechanisms of virulence.

Most strains of *C. difficile* produce both toxin A and toxin B. The two toxins show many similarities and are probably products of a gene-duplication event during the evolution of the bacterium. Both toxins affect the same target on entry into host cells[6,7], and both are lethal when injected into non-human primates[8]. They differ, however, in their target cell type, with toxin A damaging cells of the intestinal epithelium and toxin B targeting a broad range of cells[3].

Previous findings have mainly pointed to toxin A as the major culprit in causing damage during the disease, with toxin B playing a minor part. One study[9], for example, found that administering toxin A to experimental animals recapitulates many signs of *C. difficile* disease, and another[10] showed that patients with strong immune responses to this toxin were less likely to relapse with *C. difficile* disease. But it was difficult to reconcile these observations with the routine occurrence of bacterial strains lacking toxin A but expressing toxin B (A^-B^+ strains) in patients with severe *C. difficile* infection[3]. The fact that *C. difficile* can naturally cause disease without toxin A raised several questions. Could the bacterium also cause disease without toxin B? And are the toxins' functions redundant, additive or synergistic in *C. difficile* infection?

Assessing the role of toxin A in the absence of toxin B has been difficult, because naturally occurring A^+B^- strains have not been identified. The gold standard for addressing this question therefore remained an experimental system involving isogenic strains of the bacterium — strains that are genetically identical except for the disrupted target gene. Such genetic manipulation, which is routinely used for studying many bacteria, had been technically challenging in *C. difficile* until a 2006 paper[11] reported the desired targeted gene disruption. The authors of that study later used[2] this approach to produce A^+B^- and A^-B^+ isogenic *C. difficile* strains. When they tested these strains in a hamster model of *C. difficile* infection, they found that only toxin B was required for the bacterium to cause disease.

Kuehne and colleagues' findings[1] (page 711) seem to contradict those earlier results[2]. They indicate that, in hamsters, although the A^+B^- strains are less virulent than the A^-B^+ strains, they can cause illness. The investigators also generated a strain of *C. difficile* that lacks both toxins, and, as predicted, this strain is not virulent.

As is often the case with conflicting findings, there are differences between the two experimental approaches used[1,2]: the teams' isogenic strains were generated using different systems, and the sites of genetic disruption were not the same. Moreover, whereas the earlier paper examined two independent mutants, Kuehne *et al.* tested one mutant, but used a system that generates more stable mutants unable to revert to their original toxin-positive state. Furthermore, Kuehne and co-workers used physical appearance and physiological indicators to develop a scoring system for predicting death, rather than actual death of the animal as the endpoint. Whether the more subjective method for scoring death had any

impact on their data is difficult to know.

Both groups used strain 630 of *C. difficile*, a well-characterized clinical isolate. Kuehne *et al.* propose, however, that differences in the number of times this strain has been subcultured (passaged) in the two labs could account for the variation in the results. So it's conceivable that other factors — apart from the two toxins — changed during laboratory handling of the organism by the two teams, and that these factors are relevant to virulence.

It should also be noted that neither study could deliver the toxin-encoding genes back into the mutant strains of *C. difficile* to restore the virulent phenotypes. Without this, a variety of genetic explanations for the differences in the two outcomes remain. Indeed, analysis of more than one mutant is a notable strength of the earlier study[2], because it reduces concerns that other mutations could have been affected during manipulation of the organism. Ultimately, it will require further scrutiny to determine whether the overall experimental variations are significant enough to account for the disagreement in the findings.

Investigations should also be carried out to determine which of the two conclusions[1,2] holds up in other strains of *C. difficile*, in particular the hyper-virulent strains. And it would be useful if the two groups exchanged reagents in an effort to explain the differences between their data. At this point, however, perhaps it is especially necessary to emphasize the common finding of the two studies: toxin B is a crucial virulence factor of *C. difficile* and its presence alone is enough to cause disease. On the basis of both sets of results, therefore, targeting only toxin A might be a poor treatment strategy. *Clostridium difficile* is a complex bacterium, and, despite their different conclusions, both teams' contribution will help to advance the understanding of this life-threatening human pathogen. ■

Jimmy D. Ballard *is in the Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104, USA.*
e-mail: jimmy-ballard@ouhsc.edu

1. Kuehne, S. A. *et al. Nature* **467,** 711–713 (2010).
2. Lyras, D. *et al. Nature* **458,** 1176–1179 (2009).
3. Voth, D. E. & Ballard, J. D. *Clin. Microbiol. Rev.* **18,** 247–263 (2005).
4. Rupnik, M., Wilcox, M. H. & Gerding, D. N. *Nature Rev. Microbiol.* **7,** 526–536 (2009).
5. Gerding, D. N., Muto, C. A. & Owens, R. C. Jr *Clin. Infect. Dis.* **46,** S32–S42 (2008).
6. Just, I. *et al. Nature* **375,** 500–503 (1995).
7. Just, I. *et al. J. Biol. Chem.* **270,** 13932–13936 (1995).
8. Arnon, S. S. *et al. J. Pediatr.* **104,** 34–40 (1984).
9. Lyerly, D. M., Phelps, C. J., Toth, J. & Wilkins, T. D. *Infect. Immun.* **54,** 70–76 (1986).
10. Kyne, L., Warny, M., Qamar, A. & Kelly, C. P. *Lancet* **357,** 189–193 (2001).
11. O'Connor, J. R. *et al. Mol. Microbiol.* **61,** 1335–1351 (2006).

NANOFLUIDICS

Tiny electrostatic traps

Methods for trapping tiny particles are increasingly needed, especially for biological assays, but they often involve complicated apparatus. An approach has been discovered that could simplify matters considerably. **SEE LETTER P.692**

JAN C. T. EIJKEL & ALBERT VAN DEN BERG

We've all seen images of astronauts in their spaceship cabins at zero gravity. Moving freely through the interior space, nothing inhibits them but the insubstantial friction of the air. Now imagine that they are repelled by the cabin walls. Instead of freely floating around, the hapless space travellers would be pushed to the place that is farthest away from all the walls. There they would be stuck, captured in a repulsive trap. Reporting in this issue (page 692), Krishnan *et al.*[1] describe the realization of this scenario at the nanometre scale: an electrostatic trap for nanoparticles.

The authors have used micromachining techniques to create structures that contain interconnected compartments, each measuring a few hundreds of nanometres in height (Fig. 1). This is advantageous for manipulating nanoparticles, because the scaling of natural laws makes gravity totally irrelevant at these dimensions. Nanoparticles suspended in water within the compartments therefore float around like astronauts at zero gravity, never sinking to the bottom.

The second ingredient used by Krishnan *et al.* in their trap is the strong repulsion between particles and the walls of the compartments, which originates from the negative charge that both carry. The glass walls of the device naturally acquire negative charge on exposure to water. This means that, when water containing negatively charged particles is flushed through the system, the particles are pushed away from the walls and become trapped in those locations farthest removed from all of the walls — that is, in the compartment centres. Although Brownian motion bounces particles around in the traps, the repulsive forces are stronger than the Brownian forces, confining the particles for hours on end.

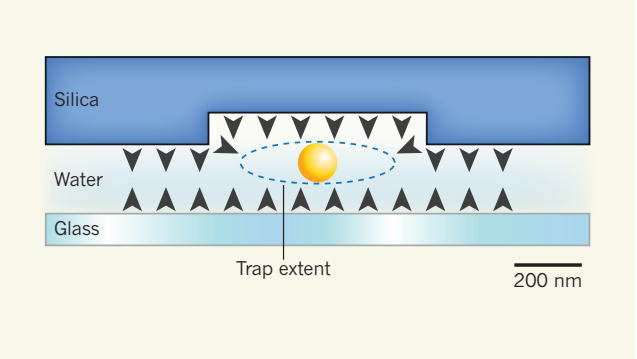


Figure 1 | Caught in a trap. Krishnan *et al.*[1] report a system in which nanometre-scale particles can be trapped. The traps consist of compartments 300 nanometres high, etched into the gap between a silica surface and a glass surface. The surfaces of the compartments become negatively charged on contact with water. Negatively charged particles (20–100 nm in diameter) suspended in the water become trapped in the compartments by particle–wall repulsion. Arrows indicate repulsive forces, and the broken line indicates the extent of the electrostatic trap.

An advantage of these electrostatic traps is that they are indifferent to the particle material. The authors show that they can trap particles of widely different composition — such as gold nanoparticles, polystyrene beads and lipid vesicles — and of sizes ranging from 20 to 100 nanometres. The only prerequisite is that there should be sufficient surface charge on the particle to be trapped, which isn't a problem because almost all materials carry some surface charge. Another attractive feature is that no external equipment needs to be used. The equipment needed for other methods of particle trapping can be very expensive and complicated, as in the case of optical trapping (in which microscopic objects are held in place by a highly focused laser beam).

A final alluring aspect of Krishnan and colleagues' system is the extremely easy way in which the shape of the traps can be altered and the number of traps can be increased. Currently available micromachining techniques would allow millions of traps to be constructed on only a few square centimetres of a substrate, at a density comparable to that of the DNA arrays used for biological assays, or even higher. We could do many things with such tiny arrays of traps. Particle-based assays are widely used in clinical chemistry and biology, for example. The use of trap arrays would enable us to trap individual particles and observe

them over time, thus allowing single binding events to be distinguished, or processes taking place in single vesicles to be observed.

Of course, many other particle-trapping methods are available[2], such as the popular optical trapping and the versatile dielectrophoretic trapping (in which particles are captured and manipulated by electric fields). How does electrostatic trapping compare? One large advantage over other methods is that electrostatic trapping occurs automatically when a particle suspension is flushed through the trap. As noted earlier, the absence of any specialist equipment makes life very easy for the operator. This comes with a considerable drawback, however: the particles are ensnared at fixed positions that cannot be changed at will, as can be done with optical trapping.

But perhaps the biggest limitation of the electrostatic trapping mechanism at present is the need for extremely low salt concentrations in the particle-carrying liquid. Given that biological fluids have high salt concentrations, if left unsolved this problem would undoubtedly put a damper on many biological applications. Indeed, a key objective of electrostatic trapping would be to capture single protein molecules at physiological salt concentrations.

To address this issue, Krishnan *et al.*[1] derived equations to describe the forces in their traps, and used these to show that the detrimental effect of high salt concentrations can be countered by reducing the height of the traps. They therefore calculated that, in principle, it should be possible to capture strongly charged proteins if traps 10 nm high could be made (for comparison, the traps used in their experiments were 300 nm high). It is worth noting, however, that the proteins considered by the authors have roughly the same dimensions as the traps required to catch them, which might allow short-range, attractive van der Waals forces between the protein and the trap's walls to negate the desired electrostatic repulsion. If so, then perhaps this problem could be prevented by attaching a polymer layer to the trap's surfaces — an approach that has been used successfully to overcome van der Waals forces to allow the levitation of colloidal microparticles above a surface on which these (larger) particles would otherwise settle by gravity[3].

Many further studies involving electrostatic trapping spring to mind. To deepen our understanding of the mechanism, experiments at different pH values, and using positively charged surfaces and particles, will be necessary. Electrodes could also be built into the trap walls to actively modify the surface charge, so that trapped particles could be moved, overcoming one of the present limitations[4,5]. Electrostatic trapping certainly opens up an intriguing avenue of investigation in the manipulation of particle systems, and may well prove to have a great impact on biological research. ■

Jan C. T. Eijkel *and* **Albert van den Berg** *are at the MESA+ Institute for Nanotechnology, University of Twente, PO Box 217, 7500 AE Enschede, the Netherlands.*
e-mail: j.c.t.eijkel@utwente.nl

1. Krishnan, M., Mojarad, N., Kukura, P. & Sandoghdar, V. *Nature* **467,** 692–695 (2010).
2. Castillo, J., Dimaki, M. & Svendsen, W. E. *Integr. Biol.* **1,** 30–42 (2009).
3. Everett, W. N., Wu, H.-J., Anekal, S. G., Sue, H.-J. & Bevan, M. A. *Biophys. J.* **92,** 1005–1013 (2007).
4. Schasfoort, R. B. M., Schlautmann, S., Hendrikse, J. & van den Berg, A. *Science* **286,** 942–945 (1999).
5. Karnik, R. *et al. Nano Lett.* **5,** 943–948 (2005).

DNA REPAIR

A protein giant in its entirety

Purification of the human tumour-suppressor protein BRCA2, which is crucial for DNA repair, has been a formidable challenge owing to its large size. That mission is now accomplished, providing biochemical insight. SEE ARTICLE P.678

LEE ZOU

A crucial mechanism by which breaks in double-stranded DNA are repaired is homologous recombination. BRCA2, a large protein of more than 400 kilodaltons in size, is a key player in this process. Three papers[1–3], including one by Jensen *et al.*[1] on page 678 of this issue, report the purification and biochemical characterization of full-length human BRCA2, providing the most comprehensive picture yet of how this protein giant functions in DNA repair.

BRCA2 is a suppressor protein of ovarian and breast tumours, and the gene encoding it is frequently mutated in patients with familial breast cancer and those with Fanconi anaemia, a disease that increases susceptibility to cancer. In cells, the absence of functional BRCA2 severely compromises the repair of DNA double-strand breaks (DSBs) by homologous recombination[4]. Biochemical studies[4] using fragments of human BRCA2, or BRCA2-like proteins from a fungus and from worms, have suggested that BRCA2 recruits another protein, RAD51, to the single-stranded DNA (ssDNA) that is generated at sites of DSBs. The resulting RAD51–ssDNA filaments mediate the search for homologous DNA sequences and 'strand invasion' — two crucial steps in homologous recombination. Owing to the difficulty in purifying full-length human BRCA2, however, the biochemical properties of this large protein have not previously been studied directly.

The three teams[1–3] have successfully expressed full-length recombinant BRCA2 using different strategies. With purified BRCA2 in hand, the researchers then sought to answer three questions: how does it interact with DNA; how does it interact with RAD51; and how does it regulate the formation of RAD51–ssDNA filaments?

Previous studies using BRCA2 fragments indicated that the protein might bind either to ssDNA or to junctions of ssDNA and double-stranded DNA (dsDNA)[5]. Full-length BRCA2 also seems to prefer ssDNA to dsDNA[1,2]. Moreover, electron microscopy data[2] show that BRCA2 specifically recognizes dsDNA with ssDNA tails, but not the blunt ends of dsDNA. Intriguingly, on binding to DNA ends, BRCA2 forms a rod-shaped complex, the dimension and mass of which are consistent with BRCA2 dimers[2].

Unlike its related proteins in fungus and worms, human BRCA2 contains eight BRC repeats, which can bind to RAD51. But how these repeats are used in full-length BRCA2 was not clear. It is now estimated[1,3] that each full-length BRCA2 molecule can bind to up to six RAD51 molecules. The ability of BRCA2 to simultaneously bind to several RAD51 proteins may facilitate RAD51 binding to ssDNA and/or formation of the helical RAD51–ssDNA filaments[6]. Although BRCA2 can bind to both ssDNA and RAD51, the researchers[2] did not detect a ternary complex of BRCA2–RAD51–ssDNA. This suggests that BRCA2 delivers RAD51 onto ssDNA, but does not become a stable part of the RAD51–ssDNA filament.

So how exactly does BRCA2 promote the formation of RAD51–ssDNA filaments? The collective results of the three papers suggest that BRCA2 has at least four distinct roles in this process (Fig. 1, overleaf). It prevents RAD51 from binding to dsDNA, which would inhibit homologous recombination[1,2]. It stimulates the binding of RAD51 to ssDNA or to dsDNA with ssDNA tails[1,2]. It enables RAD51 to bind to ssDNA even in the presence of RPA — a high-affinity ssDNA-binding protein that at later steps of homologous recombination inhibits RAD51 binding to ssDNA[1–3]. Finally, it inhibits RAD51–ssDNA dissociation by preventing the hydrolysis of ATP molecules[1,3]. Through these four distinct

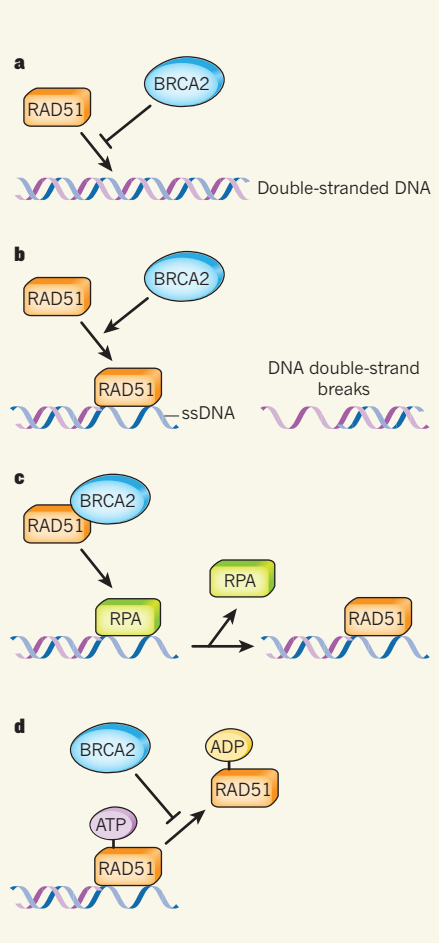


Figure 1 | The multifaceted role of BRCA2 in homologous recombination. BRCA2 mediates the formation of RAD51–single-stranded DNA (ssDNA) in four ways[1–3]. **a**, It inhibits the binding of RAD51 to double-stranded DNA. **b**, It stimulates the binding of RAD51 to single-stranded DNA tails at DNA double-strand breaks. **c**, It helps RAD51 to bind to ssDNA in the presence of the RPA protein. **d**, It inhibits RAD51 dissociation from ssDNA mediated by ATP hydrolysis to ADP.

functions, BRCA2 promotes specific and stable assembly of RAD51 on ssDNA, and thus RAD51-mediated strand exchange between homologous DNA sequences[1,2].

Jensen *et al.*[1] also reveal an intriguing divergence between the functions of the human BRCA2 and another protein, RAD52. Yeast lacks a BRCA2-like protein, and instead Rad52 promotes the formation of Rad51–ssDNA filaments. Moreover, unlike BRCA2, yeast Rad52 can anneal complementary ssDNA sequences into dsDNA, even in the presence of RPA. The human RAD52 can anneal ssDNA, but it cannot promote RAD51–ssDNA assembly. It seems, therefore, that during evolution BRCA2 has taken over the function of RAD52 in the formation of RAD51–ssDNA filaments.

Do we now have a complete picture of how BRCA2 functions in homologous recombination? The answer, clearly, is no. BRCA2 is regulated by its binding partners — as exemplified by stimulation of BRCA2-mediated binding of RAD51 to RPA-coated ssDNA by the BRCA2-associated protein DSS1 (ref. 3). Moreover, BRCA2 interacts with several other mediators of DNA repair, including BRCA1 and PALB2 — two proteins that are also frequently mutated in patients with familial breast cancer[7]. (In cells, PALB2 has a role in the localization of BRCA2 to DSBs and so in successful homologous recombination[7].) The existence of several RAD51 family members[8] further increases the complexity of homologous recombination. Finally, mounting evidence suggests that chromatin (complexes of DNA and histone proteins) plays a crucial part in regulating homologous recombination[9]. Purification of the entire BRCA2 molecule[1–3] sets the stage for future studies to reveal the full function of this large protein in homologous recombination. ■

Lee Zou *is at the Massachusetts General Hospital Cancer Center and the Department of Pathology, Harvard Medical School, Boston, Massachusetts 02129, USA.*
e-mail: zou.lee@mgh.harvard.edu

1. Jensen, R. B., Carreira, A. & Kowalczykowski, S. C. *Nature* **467,** 678–683 (2010).
2. Thorslund, T. *et al. Nature Struct. Mol. Biol.* doi:10.1038/nsmb.1905 (2010).
3. Liu, J., Doty, T., Gibson, B. & Heyer, W.-D. *Nature Struct. Mol. Biol.* doi:10.1038/nsmb.1904 (2010).
4. Thorslund, T. & West, S. C. *Oncogene* **26,** 7720–7730 (2007).
5. Yang, H. *et al. Science* **297,** 1837–1848 (2002).
6. Shin, D. S. *et al. EMBO J.* **22,** 4566–4576 (2003).
7. Moynahan, M. E. & Jasin, M. *Nature Rev. Mol. Cell Biol.* **11,** 196 –207 (2010).
8. San Filippo, J., Sung, P. & Klein, H. *Annu. Rev. Biochem.* **77,** 229 –257 (2008).
9. Harper, J. W. & Elledge, S. J. *Mol. Cell* **28,** 739 –745 (2007).

ATMOSPHERIC PHYSICS

Solar surprise?

The detection of unexpected changes in the Sun's spectral irradiance during the declining phase of the most recent solar cycle, and their implications for Earth's atmosphere, are intriguing. But they must be viewed as provisional. SEE LETTER P.696

ROLANDO R. GARCIA

Last year, Harder *et al.*[1] reported observations made by the satellite-borne Spectral Irradiance Monitor (SIM). They revealed unexpected behaviour in solar irradiance, resolved according to wavelength, during the declining phase of solar cycle 23 — unexpected, that is, compared with the currently accepted understanding of how irradiance varies during the 11-year solar cycle. On page 696 of this issue, Haigh *et al.*[2] go on to examine the surprising implications of the SIM measurements for stratospheric ozone and surface climate.

The SIM observations show very large increases in irradiance for the years 2004 minus 2007 at ultraviolet (UV) wavelengths, together with sizeable decreases at wavelengths longer than 400 nanometres, in the visible range. This result is depicted in Figure 1 of Haigh and colleagues' paper (page 696), in which SIM data are compared with the empirical model of Lean[3]. The latter, which is typical of models of this type[4–6], predicts irradiance variations on the basis of changes in solar structures such as sunspots and faculae (regions of relatively low and high brightness in the Sun's photosphere).

The model results (shown in black in the figure) indicate a modest difference in irradiance, for 2004 minus 2007, of less than 0.5 milliwatts per square metre per nanometre at all wavelengths shown. By contrast, SIM (shown in blue) measures changes of up to 5 mW m^{-2} nm^{-1} at UV wavelengths, and −1.5 mW m^{-2} nm^{-1} in the visible range beyond 400 nm. Although the irradiance changes measured by SIM are very different from those predicted by the model at almost all wavelengths, when these changes are integrated across the spectrum, they yield similar changes in the total downward solar flux at the top of Earth's atmosphere for 2004 minus 2007 (0.09 W m^{-2} for SIM and 0.11 W m^{-2} for Lean's model). This complicates the evaluation of the SIM observations because solar-cycle variability of the total solar irradiance, which is much better known than changes in wavelength-resolved irradiance, and which is reproduced by models such as Lean's, cannot be used to assess the validity of the observations.

Haigh *et al.*[2] used a photochemical–dynamical model to estimate the changes in ozone levels that would be produced by the irradiance variability observed by SIM. For ozone, they obtained changes from 2004 to 2007 of up to 2% at about 30 kilometres above Earth's surface. Furthermore, whereas the changes at 30 km altitude are in phase with solar activity, those above about 45 km are out of phase, such that lower ozone levels are obtained in 2004 than in 2007. This happens because the large variability observed by SIM in the UV leads to a large increase in the calculated abundance of hydrogen radicals, which are potent catalysts of ozone destruction, in the upper stratosphere (35–50 km) and in the mesosphere (above 50 km). On the other hand, the variability in irradiance predicted by Lean's model produces

much smaller changes in ozone levels, up to 0.8% at 40 km altitude, and these changes are in phase with solar activity throughout the stratosphere and lower mesosphere.

Haigh *et al.* argue that the pattern of ozone change implied by SIM is consistent with ozone observations made by the Microwave Limb Sounder (MLS) instrument onboard the EOS Aura satellite[7]. They carried out a multiple-regression fit to MLS ozone data in the tropics at two altitudes, centred near 30 km and 55 km. The regression predictors included two orthogonal indices of the tropical quasi-biennial oscillation (QBO, a wind oscillation that dominates the tropical stratosphere and affects the abundance of ozone therein)[8] and a solar activity index (SIM irradiances averaged over 200–400 nm). The resulting ozone–solar regression coefficient is positive at 30 km and negative at 55 km, consistent with their model-based predictions.

If the assessment by Haigh and her colleagues[2] is correct, the SIM observations would necessitate a re-evaluation of the mechanisms that produce solar variability, and how that variability affects the atmosphere. In addition to the impact on ozone, changes in SIM irradiance in the visible range, which are out of phase with the solar cycle, would require reconsideration of the role of solar variability in climate. Conventional wisdom[9] holds that long-term increases in solar activity should warm the climate. But the opposite would be true if long-term irradiance variability behaved like that measured by SIM over the declining phase of solar cycle 23.

Should these results send researchers back to their laboratories to rethink their theories and rewrite their computer models? It is probably too early for that. It remains to be convincingly shown that the wavelength-resolved irradiance observations are accurate and, if they are, that they apply to periods other than the declining phase of solar cycle 23.

In the first place, the length of the record used by Haigh *et al.* in their multiple-regression analysis is too short to yield unambiguous results: the two predictors used have timescales (28 months for the QBO, 11 years for the solar cycle) that are comparable to, or longer than, the length of the period analysed. Under these circumstances, there is no guarantee that multiple regression yields a physically meaningful apportionment of ozone variance among the predictors. Moreover, other regression analyses, based on ozone observations spanning more than 20 years[10,11], yield results that are broadly consistent with the conventional picture of solar variability. These analyses use data sets that do not extend much beyond the stratopause, the boundary — at about 50 km altitude — between the stratosphere and mesosphere. But they show no indication of a reversal in the sign of the relationship between solar activity and ozone levels in the upper stratosphere.

Haigh *et al.*[2] point out that at present there is insufficient evidence to validate the SIM observations. In fact, we will probably need SIM irradiance data and independent ozone observations over a much longer period, at least the length of one solar cycle, before we can begin to be sure that the surprising irradiance variability measured by SIM does not arise from instrumental drifts or corrections thereof. Even if this is not the case, it will be necessary to reconcile the implications of the SIM observations with previous analyses of ozone changes during the 11-year solar cycle that are not consistent with the implied effects of SIM-like irradiance changes. ■

Rolando R. Garcia *is in the Atmospheric Chemistry Division, National Center for Atmospheric Research, Boulder, Colorado 80305, USA.*
e-mail: rgarcia@ucar.edu

1. Harder, J. W., Fontenla, J. M., Pilewskie, P., Richard, E. C. & Woods, T. N. *Geophys. Res. Lett.* **36,** L07801, doi:10.1029/2008GL036797 (2009).
2. Haigh, J. D., Winning, A. R., Toumi, R. & Harder, J. W. *Nature* **467,** 696–699 (2010).
3. Lean, J. *Geophys. Res. Lett.* **27,** 2425–2428 (2000).
4. Krivova, N. A., Solanki, S. K. & Floyd, L. *Astron. Astrophys.* **452,** 631–639 (2006).
5. Krivova, N. A. & Solanki, S. K. *J. Astrophys. Astron.* **29,** 151–158 (2008).
6. Unruh, Y. C., Krivova, N. A., Solanki, S. K., Harder, J. W. & Kopp, G. *Astron. Astrophys.* **486,** 311–323 (2008).
7. Froidevaux, L. *et al. J. Geophys. Res.* **113,** D15S20, doi:10.1029/2007JD008771 (2008).
8. Baldwin, M. P. *et al. Rev. Geophys.* **39,** 179–229 (2001).
9. Lean, J. L. & Rind, D. H. *Geophys Res. Lett.* **36,** L15708, doi:10.1029/2009GL038932 (2009).
10. Randel, W. J. & Wu, F. *J. Geophys. Res.* **112,** D06313, doi:10.1029/2006JD007339 (2007).
11. Soukharev, B. E. & Hood, L. L. *J. Geophys. Res.* **111,** D020314, doi:10.1029/2006JD007107 (2006).

MYCOLOGY

Spores ride a cooperative wind

This beautiful image of synchronized spore discharge from an ascomycete fungus comes from a book published in 1791. A cross-disciplinary group of researchers led by Marcus Roper and Agnese Seminara have now brought twenty-first-century approaches, including algorithms used to model the behaviour of droplets in clouds, to bear on study of this phenomenon (M. Roper *et al. Proc. Natl Acad. Sci. USA* doi:10.1073pnas.1003577107; 2010). They find that simultaneous discharge in itself creates an air flow, a cooperatively generated wind, that allows the spores to travel much farther than if ejected alone — so enhancing their prospects of wafting farther afield.

Roper, Seminara and colleagues used a combination of simulations, analytical models and experiments to investigate spore release from species of ascomycetes, in which spores develop in sacs (asci) in a cup-shaped structure called the apothecium. Their subjects included species of *Sclerotinia* (a plant pathogen) and *Ascobolus* (a dung fungus).

Spores of *Sclerotinia*, for example, have to rise from the fruiting body on the ground to infect plant flowers. In experiments, confirmed by simulations, spores riding a cooperative wind behaved much like "frictionless projectiles". They travelled 10 centimetres or more, with the range probably being limited by gravity, compared with the 3 millimetres of those ejected on their own, which are soon halted by viscous drag. Moreover, if the cooperative spore plume hit an obstacle (which in experiments was mimicked by a glass slide, but in a natural setting might be a leaf), pressure differences in the plume resulted in spore movement around it.

The authors also used high-speed imaging to see how spore release is coordinated, and looked at various apothecial species. Their data show that the process is self-organized. In *Ascobolus*, the process of ejection is initiated in a few asci, perhaps by a highly local change in air pressure. A wave of spore discharge across the apothecium then ensues, possibly driven by an alteration in

COURTESY FARLOW LIBRARY OF CRYPTOGAMIC BOTANY, HARVARD UNIV.

elastic stress, that may arise from changes in the turgor pressure of cells that are interspersed among the asci.

As well as the practical aspect of providing insight into the dispersal dynamics of a plant pathogen (the species studied, *Sclerotinia sclerotiorum*, infects and damages many different crops), there is another angle to this line of research. The authors point out that synchronized spore discharge might catch the fancy of biologists interested in the evolution of self-organized cooperative behaviour. **Tim Lincoln**

50 Years Ago

An Introduction to the Logic of the Sciences. By R. Harré — This is a very welcome book. It should be said at the outset that the author's intention to write largely for undergraduates in science may prove a little on the modest side, since many students working for higher degrees would probably produce substantially better theses if they could find time to read what Dr. Harré has to relate … The grand point is that — from the aspect of discovery — disciplined insight came first, and the application of mathematical analysis afterwards. Essential as the latter is, momentous advances usually begin with remarkably simple premises. Incidentally, Max Planck is known to have fought long and hard in his mind against the consequences of his own quantum concept. The statistical and indiscriminate nature of much of modern physics was not to his liking. But that is the penalty of greatness. Questions like these are ably handled by Dr. Harré, and the moral is driven home.
From *Nature* 8 October 1960

100 Years Ago

Beet Sugar Making and its Chemical Control. By Y. Nikaido — In principle, the production of sugar from beetroots is a simple matter. The sugar and other soluble bodies are extracted from the sliced roots by diffusion in water; the juice thus obtained is purified from acids and other objectionable matter by "defecation" with lime, and after the excess of lime has been removed by treatment with carbonic acid, the liquor is concentrated by evaporation until the sugar crystallises out. Whilst, however, there is nothing complicated about the principle, successful and profitable production depends upon close attention to a number of points in respect of which the chemist's help is needed.
From *Nature* 6 October 1910

VISION

Neurons show their true colours

How do we tell red from green? Work on the primate retina shows how neural circuitry combines signals from individual cone photoreceptor cells to provide the basic building blocks for colour vision. SEE ARTICLE P.673

JONATHAN B. DEMB & DAVID H. BRAINARD

The processing of visual information begins in the retina, where specialized neurons called photoreceptors absorb light and stimulate multiple neural circuits. Each circuit generates specific patterns of electrical activity and converges on one of about 20 types of retinal ganglion cell. These cells' axons — the optic nerve fibres — then convey signals to various brain targets. For colour vision, specific retinal circuits compare the activity levels of different types of cone photoreceptor cells that have different spectral sensitivities[1]. On page 673 of this issue, Field *et al.*[2] describe simultaneous recordings from hundreds of ganglion cells, and a new method to map the inputs that these cells receive from individual cones. The results provide insight into the initial stages of colour vision.

Old World primates, including humans, have three types of cone photoreceptors that are maximally sensitive to long (L), middle (M) or short (S) wavelengths of light. In isolation, however, each cone is colour-blind because its activity depends on both the wavelength and intensity of incident light[1]. For example, an M cone's activity would be the same for a dim green light and a bright red light. Neural circuits derive a colour signal by comparing the activity of different cone types — cone opponency. In primate retinas, there are two broad classes of opponency: S–(L+M) opponency contrasts the activity of S cones with the combined activity of L and M cones, whereas L–M opponency contrasts the activity of L and M cones.

Primate cones are arranged in a mosaic, presenting several challenges to implementing cone opponency (Box 1). First, only a single cone exists at each location in the mosaic, so comparing cone signals confounds spatial and spectral information. Second, S cones are sparse, limiting the spatial resolution of S–(L+M) opponency. Third, the M/L cone arrangement is random or nearly random, leading to 'clumps' of either cell type[2,3]. This limits the spatial resolution of L–M opponent signals. For example, L–M resolution must be coarser at the centre of a clump of L cones than in a region where L and M cones alternate.

S–(L+M) opponency is an ancient and well-developed subsystem of colour vision. The genes encoding the S and M/L cone opsins — proteins that determine spectral sensitivity — diverged more than 500 million years ago[4]. Moreover, the primate retina contains specialized ganglion cells for computing S–(L+M) opponent signals. The small bistratified ganglion cells, for example, receive excitatory inputs from S cones through S-cone bipolar cells, and (L+M) cone signals oppose the S-cone signals by means of two distinct retinal pathways[5]. Similar cone opponency is mediated by ganglion cells in other mammals[6,7].

As for L–M opponency, this depends on a third opsin that arose in Old World primates less than 40 million years ago[4]. The neural mechanisms underlying L–M opponency have remained elusive. It could be that a specialized ganglion-cell type analogous to the small bistratified cells collects pure antagonistic signals from L and M cones[8]. So far, however, there has been no definitive identification of such cells.

Alternatively, L–M opponency might arise in other cells that existed before the emergence of the third cone opsin. A favourite candidate is the midget ganglion cell, named for its small size in the fovea — the central part of the retina[1]. In the fovea, each of these cells connects through a midget bipolar cell to a single cone, and thus receives an excitatory 'centre' signal that is selective for L or M. The centre signal is opposed by a 'surround' signal that is driven by the surrounding cones. Consequently, even if the surround draws randomly on L and M cones, most foveal midgets will exhibit a degree of L–M opponency because the centre signal is always pure[9] (Box 1).

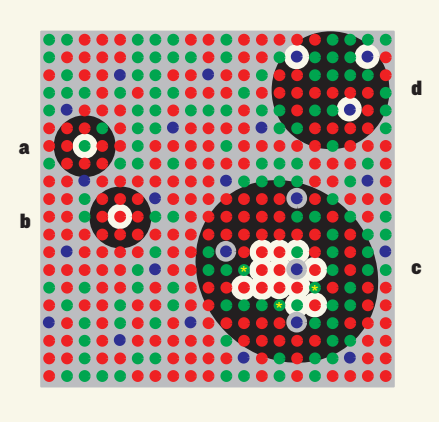
Each ganglion-cell type tiles the retina completely, but is larger in the retinal periphery than in the fovea[1]. The centre signals of the larger midget cells in the periphery therefore connect to a dozen or more cones, and analysis of these connections should provide information about the M/L selectivity of peripheral midget cells. The selective-wiring hypothesis predicts selective connections between each midget cell and either M or L cones; the random-wiring hypothesis, by contrast, predicts random connections[1]. Evidence has been reported in favour of both hypotheses[10,11], making it difficult to rule either out definitively. To date, one roadblock has been that inputs to

BOX 1

Cone opponency in the primate retina

A schematic representation of the mosaic arrangement of cones in the monkey retina that are sensitive to long (L, red), middle (M, green) and short (S, blue) wavelengths of light. S cones are rare, constituting 5–10% of the mosaic, and L cones outnumber M cones by about 2 to 1. L and M cones are arranged near randomly. For simplicity, the cone mosaic is shown as rectangular in arrangement. The actual mosaics are more hexagonal and also less regular[3]: in the central all-cone fovea, cones are tightly packed, whereas in the retinal periphery they are separated by the more numerous rod photoreceptors, which are used for seeing in dim light.

a, In the central fovea, a midget ganglion cell combines an excitatory centre-region signal driven by a single cone — outlined in white — with an opposing signal from the surrounding region driven by surrounding cones (outlined in black). This cell would have M–L opponency.



b, A midget ganglion cell that aligns with a clump of L cones, however, would lack opponency, because the centre and surround are driven by the same class of cones.

c, In contrast to foveal midget ganglion cells, a peripheral midget cell may receive excitatory input from a dozen cones. There is a bias, measurable across the population of cells, in which the centre region of the receptive field is wired to favour one cone type, tending towards cone purity. In this case, for example, the midget cell avoids some M cones near those that comprise the centre (marked with a yellow asterisk) in favour of the L cones that make up the majority. This cell will have L–M opponency. In this example, the S cones are skipped, but a subgroup of midget cells is connected with S cones[2].

d, For S−(L+M) opponency, a small bistratified ganglion cell receives an S-cone input that is opposed by M/L cone inputs. Because of the irregular layout of S cones and other features of the underlying neural circuitry, these cells lack the canonical centre–surround structure found in many other ganglion-cell types, including midget cells[1,5,8]. **J.B.D. & D.H.B.**

ganglion cells could not be characterized on a cone-by-cone basis.

In a technical tour de force, Field *et al.*[2] used a multi-electrode array to generate very high-resolution centre–surround receptive-field maps simultaneously from hundreds of macaque retinal ganglion cells *in vitro*. The array has hundreds of metal electrodes that detect the rates of action potentials that normally travel down the optic nerve. With recordings made in as little as one hour, the authors could not only measure inputs to ganglion-cell centres and surrounds at the resolution of individual cones, but also determine the type of each cone with high accuracy. Their data show how individual peripheral midget cells draw on cones of different types.

Across the midget-cell population, Field and co-workers observed a continuum of cone selectivity in the cell centres, ranging from pure M or L input to a random mix of the two. An elegant re-sampling analysis[2] showed that although the random-wiring hypothesis provides a first-order account of the chromatic properties of the midget-cell population[11], it fails to account for these properties in detail. That is, if the midget centres drew randomly on cones — irrespective of the cone type — there would be less opponency in the population as a whole than is actually observed. Instead, selectivity in the weighting of cone inputs biases the centres towards M or L purity. This bias enhances L–M opponency across the midget-cell population, and could thereby improve the quality of red–green colour vision in the retinal periphery.

Of the questions that Field and colleagues' study highlights, two are particularly fascinating. First, how does the observed specificity of cone inputs come about? Expression by gene therapy of a third cone opsin in the retinas of adult New World monkeys that had been red–green colour-blind since birth[12] allowed them to distinguish between new colours. Theoretically, these new abilities could be mediated by random wiring into the cone mosaic, which would be sufficient to generate some degree of cone opponency[9]. But it would be interesting to determine this experimentally. This could be done using Field and colleagues' method for mapping the receptive fields of ganglion cells at individual-cone resolution. If these cells show random wiring following gene therapy in adulthood, similar receptive-field mapping experiments could be performed in the subset of female New World monkeys born with the three cone types[13,14], to determine whether the plasticity required for cone-selective wiring occurs during development. If biased cone connectivity never occurs in New World monkeys, it would suggest that cone-selective wiring developed in Old World primates on an evolutionary timescale.

And how crucial is the observed degree of selective cone input for red–green colour vision? Psychophysical measurements[8,15] have shown that red–green sensitivity degrades in the retinal periphery, and that its spatial resolution is coarse. The question is whether this performance is better than would be expected from random wiring alone[9]. Field and co-workers' physiological measurements[2], which describe the properties of the entire population of midget cells in a peripheral region of the retina, may help us to more sharply analyse such psychophysical data to understand the precise functional implications of the bias towards selective wiring. ■

Jonathan B. Demb *is in the Department of Ophthalmology and Visual Sciences, and the Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, Michigan 48105, USA.*
David H. Brainard *is in the Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.*
e-mails: jdemb@umich.edu; brainard@psych.upenn.edu

1. Dacey, D. M. *Annu. Rev. Neurosci.* **23,** 743–775 (2000).
2. Field, G. D. *et al. Nature* **467,** 673–677 (2010).
3. Hofer, H., Carroll, J., Neitz, J., Neitz, M. & Williams, D. R. *J. Neurosci.* **25,** 9669–9679 (2005).
4. Nathans, J., Thomas, D. & Hogness, D. S. *Science* **232,** 193–202 (1986).
5. Crook, J. D. *et al. J. Neurosci.* **29,** 8372–8387 (2009).
6. Yin, L., Smith, R. G., Sterling, P. & Brainard, D. H. *J. Neurosci.* **29,** 2706–2724 (2009).
7. Li, W. & DeVries, S. H. *Nature Neurosci.* **9,** 669–675 (2006).
8. Calkins, D. J. & Sterling, P. *Neuron* **24,** 313–321 (1999).
9. Lennie, P., Haake, P. W. & Williams, D. R. in *Computational Models of Visual Processing* (eds Landy, M. S. & Movshon, J. A.) 71–82 (MIT Press, 1991).
10. Martin, P. R. *et al. Nature* **410,** 933–936 (2001).
11. Diller, L. *et al. J. Neurosci.* **24,** 1079–1088 (2004).
12. Mancuso, K. *et al. Nature* **461,** 784–787 (2009).
13. Jacobs, G. H. *Phil. Trans. R. Soc. B* **364,** 2957–2967 (2009).
14. Blessing, E. M. *et al. J. Physiol. (Lond.)* **557,** 229–245 (2004).
15. Hansen, T., Pracejus, L. & Gegenfurtner, K. R. *J. Vis.* **9**(4), article 26 (2009).

doi:10.1038/nature09424

Functional connectivity in the retina at the resolution of photoreceptors

Greg D. Field[1]*, Jeffrey L. Gauthier[1]*†, Alexander Sher[2]*, Martin Greschner[1], Timothy A. Machado[1]†, Lauren H. Jepson[1], Jonathon Shlens[1], Deborah E. Gunning[3], Keith Mathieson[3], Wladyslaw Dabrowski[4], Liam Paninski[5], Alan M. Litke[2] & E. J. Chichilnisky[1]

To understand a neural circuit requires knowledge of its connectivity. Here we report measurements of functional connectivity between the input and ouput layers of the macaque retina at single-cell resolution and the implications of these for colour vision. Multi-electrode technology was used to record simultaneously from complete populations of the retinal ganglion cell types (midget, parasol and small bistratified) that transmit high-resolution visual signals to the brain. Fine-grained visual stimulation was used to identify the location, type and strength of the functional input of each cone photoreceptor to each ganglion cell. The populations of ON and OFF midget and parasol cells each sampled the complete population of long- and middle-wavelength-sensitive cones. However, only OFF midget cells frequently received strong input from short-wavelength-sensitive cones. ON and OFF midget cells showed a small non-random tendency to selectively sample from either long- or middle-wavelength-sensitive cones to a degree not explained by clumping in the cone mosaic. These measurements reveal computations in a neural circuit at the elementary resolution of individual neurons.

Colour vision requires neural circuitry to compare signals from spectrally distinct cone types. For example, the signature of primate colour vision—red–green and blue–yellow colour opponency—implies that neural circuits pit signals from different cone types against one another. However, the pattern of connectivity between the long- (L), middle- (M) and short (S)-wavelength-sensitive cones and various retinal ganglion cell (RGC) types, which determines how colour signals are transmitted in parallel pathways to the brain, remains incompletely understood[1–9]. To probe the circuitry for colour vision more fully, the pattern of connectivity between the full lattice of cone photoreceptors and complete populations of RGCs of several types was measured in primate retina.

Hundreds of RGCs were simultaneously recorded in the peripheral macaque retina using large-scale electrophysiological recordings[10–12]. The light responses of each cell were characterized by computing the spike-triggered average of a spatio-temporal white noise stimulus (see Methods). From the spike-triggered average, several features of light response were identified, including the spatial receptive field and the response time course. Classification based on these properties was used to identify functionally distinct RGC classes (Fig. 1a, centre). The receptive fields of each cell class formed a regular mosaic covering the region of retina recorded[12–15]. This revealed that each functionally defined cell class corresponded to one RGC type, because the dendrites of each RGC type uniformly tile the retinal surface[16,17]. Density and light response properties were used to identify the ON and OFF midget, ON and OFF parasol and small bistratified cell types, which collectively account for ~75% of RGCs[5]. In many cases, receptive-field mosaics had few or no gaps, indicating that nearly every cell was recorded.

To resolve the fine structure of receptive fields, stimuli with tenfold smaller pixels (5 μm × 5 μm) were used. At this resolution, receptive fields did not conform to the smooth Gaussian approximation used in Fig. 1a (centre panel) and in previous studies[18]. Instead, each receptive field was composed of punctate islands of light sensitivity (Fig. 1a, outer panels). The separation between islands was roughly equal to the spacing of the cone lattice, consistent with the idea that each island reflected the contribution of a single cone[10,19]. To test this hypothesis, locations of islands were compared to photographs of cone outer segments labelled with peanut agglutinin; a close alignment was observed (Fig. 1b and Supplementary Methods).

The spectral type of each cone—L, M or S—was identified using the relative magnitudes of the three display primaries in the spike-triggered average at its location (Fig. 2a). These values, accumulated across all cones in a recording, formed three distinct clusters (Fig. 2b) aligned with the spectral sensitivities of the macaque cones (coloured lines)[20]. S cones were easily identifiable, L and M cones were somewhat less so (Fig. 2b, c) because of their overlapping spectral sensitivities.

The full cone mosaic was visualized by pooling information from all recorded RGCs. This was accomplished by fitting the receptive fields of all RGCs with a model in which each receptive field is approximated by a weighted sum of Gaussian functions centred on the locations of cones (Supplementary Methods). This approach revealed nearly complete cone mosaics (Fig. 2d, e). The relative frequencies of L, M and S cones were in a ratio of roughly 8:4:1 (average of six data sets)[21].

The functional connectivity between each RGC and the cones within its receptive field was summarized by assigning an input strength to each cone, equal to the weight in the model fit derived from the spike-triggered average (Supplementary Methods). This permitted well-constrained estimation of the inputs of weak cones, including those in the receptive-field surround (Fig. 2g), because cone locations were robustly identified using data from multiple cells. The receptive field of each RGC was summarized graphically with a collection of radiating lines connecting to cones: the thickness of each line is proportional to the weight; white lines represent the centre and black lines the surround of the receptive field (Fig. 2h). This representation was used to visualize several complete mosaics of RGCs

[1]Systems Neurobiology Laboratories, Salk Institute for Biological Studies, La Jolla, California 92037, USA. [2]Santa Cruz Institute for Particle Physics, University of California, Santa Cruz, California 95064, USA. [3]Department of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK. [4]Faculty of Physics and Applied Computer Science, AGH University of Science and Technology, 23 30-059 Krakow, Poland. [5]Department of Statistics and Center for Theoretical Neuroscience, Columbia University, New York, New York 10027, USA. †Present addresses: Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey 08544, USA (J.L.G.); Department of Statistics and Center for Theoretical Neuroscience, Columbia University, New York, New York 10027, USA (T.A.M.).
*These authors contributed equally to this work.

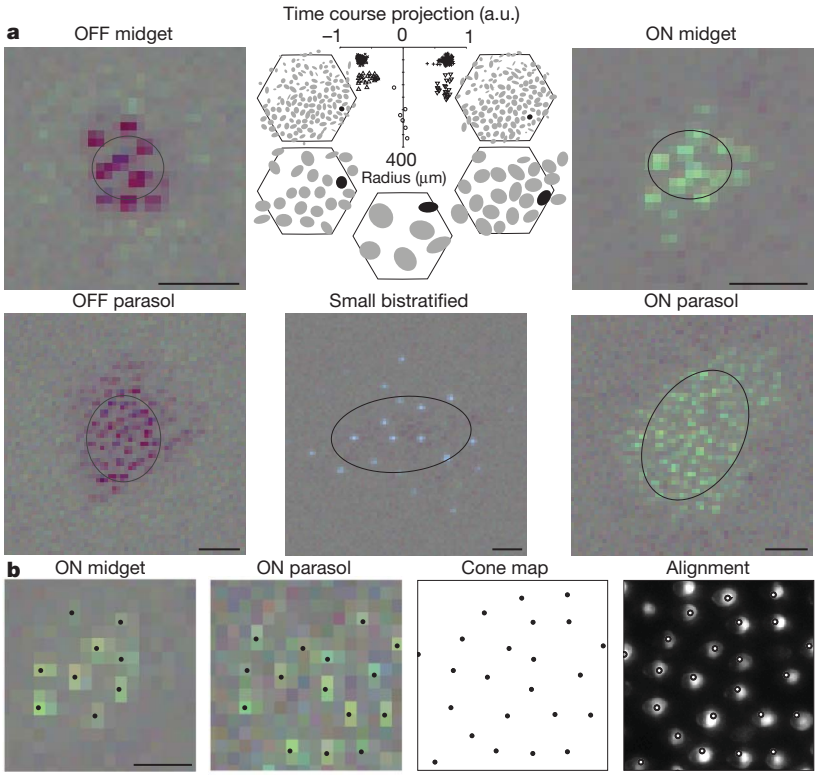
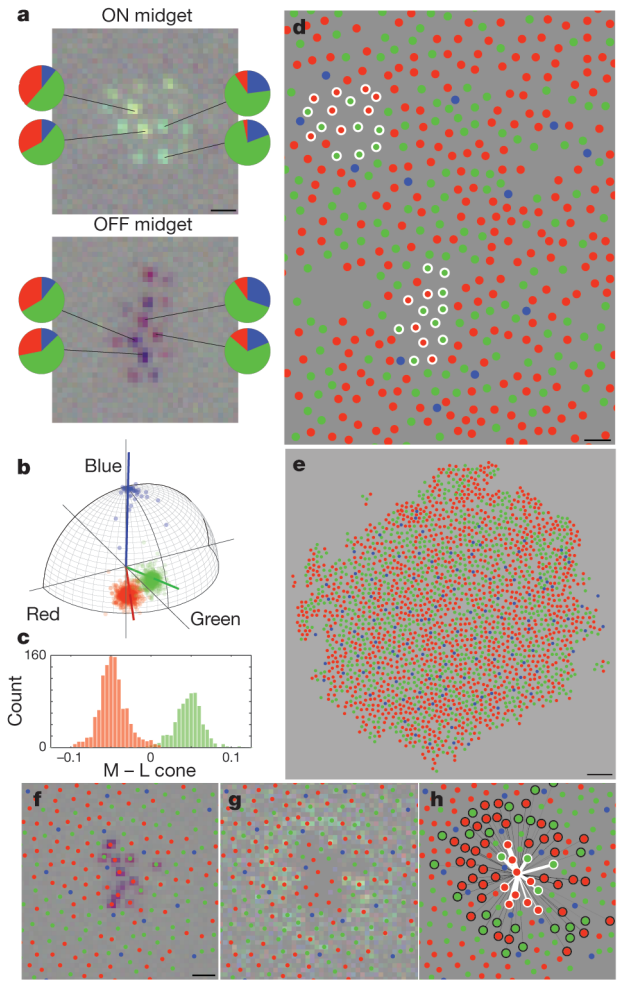


Figure 1 | Cell-type classification and receptive fields at single-cone resolution. **a**, Receptive fields of 323 RGCs recorded simultaneously from isolated macaque retina were measured using reverse correlation with white noise stimuli. Centre panel shows receptive-field radius versus first principal component of response time course; clusters reveal distinct cell types. a.u., arbitrary units. Hexagons surrounding centre panel show outline of electrode array and ellipses show Gaussian fits to receptive fields of cells from each cluster. The outer panels show fine-grained spatial receptive-field profiles for highlighted cells. Scale bars, 50 μm. **b**, From left to right, panels 1 and 2 show spatial receptive-field profiles of two cells, with putative locations of cones (black dots) identified by thresholding. Scale bar, 25 μm. Panel 3 shows the putative cone map accumulated across cells and panel 4 shows putative cone map overlaid on a photograph of cone outer segments labelled with peanut agglutinin.



receiving input from a complete mosaic of cones (Fig. 3). A total of 1,961 RGCs receiving input from a total of 17,380 cones in seven preparations were examined.

These connectivity diagrams provide insight into the specificity of L-, M- and S-cone inputs to the RGC types mediating high-resolution vision and colour vision. This specificity has been a source of controversy in previous work (see Supplementary Discussion).

Previous studies provide conflicting accounts of S-cone inputs to midget and parasol cells[1,3,4,6,8,22–28]. In the present data OFF midget cells frequently received at least one strong functional S-cone input, whereas ON midget, ON parasol and OFF parasol cells did so much less frequently (for example, see Fig. 3). At the same time, all four RGC types sampled essentially the entire mosaic of L and M cones. For example, in one recording an ON midget cell received no input from three S cones located within its receptive-field centre (Fig. 4a). Yet, two of these S cones were sampled by nearby OFF midget cells (Fig. 4b, c). On average, S cones were strongly sampled by OFF midget cells about five times more frequently than by ON midget and parasol cells (Fig. 4d). Further analysis showed that ON midget cells had a tendency to sample weakly from S cones (Supplementary Methods). The sampling of S cones by OFF midget cells confirms a prediction from anatomical work[25]: OFF midget bipolar cells contact S cones in

Figure 2 | Cone-type identification and inputs to RGCs. **a**, The spectral sensitivity of cones providing input to two cells is represented by the relative magnitude of the red, green and blue spike-triggered average values (a.u.) at their locations. **b**, For every cone in one recording, these values are shown as points on a sphere. Coloured lines indicate spectral sensitivity of macaque cones. Point colour indicates classification as L (red), M (green), or S (blue). **c**, L- and M-cone discriminability quantified by projection along the line joining L- and M-cone loci. Bar colour indicates classification. S cones excluded. **d**, Assembled cone mosaic from all RGCs over a region. Cones from **a** are circled. **e**, Full mosaic of 2,373 cones from one recording. **f**, Cone mosaic overlaid on STA, showing the strength of cone inputs. **g**, Weaker cone inputs in receptive-field surround revealed by truncating positive values and renormalizing. **h**, Connectivity diagram, with line thickness proportional to strength of each cone input. Surround (black) line thicknesses were increased fivefold relative to centre (white) line thicknesses for visibility. Scale bars, 25 μm (**a**, **d**, **f**), 75 μm (**e**).

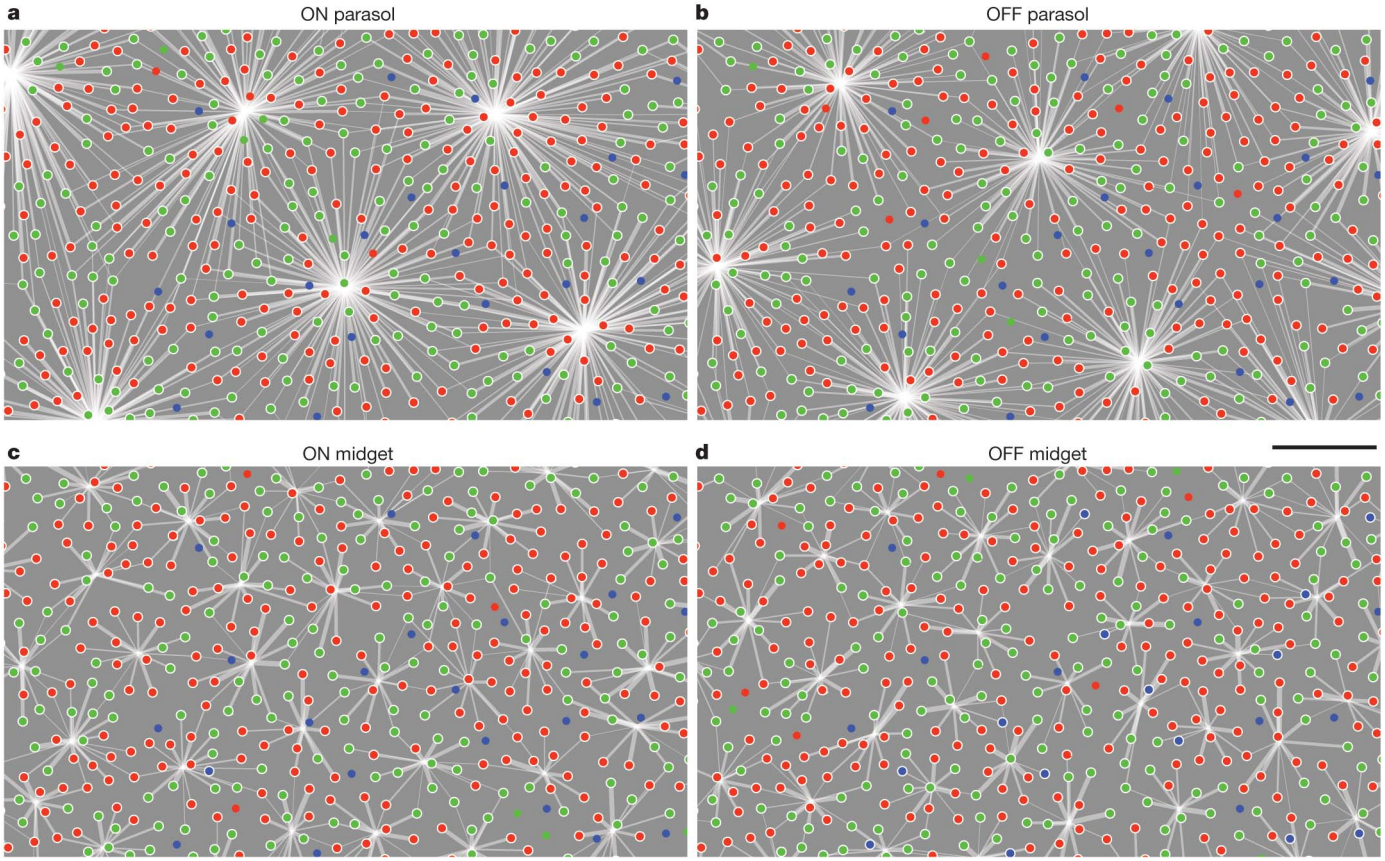


Figure 3 | Full functional sampling of cone lattice by four RGC types. **a–d**, Each panel shows cones identified in a single recording (red, green and blue dots) sampled by receptive-field centres of RGCs of a single type. Cones are identical in all panels. Cones providing input to at least one RGC are highlighted with an annulus. Scale bar, 50 μm.

the central retina, therefore, OFF midget RGCs should receive S-cone input. The absence of S-cone input to parasol cells also confirms recent findings[6]. An important question for future work is whether the S-cone signals carried by OFF midget cells contribute to blue–yellow and red–green opponent colour vision.

The specificity of L- and M-cone inputs to peripheral midget cells, which is thought to underlie red–green opponent colour vision, has also been debated[1,8,9,29–39]. One study suggested that midget cells tend to selectively sample from either L or M cones in the receptive-field centre, producing red–green colour opponency by pitting relatively pure L- or M-cone centre signal against a mixture of L- and M-cone signals from the surround[2]. Another study suggested that the receptive-field surround may enhance opponency by sampling predominantly from the cone type less strongly sampled by the centre[7], consistent with previous work[3,39]. Yet another study found no evidence for colour opponency in peripheral midget cells[40], indicating that cone sampling is random in both receptive-field centre and surround.

In the present data, a significant fraction of peripheral midget cells showed red–green colour opponency (Fig. 4h). Opponency was quantified by calculating the relative strengths of the total input from L, M and S cones, obtained with cone-isolating stimuli (Supplementary Methods)[1]. To examine separately the roles of the receptive-field centre and surround in opponency, cones were defined as contributing primarily to the receptive-field centre or surround based on the sign of their input and their location (Supplementary Methods). Interestingly, opponency was often strong in those midget cells that sampled either L or M cones dominantly or exclusively in the receptive-field centre, whereas in the receptive-field surround cone sampling seemed indiscriminate (for example, Fig. 4e–g). These observations are consistent with the hypothesis that sampling bias towards either L or M cones in the receptive-field centre mediates opponency. However, across the population of midget cells the purity of cone input to the receptive-field centre varied widely (for example, see Fig. 3c, d), raising the alternative possibility of random sampling in both centre and surround.

To test the randomness of L- and M-cone sampling quantitatively, statistical analysis was performed, beginning with cones in the receptive-field centre. First, an index of cone input purity was computed for all midget cells in each preparation (Fig. 4i and Methods). The width of the distribution of purity indices quantifies the diversity of cone inputs to recorded cells (Fig. 4j, top). The purity indices were then re-computed after artificially and randomly permuting the identities of L and M cones (Fig. 4j, bottom), while preserving all other aspects of the data. If the connectivity between L and M cones and midget cells were random, then permutation of cone identities would not significantly alter the distribution of purity indices. In fact, the distribution was narrower after permutation, and fewer cells with pure L- or M-cone centres were observed (index values near ± 1). This tendency was statistically significant, and was observed in nearly all of the populations of ON and OFF midget cells examined (Fig. 4k). Although these deviations from random connectivity are small, they imply that the receptive-field centres of midget cells tend to favour inputs from either L or M cones, contributing to red–green opponency. In contrast, the same analysis applied to cones in the receptive-field surround yielded results consistent with the hypothesis of random sampling (Fig. 4l).

In principle, the observed purity (width of distribution of purity indices) could be produced by clumping in the cone mosaic, that is, aggregation of cones of the same type. Clumping would increase the proportion of midget cells with centres dominated by one cone type. Evidence for a weak cone clumping was reported in central human retina and peripheral macaque retina[41,42], but the implications for colour opponency in midget cells have not been examined experimentally. In

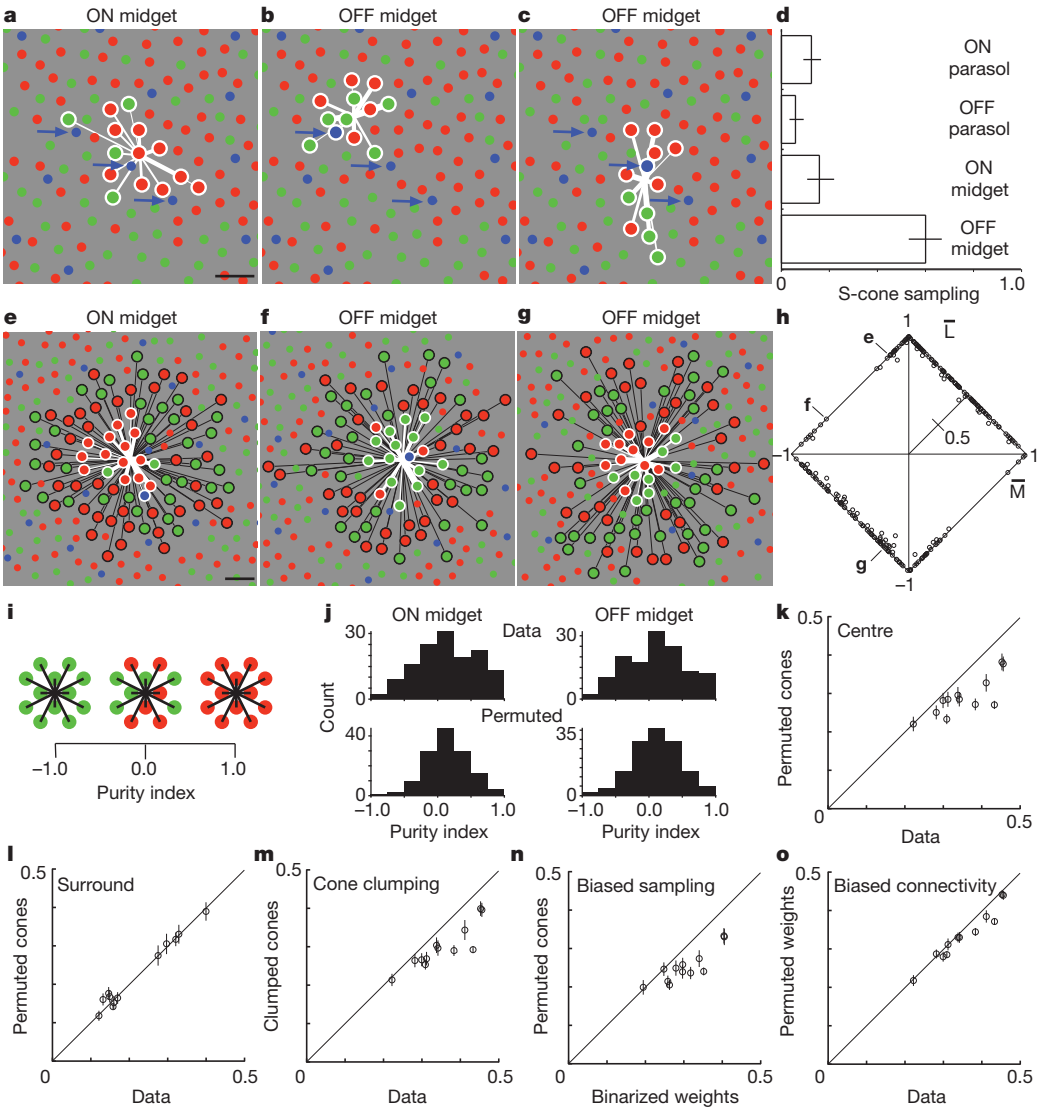


Figure 4 | Cone-type specificity. **a**, ON midget cell lacking input from nearby S cones (arrows). **b**, **c**, OFF midget cells receiving input from these cones. **d**, Frequency of strong S-cone sampling by each cell type. **e**, **f**, Two midget cells with relatively pure L- or M-cone input. **g**, Midget cell with mixed L- and M-cone input. **h**, Normalized L-, M- and S-cone inputs to all midget cells in one recording, obtained with cone-isolating stimuli. Abscissa, $\bar{M} = M/(|L|+|M|+|S|)$; ordinate, $\bar{L} = L/(|L|+|M|+|S|)$. Diagonals, no S-cone input. Upper-right and lower-left quadrants, same-sign (non-opponent) L- or M-cone input. Lower-right and upper-left quadrants, opposite-sign (opponent) L- or M-cone input. The letters in bold refer to cells from previous panels. **i**, Purity index schematic. **j**, Top, purity index for ON and OFF midget cells in one recording; width (standard deviation (s.d.)) 0.45, 0.44 respectively. Bottom, purity index after random permutation of L and M cones; width 0.37 ± 0.04 and 0.36 ± 0.04, respectively (mean $\pm$ 2 s.d. across permutations). **k**, Comparison of purity distribution width in data and permutations. Each point represents >50 simultaneously recorded ON or OFF midget cells. Error bars are 1 s.d. across permutations. **l**, As **k**, using cones from receptive-field surround. **m**, Using random cone mosaics with clumping matched to data. **n**, Using binarized cone weights (0,1). **o**, Using random permutation of cone weights in receptive-field centre. Scale bars, 25 μm (**a**–**c**, **e**–**g**).

the present data, tests for cone clumping on the scale of midget-cell receptive fields indicated a weak tendency towards clumping in three out of seven recordings (Methods). However, clumping alone cannot account for the observed purity, because artificial cone mosaics with the same degree of clumping reduced purity (Fig. 4m). Thus, the purity indicates that midget cells sample L- and M-cone inputs, through the retinal network, in a selective manner.

Selective sampling could be produced if (1) each midget cell receives inputs from one cone type more frequently than the other; and/or (2) each midget cell weights inputs from one cone type more strongly than the other. The results of statistical analyses were consistent with both factors. In model (1), the number of cones sampled by each midget cell should be skewed towards one cone type or the other. Therefore, random permutation of L- and M-cone identities should reduce purity, even if the relative weights of different cone inputs to each cell are ignored by binarizing them. This prediction was confirmed (Fig. 4n). In model (2), the weights on cone inputs to each midget cell should be skewed towards one cone type or the other. Therefore, random permutation of the strength of all the cone inputs within the receptive field of each midget cell should reduce purity. This prediction was also confirmed (Fig. 4o), although the effect was modest. Control analysis indicated that these findings were not a result of the tapering receptive-field profile of RGCs or clumping in the cone mosaic (data not shown).

Selective sampling raises questions about the mechanisms by which functional connectivity between cones and RGCs is coordinated. The divergence of the L- and M-cone photopigments in primates is relatively recent[43], and there is little evidence for segregation of L- and M-cone signals in the retinal circuitry[44]. L and M cones are apparently indistinguishable both anatomically and histochemically. Furthermore, there is only tentative anatomical evidence of differences in retinal circuits carrying L- and M-cone signals[35], in contrast to the markedly

different pathway that conveys S-cone signals[45,46]. Thus, there is no candidate structural or molecular basis for selective sampling. In principle, selective sampling could arise from activity-dependent adjustment of synaptic inputs. At the eccentricites recorded, midget bipolar cells usually contact only one cone[47], providing an opportunity for midget RGCs to selectively sample inputs from bipolar cells carrying signals from one cone type. These bipolar cells could be distinguished by the statistics of their responses to natural scenes[48]. The possibility of such an adaptive mechanism is broadly consistent with recent observations of long-term adaptability in retinal signals[49] and colour vision[50].

METHODS SUMMARY

Extracellular multi-electrode recordings were obtained from ganglion cells of isolated retinas taken from macaque monkeys (*Macaca fascicularis* and *Macaca mulatta*) used in other laboratories[10]. Spikes from several hundred cells were segregated off-line[11]. Reverse correlation of spike times with white noise checkerboard stimuli focused on the retina was used to obtain receptive-field maps for all cells (Fig. 1). ON and OFF midget and parasol cells and small bistratified cells were classified according to their characteristic light responses and density (Fig. 1)[12,14,15]. Locations, spectral sensitivity, and input strengths of L, M and S cones to these cell types were obtained from the fine-grained receptive-field maps (Figs 2 and 3). To test for selective functional connectivity, the measured spatial arrangement and input strengths of the three cone types to RGCs were compared to artificially modified representations (Fig. 4).

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

Received 1 March; accepted 11 August 2010.

1. Derrington, A. M., Krauskopf, J. & Lennie, P. Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol. (Lond.)* **357,** 241–265 (1984).
2. Martin, P. R., Lee, B. B., White, A. J., Solomon, S. G. & Ruttiger, L. Chromatic sensitivity of ganglion cells in the peripheral primate retina. *Nature* **410,** 933–936 (2001).
3. Reid, R. C. & Shapley, R. M. Space and time maps of cone photoreceptor signals in macaque lateral geniculate nucleus. *J. Neurosci.* **22,** 6158–6175 (2002).
4. Chatterjee, S. & Callaway, E. M. S cone contributions to the magnocellular visual pathway in macaque monkey. *Neuron* **35,** 1135–1146 (2002).
5. Dacey, D. M. in *The Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 281–301 (MIT Press, 2004).
6. Sun, H., Smithson, H. E., Zaidi, Q. & Lee, B. B. Specificity of cone inputs to macaque retinal ganglion cells. *J. Neurophysiol.* **95,** 837–849 (2006).
7. Buzas, P., Blessing, E. M., Szmajda, B. A. & Martin, P. R. Specificity of M and L cone inputs to receptive fields in the parvocellular pathway: random wiring with functional bias. *J. Neurosci.* **26,** 11148–11161 (2006).
8. Wiesel, T. N. & Hubel, D. H. Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *J. Neurophysiol.* **29,** 1115–1156 (1966).
9. Jacobs, G. H. & De Valois, R. L. Chromatic opponent cells in squirrel monkey lateral geniculate nucleus. *Nature* **206,** 487–489 (1965).
10. Chichilnisky, E. J. & Baylor, D. A. Receptive-field microstructure of blue-yellow ganglion cells in primate retina. *Nature Neurosci.* **2,** 889–893 (1999).
11. Litke, A. M. *et al.* What does the eye tell the brain? Development of a system for the large scale recording of retinal output activity. *IEEE Trans. Nucl. Sci.* **51,** 1434–1440 (2004).
12. Frechette, E. S. *et al.* Fidelity of the ensemble code for visual motion in primate retina. *J. Neurophysiol.* **94,** 119–135 (2005).
13. Devries, S. H. & Baylor, D. A. Mosaic arrangement of ganglion cell receptive fields in rabbit retina. *J. Neurophysiol.* **78,** 2048–2060 (1997).
14. Chichilnisky, E. J. & Kalmar, R. S. Functional asymmetries in ON and OFF ganglion cells of primate retina. *J. Neurosci.* **22,** 2737–2747 (2002).
15. Field, G. D. *et al.* Spatial properties and functional organization of small bistratified ganglion cells in primate retina. *J. Neurosci.* **27,** 13261–13272 (2007).
16. Wässle, H., Peichl, L. & Boycott, B. B. Dendritic territories of cat retinal ganglion cells. *Nature* **292,** 344–345 (1981).
17. Dacey, D. M. The mosaic of midget ganglion cells in the human retina. *J. Neurosci.* **13,** 5334–5355 (1993).
18. Rodieck, R. W. *The First Steps in Seeing* (Sinauer, 1998).
19. Sincich, L. C., Zhang, Y., Tiruveedhula, P., Horton, J. C. & Roorda, A. Resolving single cone inputs to visual receptive fields. *Nature Neurosci.* **12,** 967–969 (2009).
20. Baylor, D. A., Nunn, B. J. & Schnapf, J. L. Spectral sensitivity of cones of the monkey Macaca fascicularis. *J. Physiol. (Lond.)* **390,** 145–160 (1987).
21. Roorda, A. & Williams, D. R. The arrangement of the three cone classes in the living human eye. *Nature* **397,** 520–522 (1999).
22. Lee, B. B., Martin, P. R. & Valberg, A. The physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina. *J. Physiol. (Lond.)* **404,** 323–347 (1988).
23. Kaiser, P. K., Lee, B. B., Martin, P. R. & Valberg, A. The physiological basis of the minimally distinct border demonstrated in the ganglion cells of the macaque retina. *J. Physiol. (Lond.)* **422,** 153–183 (1990).
24. Lee, B. B., Pokorny, J., Smith, V. C. & Kremers, J. Responses to pulses and sinusoids in macaque ganglion cells. *Vision Res.* **34,** 3081–3096 (1994).
25. Klug, K., Herr, S., Ngo, I. T., Sterling, P. & Schein, S. Macaque retina contains an S-cone OFF midget pathway. *J. Neurosci.* **23,** 9881–9887 (2003).
26. Tailby, C., Szmajda, B. A., Buzas, P., Lee, B. B. & Martin, P. R. Transmission of blue (S) cone signals through the primate lateral geniculate nucleus. *J. Physiol. (Lond.)* **586,** 5947–5967 (2008).
27. Lee, S. C. & Grunert, U. Connections of diffuse bipolar cells in primate retina are biased against S-cones. *J. Comp. Neurol.* **502,** 126–140 (2007).
28. Solomon, S. G., Lee, B. B., White, A. J., Ruttiger, L. & Martin, P. R. Chromatic organization of ganglion cell receptive fields in the peripheral retina. *J. Neurosci.* **25,** 4527–4539 (2005).
29. De Valois, R. L., Abramov, I. & Jacobs, G. H. Analysis of response patterns of LGN cells. *J. Opt. Soc. Am.* **56,** 966–977 (1966).
30. de Monasterio, F. M., Gouras, P. & Tolhurst, D. J. Trichromatic colour opponency in ganglion cells of the rhesus monkey retina. *J. Physiol. (Lond.)* **251,** 197–216 (1975).
31. Lee, B. B. Receptive field structure in the primate retina. *Vision Res.* **36,** 631–644 (1996).
32. Dacey, D. M. Parallel pathways for spectral coding in primate retina. *Annu. Rev. Neurosci.* **23,** 743–775 (2000).
33. Gegenfurtner, K. R. Cortical mechanisms of colour vision. *Nature Rev. Neurosci.* **4,** 563–572 (2003).
34. Solomon, S. G. & Lennie, P. The machinery of colour vision. *Nature Rev. Neurosci.* **8,** 276–286 (2007).
35. Calkins, D. J., Schein, S. J., Tsukamoto, Y. & Sterling, P. M and L cones in macaque fovea connect to midget ganglion cells by different numbers of excitatory synapses. *Nature* **371,** 70–72 (1994).
36. Jusuf, P. R., Martin, P. R. & Grunert, U. Random wiring in the midget pathway of primate retina. *J. Neurosci.* **26,** 3908–3917 (2006).
37. Ghosh, K. K., Goodchild, A. K., Sefton, A. E. & Martin, P. R. Morphology of retinal ganglion cells in a new world monkey, the marmoset Callithrix jacchus. *J. Comp. Neurol.* **366,** 76–92 (1996).
38. Lennie, P., Haake, W. & Williams, D. R. in *Computational Models of Visual Processing* (eds Landy, M. & Movshon, J. A.) 71–82 (MIT Press, 1991).
39. Reid, R. C. & Shapley, R. M. Spatial structure of cone inputs to receptive fields in primate lateral geniculate nucleus. *Nature* **356,** 716–718 (1992).
40. Diller, L. *et al.* L and M cone contributions to the midget and parasol ganglion cell receptive fields of macaque monkey retina. *J. Neurosci.* **24,** 1079–1088 (2004).
41. Hofer, H., Carroll, J., Neitz, J., Neitz, M. & Williams, D. R. Organization of the human trichromatic cone mosaic. *J. Neurosci.* **25,** 9669–9679 (2005).
42. Packer, O. S., Williams, D. R. & Bensinger, D. G. Photopigment transmittance imaging of the primate photoreceptor mosaic. *J. Neurosci.* **16,** 2251–2260 (1996).
43. Nathans, J. The evolution and physiology of human color vision: insights from molecular genetic studies of visual pigments. *Neuron* **24,** 299–312 (1999).
44. Calkins, D. J. & Sterling, P. Evidence that circuits for spatial and color vision segregate at the first retinal synapse. *Neuron* **24,** 313–321 (1999).
45. Kouyama, N. & Marshak, D. W. Bipolar cells specific for blue cones in the macaque retina. *J. Neurosci.* **12,** 1233–1252 (1992).
46. Dacey, D. M. & Lee, B. B. The 'blue-on' opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. *Nature* **367,** 731–735 (1994).
47. Wassle, H., Grunert, U., Martin, P. R. & Boycott, B. B. Immunocytochemical characterization and spatial distribution of midget bipolar cells in the macaque monkey retina. *Vision Res.* **34,** 561–579 (1994).
48. Wachtler, T., Doi, E., Lee, T. & Sejnowski, T. J. Cone selectivity derived from the responses of the retinal cone mosaic to natural scenes. *J. Vis.* **7,** doi:10.1167/7.8.6 (2007).
49. Hosoya, T., Baccus, S. A. & Meister, M. Dynamic predictive coding by the retina. *Nature* **436,** 71–77 (2005).
50. Neitz, J., Carroll, J., Yamauchi, Y., Neitz, M. & Williams, D. R. Color perception is mediated by a plastic neural mechanism that is adjustable in adults. *Neuron* **35,** 783–792 (2002).

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

Acknowledgements This work was supported by the Helen Hay Whitney Foundation (G.D.F.), German Research Foundation (M.G.), National Institutes of Health (NIH) National Research Service Award (NS054519-01) and Chapman Foundation (J.L.G.), Miller Institute for Basic Research in Science (J.S.), Polish Ministry of Science and Higher Education (W.D.), Burroughs Wellcome Fund Career Award at Scientific Interface (A.S.), Engineering and Physical Sciences Research Council (D.E.G.), Royal Society of Edinburgh (K.M.), McKnight Foundation (A.M.L. and E.J.C.), NSF Grant PHY-0750525 (A.M.L.), a Sloan Research Fellowship and NIH Grant EY13150 (E.J.C). We thank C. K. Hulse for technical assistance; M. I. Grivich, D. Petrusca, A. Grillo, P. Grybos, P. Hottowy and S. Kachiguine for technical development; H. Fox, M. Taffe, E. Callaway and K. Osborn for providing access to retinas; S. Barry for machining; F. Rieke and T. Sejnowski for providing comments on the manuscript. We thank the San Diego Supercomputer Center and the National Science Foundation (Cooperative Agreements 05253071 and 0438741) for large-scale data storage.

Author Contributions G.D.F., J.L.G., A.S. and E.J.C. conceived the experiments. G.D.F., J.L.G., A.S., M.G., J.S., L.H.J. and E.J.C. performed the electrophysiological experiments. G.D.F., J.L.G., A.S., M.G., T.A.M., E.J.C. and L.P. analysed the data. A.S., D.E.G., K.M., W.D. and A.M.L. provided and supported the large-scale multielectrode array system. G.D.F. and E.J.C. wrote the manuscript.

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

METHODS

Most procedures are described in Supplementary Methods. Electrophysiological recordings from the retina of macaque monkeys (*Macaca fascicularis* and *Macaca mulatta*), spike identification, stimulus generation and calibration, receptive-field characterization, and cell-type classification were performed using previously described methods[10–12,15]. Imaging of cone outer segments was performed using standard fluorescence microscopy. New methods were developed for cone identification and classification. Here, details are given about analysis of cone-type specificity.

Analysis of cone-type purity. The degree of purity of L- and M-cone input to RGCs was quantified using an index:

$$\frac{\sum_i L_i - \sum_j M_j}{\sum_i |L_i| + \sum_j |M_j|} \tag{1}$$

where L_i and M_j represent the weights on distinct L and M cones in the receptive field. Thus, if all cones sampled were L (M) cones, the index would be 1 (−1) (Fig. 4i). If the weighted input were equal for L and M cones, the index would be zero. This index was computed separately for cones in the receptive-field centre and surround.

To examine purity in populations, the distribution of purity indices was generated for all ON midget cells and all OFF midget cells (Fig. 4j). The width (s.d.) of this distribution was compared to that of a null distribution, which was generated by artificially and randomly permuting cone types in the cone mosaic, recomputing purity indices for all RGCs, and calculating the s.d. of the resulting distribution. This value represents the expectation if cone types were randomly spatially distributed and randomly sampled by RGCs. To statistically compare the permutation to the data (Fig. 4k, error bars), cone mosaics were re-permuted repeatedly and the s.d. across permutations was calculated.

Several control analyses were performed to verify the finding that midget cells showed non-random cone purity (Fig. 4k). First, the analysis was run using a more stringent criterion, and a more permissive criterion, for the identification of cones in the data set, by altering the stopping point used in the automated cone identification algorithm (Supplementary Methods). Second, the analysis was run for each RGC type (for example, ON midget) using cones identified only from the receptive fields of the other RGC types (for example, OFF midget, ON parasol, OFF parasol, small bistratified). Third, analysis was performed using an entirely different procedure to identify cones (Supplementary Methods). These variants did not alter the conclusions. A fourth control analysis indicated that a substantial overestimate (25–50%) of the M- relative to L-cone-input weights to midget cells could artificially indicate purity of the magnitude observed. In the data, estimated M-cone weights were generally smaller than estimated L-cone weights (5 ± 5% across data sets). Assuming that the true M-cone weights are not substantially smaller than the true L-cone weights, the direction and magnitude of the weight difference in the data cannot account for the purity finding. The analysis of binarized weights (Fig. 4o) also indicates that the purity finding cannot be explained by biased estimates of cone weights.

Analysis of cone clumping and effects on purity. To test whether mosaics of L and M cones showed clumping, an index was calculated that summarized the degree of aggregation of each cone type on the spatial scale of the midget-cell receptive field:

$$c = \sum_{i \neq j} \exp\left[-\frac{1}{2}(D_{ij}/r)^2\right] \tag{2}$$

where D_{ij} is the distance between the ith and jth cone, r is the average receptive-field radius of midget cells, and summation is performed over all cone pairs of the same type (either L or M). If the cone mosaics in the data were clumped on the spatial scale of the midget-cell receptive field, then randomly permuting L- and M-cone identities in the data should disrupt this clumping and reduce the index. In four data sets examined, the degree of clumping in the data was within 1 s.d. of the distribution of clumping values obtained with repeated permutations of cone types, whereas in the remaining three data sets the value was 2–3 s.d. of permutations. Similar results were obtained with spatial scales r twice and half as large. Simulations indicated that these results could be explained by systematic errors in L- and M-cone classification of 2–5%. Estimated L- and M-cone classification error rates ranged from 1–6% (mean = 3.4%) across seven preparations. Thus, the weak evidence for clumping was consistent with an origin in small-cone-type classification errors.

Three further statistical tests of clumping were performed as controls. These have the advantage of testing multiple spatial scales. The three tests compared cumulative distributions of three distance measurements from an observed mosaic of L and M cones to values computed from the same mosaic after cone type permutation. The three tests were based on the distribution of inter-cone distances[41], nearest neighbour distances, and distances from cones to an artificial square lattice of points (statistics H, G and F respectively from ref. 51). For these three tests respectively, 1/7, 3/7 and 2/7 of cone mosaics examined showed evidence for clumping of either L or M cones at $P < 0.01$ at some spatial scale. There was no spatial scale that consistently yielded indications of clumping, and only 1/7 data sets indicated clumping in all three tests. Thus, these tests provided at most weak evidence for clumping of L or M cones.

However, even weak clumping could contribute to apparent cone-type purity in the receptive-field centre. To test this possibility, artificial cone mosaics with a clumping index matching the data were generated by first randomly permuting the cone labels of an observed cone mosaic, then swapping the labels of randomly selected L–M cone pairs if the swap increased the index, iterating until the clumping index matched that of the data. Many such artificially clumped cone mosaics were generated, and the purity of midget-cell receptive fields was compared with real and artificial mosaics (Fig. 4m).

Control analysis was performed by shifting the cone mosaic with respect to RGCs instead. This was done by selecting a collection of RGCs, translating the cone lattice with respect to them, and then associating with each RGC the set of cones closest to the locations of the original cones in its original receptive field. Across many randomly selected shifts of the cone mosaic, this manipulation also reduced purity (Supplementary Methods), confirming that the observed purity is not explained purely by clumping in the cone mosaic.

Analysis of biased sampling and weighting. Variations of the purity analysis tested the contribution of biased sampling and biased weighting to the purity of midget-cell receptive-field centres. To test biased sampling, the same analysis described earlier was performed; however, the weights of cones feeding the receptive-field centre of each RGC were first set to one, other weights were set to zero (Fig. 4n). This produced a purity analysis based on the collection of cones feeding the receptive-field centre while ignoring the weights on those cones. To test biased weighting, the weights of cones providing input to each RGC were permuted (Fig. 4o). This kept the collection of cones feeding each RGC constant, but eliminated any relationship between cone type and cone weight.

51. Diggle, P. J. *Statistical Analysis of Spatial Point Patterns* (Academic, 1983).

doi:10.1038/nature09399

Purified human BRCA2 stimulates RAD51-mediated recombination

Ryan B. Jensen[1,2], Aura Carreira[1,2] & Stephen C. Kowalczykowski[1,2]

Mutation of the breast cancer susceptibility gene, *BRCA2*, leads to breast and ovarian cancers. Mechanistic insight into the functions of human BRCA2 has been limited by the difficulty of isolating this large protein (3,418 amino acids). Here we report the purification of full-length BRCA2 and show that it both binds RAD51 and potentiates recombinational DNA repair by promoting assembly of RAD51 onto single-stranded DNA (ssDNA). BRCA2 acts by targeting RAD51 to ssDNA over double-stranded DNA, enabling RAD51 to displace replication protein-A (RPA) from ssDNA and stabilizing RAD51–ssDNA filaments by blocking ATP hydrolysis. BRCA2 does not anneal ssDNA complexed with RPA, implying it does not directly function in repair processes that involve ssDNA annealing. Our findings show that BRCA2 is a key mediator of homologous recombination, and they provide a molecular basis for understanding how this DNA repair process is disrupted by *BRCA2* mutations, which lead to chromosomal instability and cancer.

One of the proposed driving forces behind the tumorigenic process is the onset of genomic instability that, when coupled to repeated rounds of cell division, promotes oncogenesis[1]. A hallmark of human and mouse cells that are mutant for BRCA2 is severe chromosomal instability marked by an accumulation of chromosomal breaks, translocations, exchanges and other abnormal structures[2]. Accordingly, germline mutations in BRCA2 are associated with a highly penetrant incidence of breast and/or ovarian cancer as well as tumours in other tissues and organs[3,4]. BRCA2 possesses eight highly conserved repeated sequences, termed the BRC repeats, and a carboxy-terminal region that were shown to bind RAD51 (refs 5–7). RAD51 has a central role in recombination, assembling onto single-stranded DNA (ssDNA) as a nucleoprotein filament and catalysing the invasion and exchange of homologous DNA sequences[8,9].

At the cellular level, loss of BRCA2 function results in sensitivity to cross-linking agents, a decrease in homology-directed repair of double-stranded DNA breaks (DSBs), and defects in replication and checkpoint control[2,10–12]. BRCA2 is also required for RAD51-induced focus formation after exposure to DNA-damaging agents[13,14]. Prior studies using fragments of BRCA2, fusions of the BRC repeats with the DNA-binding domain (DBD) of BRCA2, and analysis of the *Ustilago maydis* and *Caenorhabditis elegans* orthologues, Brh2 and BRC-2 respectively, have provided a framework for understanding the mediator role that BRCA2 has in DSB repair by RAD51-driven homologous recombination[15–23] (see Supplementary Fig. 1 for a model based on this report and the work referenced above). However, the large size of human BRCA2 (3,418 amino acids), difficulty in driving high-level expression, insufficient solubility, and its propensity to degrade, have precluded isolation of the full-length BRCA2 protein and have hampered fuller understanding of its functions. Here, we describe purification of the full-length protein from human cells, and report its biochemical functions with regard to recombinational DNA repair.

Purified BRCA2 binds RAD51 and DMC1

By using a mammalian expression vector (phCMV1) driven by a CMV promoter and by adding two tandem repeats of the maltose binding protein (designated 2XMBP) to the amino terminus of human BRCA2 (470 kDa including the two MBP tags), we expressed significant amounts of protein that could be purified to near homogeneity (Fig. 1a and Supplementary Fig. 2). The identity of full-length BRCA2 was confirmed by western blotting using an antibody to the C-terminal region of BRCA2 (Fig. 1a, lane 4 and Supplementary Fig. 2b, c), an antibody to the N-terminal MBP tag, and by mass spectrometric analysis. Mass spectrometric analysis of a variable minor band directly below the full length protein confirmed the presence of a truncated BRCA2 species lacking the C terminus. The band near the 50 kDa marker (Fig. 1a, asterisk) was confirmed by mass spectrometry to be β-tubulin. The presence of this contaminant appears not to interfere with any of our *in vitro* studies. We found that 2XMBP–BRCA2 fully complemented *brca2* mutant (VC8) cells (Fig. 1b and Supplementary Fig. 3); therefore, the tag was not removed for the studies reported here. Hereafter, we refer to the N-terminal 2XMBP-tagged version of full length BRCA2 as BRCA2.

Our initial criteria for determining that purified BRCA2 was properly folded and retained biochemical function was to test its ability to bind recombination proteins that were previously reported to interact. We incubated BRCA2 with several purified candidate proteins and used the MBP tag to capture the complexes on amylose beads, and analyse the complexes on SDS–PAGE gels stained with SYPRO Orange, (Fig. 1c, d). As expected from *in vivo* pull-down assays and interaction with a fusion protein construct[5,15,24–26], human RAD51 bound to BRCA2 (Fig. 1c, lane 9). Also in agreement[27], BRCA2 bound DMC1 (Fig. 1d, lane 5), the meiotic counterpart of RAD51. BRCA2 bound to yeast Rad51 (Fig. 1c, lane 8) but, given the high homology between the two orthologues (~67% identical; 83% homologous[28,29]), this is not surprising. BRCA2 did not appreciably bind the *Escherichia coli* homologue, RecA, (Fig. 1c, lane 7) showing that interaction did not extend to the evolutionarily distant bacterial protein. We also did not detect any significant interaction with human RPA (Fig. 1d, lane 6), despite a report in the literature[30], *E. coli* SSB (Fig. 1c, lane 6), or human RAD52 (Fig. 1d, lane 7).

The BRC repeat defines a RAD51 binding motif whose number varies widely amongst different organisms ranging from one BRC repeat in *U. maydis* to 15 repeats in *Trypanosoma brucei*[31]. Human BRCA2 contains eight BRC repeats. Various studies established that

[1]Department of Microbiology, University of California, Davis, California 95616, USA. [2]Department of Molecular & Cellular Biology, University of California, Davis, California 95616, USA.

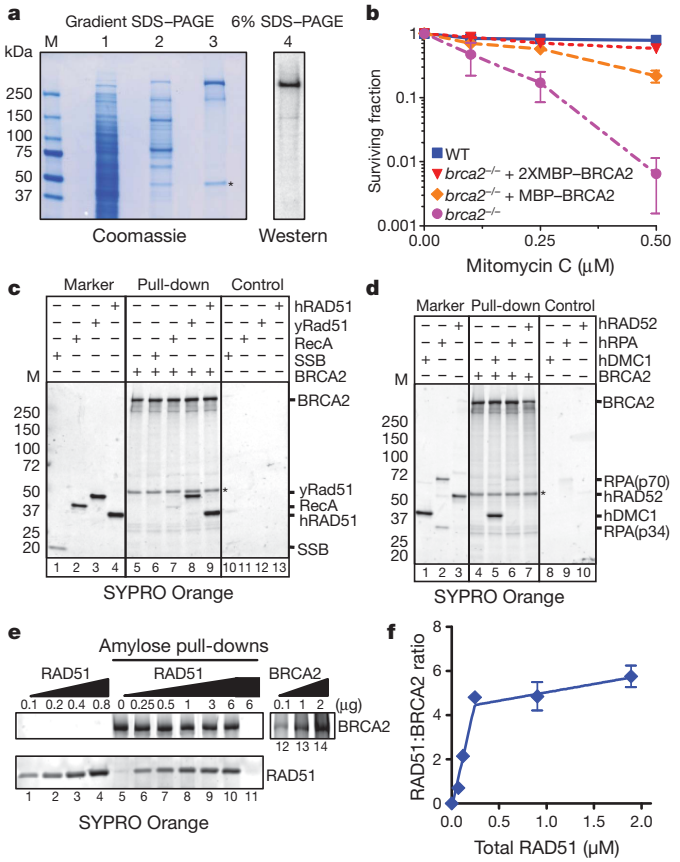


Figure 1 | Protein interactions with purified full-length human BRCA2. **a**, Lane 1, 293T cell lysate; 2, amylose eluate; 3, purified BRCA2; 4, western blot (Ab-2). M, standards. Asterisk, β-tubulin. **b**, Mitomycin C survival. Error bars, s.d. ($n = 3$). **c, d**, Protein pull-downs. 2XMBP–BRCA2 with indicated proteins. Marker, protein input. Control, proteins + amylose resin. **e**, RAD51 titration of BRCA2. Lanes 1–4, RAD51 standards; 5–11, pull-downs; 5, BRCA2 alone; 11, RAD51 alone; 12–14, BRCA2 standards. **f**, Data from **e** fit to segmental linear regression. Error bars, s.d. ($n = 2$).

most of the human BRC repeats in isolation can bind RAD51 (refs 5, 24–26); however, it remained unclear how many binding sites are occupied within the context of the full-length protein and whether the interaction affinities were comparable. Using known concentrations of purified RAD51 (Fig. 1e, lanes 1–4) and recombinant BRCA2 (Fig. 1e, lanes 12–14) as standards, and staining with SYPRO Orange, we quantified the binding. In the absence of BRCA2, RAD51 did not bind non-specifically to the amylose resin (Fig. 1e, lane 11). In the presence of BRCA2 (Fig. 1e, lanes 5–10), the amount of RAD51 bound increased linearly with concentration (indicative of tight binding in the nM range) until about 4.5 (± 0.9) RAD51 molecules were bound per BRCA2 (Fig. 1f); afterward, a weaker binding (in the μM range) was evident. At the maximum RAD51 concentration attainable, approximately six RAD51 proteins were bound to each BRCA2. The binding of RAD51 to BRCA2 was also examined under buffer conditions identical to those used for DNA strand exchange assays, and similar binding characteristics were found (Supplementary Fig. 4). The protein complexes formed between BRCA2 and RAD51 were not dependent on magnesium or calcium ions, nucleotide cofactors or the presence of DNA (data not shown).

BRCA2 prefers to bind ssDNA over dsDNA

The DNA binding domain of BRCA2 contains both oligonucleotide-binding (OB) folds and a tower domain, which engender BRCA2 with potential sites for binding both ssDNA and dsDNA[16]. Indeed, both the carboxy terminus of BRCA2 and a fusion protein containing BRC

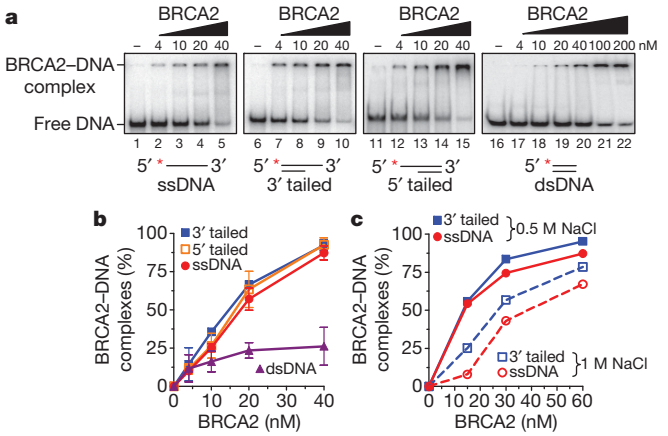


Figure 2 | BRCA2 displays a strong preference for binding tailed and ssDNA substrates over dsDNA. **a**, EMSA. BRCA2 binding ssDNA, 3′ tailed DNA, 5′ tailed DNA and dsDNA. **b**, Quantification. **c**, EMSA in 0.5 M or 1 M NaCl. Error bars, s.d. ($n = 3$).

repeats 3 and 4 linked to the DNA binding domain of BRCA2 bind both ssDNA and dsDNA[15,16]. By using electrophoretic mobility shift assays (EMSA), we tested the ability of BRCA2 to bind ssDNA, dsDNA and dsDNA with an ssDNA tail (3′ tailed or 5′ tailed). BRCA2 bound to all of these substrates; however, those containing ssDNA were strongly preferred over dsDNA (Fig. 2a, b). These results are consistent with previous reports on the DNA binding domain of BRCA2 (ref. 15). A slight preference for tailed DNA over ssDNA was revealed at higher salt concentrations (Fig. 2c); however, the difference was modest. The binding specificity for the various DNA substrates was unaltered when BRCA2 was incubated with RAD51 before binding to the DNA (Supplementary Fig. 5).

BRCA2 stimulates DNA strand exchange by RAD51

An essential function of RAD51 in recombinational DNA repair is its capacity to homologously pair and exchange DNA strands. To promote this process, RAD51 must assemble onto the 3′ ssDNA tails generated by resection of DNA breaks. To mimic the DNA intermediate generated after DSB resection *in vivo*, we used a tailed DNA substrate created by annealing a 42-mer oligonucleotide to a 167-mer to create a 42 base pair (bp) dsDNA region followed by a 125 nucleotide 3′ ssDNA overhang (Fig. 3b; termed 3′ tailed DNA). To

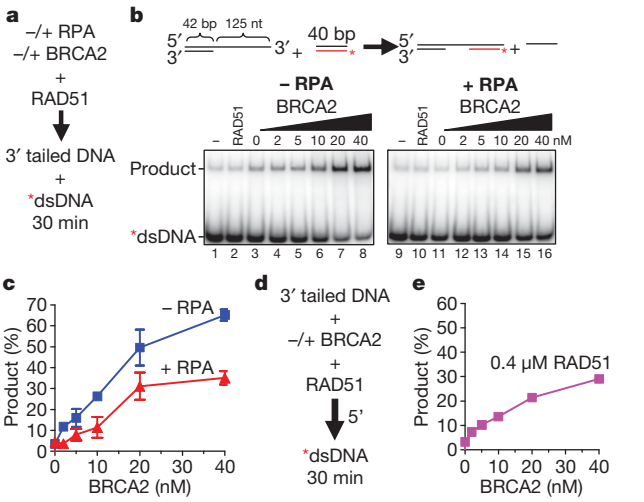


Figure 3 | BRCA2 stimulates DNA strand exchange promoted by RAD51. **a**, DNA strand exchange reaction protocol for **b** and **c**. **b**, Reaction in the absence (left) or presence (right) of RPA. **c**, Quantification of **b**. Error bars, s.d. ($n = 3$). **d**, DNA strand exchange protocol for **e** in the absence of RPA. **e**, Quantification of autoradiogram in Supplementary Fig. 8b.

validate the DNA substrate, we conducted DNA strand exchange assays as a function of RAD51 protein concentration using an optimized *in vitro* DNA strand exchange protocol (Supplementary Fig. 6a, c); optimal product formation was at a 1:3 (RAD51:nucleotide) ratio, consistent with the DNA binding stoichiometry of RAD51 (ref. 32). However, *in vivo*, filament assembly conditions are not optimal: RAD51 must compete with RPA for binding to the ssDNA[33,34]. Furthermore, RAD51 can bind to both ssDNA and dsDNA, but the binding to dsDNA is not productive and, in fact, blocks DNA strand exchange[9,34]. Thus, DNA strand exchange can be stimulated in at least two mechanistically distinct ways.

Initially, to determine whether BRCA2 affects DNA strand exchange, reactions were performed by using an optimal amount of RAD51 but introducing it concurrently to a mixture of 3′ tailed DNA and dsDNA. Indeed, as demonstrated in Fig. 3b, c, when RAD51 is permitted to assemble on both ssDNA and dsDNA, DNA strand exchange is reduced to background levels (Fig. 3b, lanes 2 and 10). However, if BRCA2 is incubated with RAD51 before mixing with DNA substrates, this inhibition is alleviated in a concentration-dependent manner (Fig. 3b, lanes 3–8; Fig. 3c), indicating that BRCA2 directs RAD51 to ssDNA or limits binding to dsDNA, or both. In the presence of RPA, stimulation by BRCA2 is maintained, although the magnitude is reduced (Fig. 3b, lanes 11–16; Fig. 3c). To confirm that the product did not result from 'melting' of the donor duplex DNA and spontaneous annealing during the de-proteinization step[35], we performed the same reaction with tenfold excess unlabelled oligonucleotide complementary to the labelled pairing strand in the stop-mix, and the results were unchanged (Supplementary Fig. 7a, lane 9). The stimulation was ATP-dependent (Supplementary Fig. 7a, lanes 11 and 12) and did not occur with a heterologous template (Supplementary Fig. 7a, lane 10); furthermore, BRCA2 alone was unable to promote DNA strand exchange (Supplementary Fig. 7a, lane 3). These results support a role for BRCA2 in targeting RAD51 to ssDNA, limiting assembly onto the dsDNA partner, or both.

BRCA2 limits assembly of RAD51 on dsDNA

To determine whether BRCA2 slowed or prevented assembly on the dsDNA partner of DNA strand exchange, reactions were performed using a concentration of RAD51 (0.4 μM) sufficient to saturate both the ssDNA and dsDNA present. At such a concentration, DNA strand exchange is inhibited due to binding of excess RAD51 to the dsDNA target[32,34] (see Supplementary Fig. 6a, lane 5 and Supplementary Fig. 6c). To optimize filament formation on the ssDNA, BRCA2 and RAD51 were incubated with the 3′ tailed ssDNA first, and then the dsDNA was added to initiate the reaction (Fig. 3d); however, the excess free RAD51 binds the dsDNA partner and inhibits the reaction (Supplementary Fig. 8b, lane 1). To eliminate complications from competition with RPA, these reactions were done in the absence of RPA. BRCA2 stimulated DNA strand exchange in a concentration-dependent manner (Fig. 3e and Supplementary Fig. 8, lanes 2–6). To provide more direct evidence that BRCA2 prevents nucleation on the dsDNA and targets RAD51 to the 3′ tailed DNA, we preincubated BRCA2 and RAD51 and then analysed RAD51–DNA complex formation by either EMSA or the partitioning of RAD51 onto biotinylated 3′ tailed DNA in the presence of excess dsDNA (Supplementary Fig. 9): BRCA2 prevented binding of RAD51 to dsDNA (Supplementary Fig. 9b) and favoured RAD51 binding to the ssDNA (Supplementary Fig. 9f). Taken together with the results of the previous section, these data support the idea that BRCA2 recruits RAD51 to ssDNA, likely by virtue of its affinity for ssDNA, and inhibits assembly of RAD51 onto dsDNA.

BRCA2 recruits RAD51 to ssDNA complexed with RPA

Previous work using a human BRCA2 polypeptide fusion and *U. maydis* Brh2 demonstrated the ability of those proteins to promote RAD51 filament formation onto RPA-coated ssDNA[15,17]. To determine whether the full-length human BRCA2 protein provides a similar avenue for stimulation, we next performed DNA strand exchange assays using an optimal amount of RAD51 but, rather than permitting filament formation on naked ssDNA, the ssDNA was first complexed with RPA (Fig. 4a). RAD51 was subsequently introduced in the presence or absence of BRCA2, and finally, the labelled duplex DNA was added to start the reaction. As expected, incubation of the ssDNA with increasing concentrations of RPA before addition of RAD51 severely impaired DNA strand exchange (Supplementary Fig. 6b, lanes 4–7). As shown in Fig. 4b, c, increasing amounts of BRCA2 stimulated DNA strand exchange as much as 20-fold, indicating that BRCA2 accelerates formation of the RAD51 nucleoprotein filament at the presynaptic stage of recombination and alleviates the inhibition posed by RPA. Stimulation by BRCA2 occurred at concentrations as low as 2 nM (Fig. 4b, c), and that were sub-stoichiometric relative to the RAD51 concentrations (approximately 100-fold less than RAD51).

To confirm a presynaptic role for BRCA2, we also performed kinetic analyses of the DNA strand exchange reactions by varying the time that BRCA2 was incubated with RAD51 and the RPA–ssDNA complex, either before (Supplementary Fig. 10a–c) or after (Supplementary Fig. 10d–f) addition of homologous dsDNA. BRCA2 imparts a significant stimulation (~20-fold) of DNA strand exchange in as little as one minute after incubation with RAD51 and the RPA–ssDNA complex (Supplementary Fig. 10c); only after 60 min does the yield without BRCA2 match that seen with BRCA2 at 1 min.

In contrast to *U. maydis* Brh2 (ref. 17) and the *E. coli* analogue, RecFOR[36], which show a strict specificity to act at a 3′ ssDNA overhang, stimulation of DNA strand exchange by human BRCA2 was the same for 3′ versus 5′ tailed DNA (Fig. 4b, compare lanes 3–8 and 11–16, and 4c). This result was consistent with our observation above (Fig. 2a, b) that the binding affinity of BRCA2 for 3′ tailed DNA and 5′ tailed DNA is the same. A bias was also not apparent at sub-stoichiometric concentrations of RAD51 (data not shown) where reduced filament occupancy revealed junction specificity for Brh2 (ref. 17). However, we did observe a consistent twofold preference for both 3′ and 5′ tailed DNA substrates over ssDNA, indicating that stimulation by BRCA2 is modestly greater for a DNA substrate containing a junction of ssDNA with dsDNA (Fig. 4c), also consistent with our binding results. These findings demonstrate that BRCA2 possesses the capacity to broadly stimulate RAD51 assembly onto ssDNA with or without a dsDNA junction, and to uniquely promote the assembly onto either 3′ or 5′ ssDNA tails.

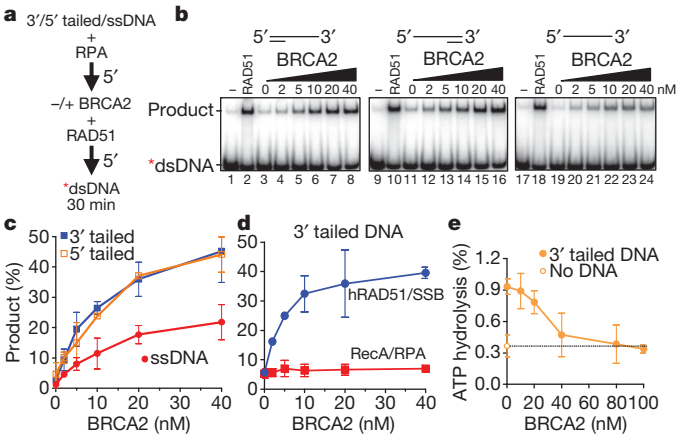


Figure 4 | BRCA2 stimulates RAD51-mediated DNA strand exchange by promoting stable RAD51–ssDNA filament formation, overcoming inhibition by RPA. **a**, DNA strand exchange reaction protocol for **b** and **c**. **b**, Autoradiograms of 3′ tailed, 5′ tailed and ssDNA substrates. **c**, Quantification of **b**. **d**, Assays as in **a**, except RecA replaced RAD51 and SSB replaced RPA. **e**, Inhibition of RAD51 ssDNA-dependent ATP hydrolysis by BRCA2. Error bars, s.d. ($n = 3$).

BRCA2 did not stimulate RecA (Fig. 4d and Supplementary Fig. 11b, left panel), consistent with the failure to pull-down RecA. When RPA was replaced by SSB, BRCA2 could still stimulate DNA strand exchange by RAD51 (Fig. 4d and Supplementary Fig. 11b, right panel), implying that neither BRCA2 nor RAD51 need to interact directly with the ssDNA-binding proteins. This idea is further bolstered by the lack of interaction between BRCA2 and either SSB or RPA in the pull-down assays (Fig. 1c, d). This finding is consistent with the behaviour of a human BRCA2 polypeptide fusion[15] but distinct from that of the bacterial RecFOR and fungal Brh2, which require their cognate ssDNA-binding proteins for stimulation[17,36]. Thus, it seems that direct interactions between human BRCA2, RAD51 and DNA are sufficient to stimulate the ability of RAD51 to gain access to the RPA- or SSB-coated ssDNA and to then displace them as the ensuing nucleoprotein filament is formed and extended.

Our previous work on the BRC repeats demonstrated that they stabilize ssDNA-RAD51 complexes by blocking the ATPase activity of RAD51 (ref. 21). To gain insight into the mechanism by which BRCA2 stimulates presynaptic complex formation, we measured its effect on the ATPase activity of RAD51. BRCA2 inhibited the ssDNA-dependent ATPase activity of RAD51 in a concentration dependent manner to the level seen in the absence of DNA (Fig. 4e). These results suggest that the same mechanism for RAD51 nucleoprotein filament stabilization ascribed to the BRC repeats applies to BRCA2[21]: namely, full length BRCA2 stabilizes the RAD51 bound to the ssDNA substrate by down-regulating its ATPase activity, a hydrolysis activity which inactivates and turns over the RAD51 protein; however, quantitatively, this control through down-regulation is achieved at much lower concentrations of BRCA2 than the BRC repeats.

BRCA2 cannot anneal RPA–ssDNA complexes

The capability of BRCA2 to accelerate displacement of RPA by RAD51 without the need for a ssDNA/dsDNA junction has some parallels to yeast Rad52 (yRad52)[37–39], but human RAD52 lacks this ability to stimulate RPA replacement[40]. However, because yeast lacks a known BRCA2 homologue whereas mammals possess both BRCA2 and RAD52, it is possible that evolutionary changes separated the functions of yRad52 into several mammalian proteins. Another important role of the multi-functional yRad52, and bacterial RecO, is the annealing of complementary ssDNA that is bound by the cognate ssDNA-binding protein[41,42]; both the *U. maydis* and *C. elegans* BRCA2 orthologues can anneal ssDNA in the presence of RPA[23,43]. Consequently, we investigated whether BRCA2 or human RAD52 possess a similar capacity. Complementary ssDNA substrates, with or without saturating human RPA, were incubated with BRCA2 or RAD52 and then mixed (Fig. 5a). Figure 5b shows that in the absence of proteins, spontaneous annealing occurred over time (lanes 2–5; quantification in Fig. 5c). BRCA2 marginally increased (lanes 6–9), and RAD52 clearly increased the rate of annealing (lanes 10–13). When RPA was added (lanes 15–18), spontaneous annealing was completely blocked. Unlike the *U. maydis* and *C. elegans* orthologues[23,43], human BRCA2 was unable to overcome this inhibition (lanes 19–22), but RAD52 readily annealed the RPA–ssDNA complexes (lanes 23–26). In contrast, when substituted for BRCA2, neither human RAD52 nor yeast Rad52 stimulated DNA strand exchange by RAD51 when the ssDNA was complexed with RPA (Supplementary Fig. 12). Taken together, these data show that BRCA2 and human RAD52 have assumed divergent roles in mammalian cells. BRCA2 has taken on the functions that stimulate joint molecule formation and DNA strand exchange, whereas RAD52 provides the ssDNA-annealing functions of recombination[44].

Discussion

Our results reveal the biochemical functions of full-length human BRCA2, and they establish that BRCA2 augments the functions of RAD51 that are essential for recombinational repair of DNA breaks

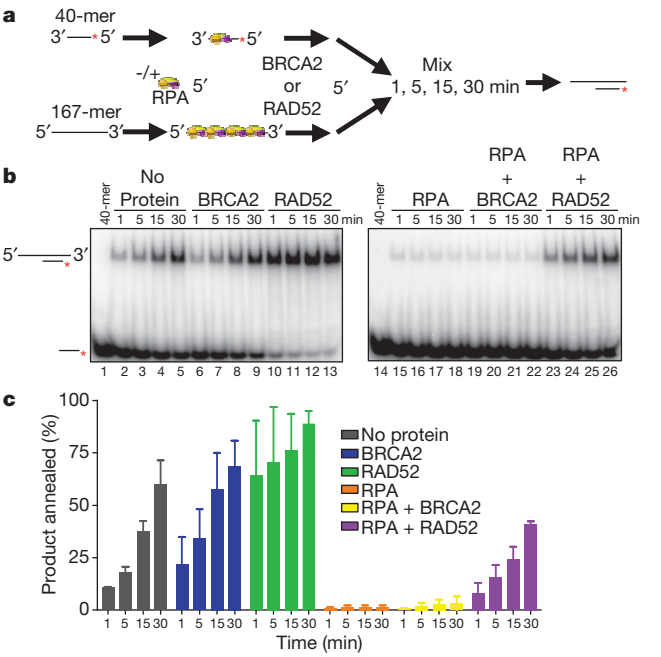


Figure 5 | BRCA2 does not anneal ssDNA complexed with RPA. **a**, Schematic of DNA annealing assays. **b**, Autoradiogram: absence of RPA, plus indicated protein (left); presence of RPA first, plus indicated protein (right). Lanes 1 and 14, radio-labelled 40-mer. **c**, Quantification of **b**. Error bars, s.d. ($n = 3$).

(Supplementary Fig. 1). Stimulation by BRCA2 is a consequence of several mutually reinforcing effects; BRCA2 (1) enforces binding of RAD51 to ssDNA, (2) accelerates the rate of RPA-displacement from ssDNA by RAD51, (3) inhibits the ATPase activity of RAD51, and (4) limits binding to dsDNA. BRCA2 focuses the assembly of RAD51 onto ssDNA and facilitates the RAD51-mediated displacement of RPA from ssDNA, which is a key regulatory step of DNA pairing. In support of this general concept, the promotion of RAD51 filament formation onto RPA-coated ssDNA was also demonstrated using a different full-length BRCA2 protein expression construct and preparation[45]. By enabling formation of the presynaptic complex, BRCA2 permits progression to the subsequent DNA pairing phase of recombinational DNA repair. Furthermore, by inhibiting the ssDNA-dependent ATP hydrolysis of RAD51, BRCA2 preserves the active and most stable form of RAD51, the ATP–RAD51–ssDNA complex[21,22]. Because the rate-limiting step in RAD51 nucleoprotein filament assembly is nucleation of the first several monomers of the filament[46–48], BRCA2 can act catalytically to stabilize a nucleus by blocking RAD51 self-inactivation and dissociation via its ATPase activity. If the RAD51 molecules bound to BRCA2 do indeed comprise the nucleus, then BRCA2 can stabilize a nascent filament of up to 4–6 RAD51 molecules. Inhibition of ATPase activity was also observed for the *C. elegans* proteins (Rad51 and BRC-2)[23], but not for the *U. maydis* orthologues (Rad51 and Brh2)[17] or the *E. coli* analogues (RecA and RecFOR)[36], indicating that this mechanism of stimulating RAD51 function is a late adaption of multicellular organisms. In addition, our results show that BRCA2 prevents or slows the assembly of RAD51 onto duplex DNA, an aspect of RAD51 filament assembly that impairs recombination reactions. Based on previous single-molecule studies with the BRC repeats[21], we propose that interaction with full-length BRCA2 slows nucleation of RAD51 onto dsDNA.

An unanticipated feature of human BRCA2 is its ability to bind and stimulate RAD51-mediated DNA strand exchange at regions of ssDNA as well as at ssDNA/dsDNA junctions of either polarity. This capability permits the BRCA2-facilitated loading of RAD51 in both the 5′→3′ and 3′→5′ directions. Although most models for BRCA2 function have focused on DSB repair, we note that these

characteristics are also consistent and suggestive of a role for BRCA2 in the recombinational repair of DNA gaps that occur during DNA replication due to damage in the template (Supplementary Fig. 1b). Given that BRCA2 facilitates growth of the filament in either polarity from internal ssDNA regions or from either junction, BRCA2 could readily contribute to daughter strand DNA gap repair as well as DSB repair. It is notable that the lack of polarity in human BRCA2 function differs from analogues such as RecFOR and Brh2, which load RecA and *U. maydis* Rad51, respectively, specifically onto the 3′ overhanging ssDNA[17,36]. The absence of a bias for pairing 3′-ends by human RAD51 may reflect its underlying intrinsic capability to assemble in both directions[49]. But because resection of a DSB *in vivo* produces 3′ tailed ssDNA, there is no compelling need for polarity enforcement by BRCA2.

Another key difference between the BRCA2 orthologues of *U. maydis* and *C. elegans*[23,43] is that human BRCA2 is unable to anneal ssDNA complexed with RPA, the physiological intermediate of recombination. However, human RAD52 does manifest this capability, implying that in mammalian cells this function is assumed by RAD52. Yet RAD52 cannot stimulate the assembly of RAD51 onto ssDNA complexed with RPA[40] (Supplementary Fig. 12), but BRCA2 can. This behaviour of RAD52 is distinct from yRad52 which displays both of these species-specific capabilities[37–39,41], explaining the essential role of yRad52 in all recombinational DNA repair processes. Thus, in vertebrates, these functions have separated: BRCA2 targets RAD51 to ssDNA to mediate DNA strand invasion into a duplex donor to produce joint molecules by a distinctive mechanism, whereas RAD52 anneals RPA–ssDNA complexes in steps or pathways of recombinational repair that could include second-end capture in DSB repair, single-strand annealing (SSA), and synthesis-dependent strand annealing. In support of this idea, error-prone repair by SSA is increased in cells lacking functional BRCA2, whereas *RAD52* complementation in *RAD52*$^{-/-}$ mouse ES cells augments the SSA pathway[44,50]. Both *in vivo* and *in vitro* studies have clarified the role of BRCA2 in catalysing the delivery of RAD51 to sites of DNA damage. Our work has confirmed and extended prior expectations that the intact human BRCA2 protein mediates the rapid and ordered assembly of the RAD51 protein onto ssDNA, and helps to explain why cells lacking functional BRCA2 would be severely impaired for formation of this critical intermediate in recombinational repair. As a consequence of BRCA2 loss, DNA break repair mediated through template-directed repair from homologous sequences within an intact homologue or sister chromatid would be disrupted, leading to error prone repair and potential chromosomal instability. The ability to purify full-length human BRCA2, a protein directly responsible for genetically predisposing individuals to substantially high risks for cancer, should open a whole new venue for understanding this very large and complex protein.

METHODS SUMMARY

Expression and purification of full-length BRCA2. Human BRCA2 cDNA was cloned into phCMV1 with two repeats of MBP at the N terminus. 293TD cells were transfected and harvested 31 h post-transfection. Extracts were batch bound to amylose resin. The protein was eluted with 10 mM maltose, loaded onto HiTrap Q, and step eluted at 450 mM NaCl.

DNA substrates. Oligonucleotides were PAGE-purified. The 3′ tailed, 5′ tailed and dsDNA were annealed at 1:1 molar ratio, and ^{32}P-labelled at the 5′-end.

DNA strand exchange assays. Unless otherwise indicated, reactions were at 37 °C for 30 min with RAD51 (0.22 μM); RPA (0.1 μM); 3′ tailed, 5′ tailed, or ssDNA (4 nM molecules) and dsDNA (4 nM molecules). Reactions were terminated with proteinase K-SDS, and analysed by electrophoresis (6% polyacrylamide) and phosphor-imaging.

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

Received 31 May; accepted 11 August 2010.
Published online 22 August 2010.

1. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396,** 643–649 (1998).
2. Yu, V. P. *et al.* Gross chromosomal rearrangements and genetic exchange between nonhomologous chromosomes following BRCA2 inactivation. *Genes Dev.* **14,** 1400–1406 (2000).
3. Wooster, R. *et al.* Identification of the breast cancer susceptibility gene *BRCA2*. *Nature* **378,** 789–792 (1995).
4. Phelan, C. M. *et al.* Mutation analysis of the *BRCA2* gene in 49 site-specific breast cancer families. *Nature Genet.* **13,** 120–122 (1996).
5. Wong, A. K. C., Pero, R., Ormonde, P. A., Tavtigian, S. V. & Bartel, P. L. RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene *brca2*. *J. Biol. Chem.* **272,** 31941–31944 (1997).
6. Bignell, G., Micklem, G., Stratton, M. R., Ashworth, A. & Wooster, R. The BRC repeats are conserved in mammalian BRCA2 proteins. *Hum. Mol. Genet.* **6,** 53–58 (1997).
7. Esashi, F. *et al.* CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. *Nature* **434,** 598–604 (2005).
8. Bianco, P. R., Tracy, R. B. & Kowalczykowski, S. C. DNA strand exchange proteins: a biochemical and physical comparison. *Front. Biosci.* **3,** D570–D603 (1998).
9. Baumann, P., Benson, F. E. & West, S. C. Human Rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions *in vitro*. *Cell* **87,** 757–766 (1996).
10. Moynahan, M. E., Pierce, A. J. & Jasin, M. BRCA2 is required for homology-directed repair of chromosomal breaks. *Mol. Cell* **7,** 263–272 (2001).
11. Chen, C. F., Chen, P. L., Zhong, Q., Sharp, Z. D. & Lee, W. H. Expression of BRC repeats in breast cancer cells disrupts the BRCA2-Rad51 complex and leads to radiation hypersensitivity and loss of G_2/M checkpoint control. *J. Biol. Chem.* **274,** 32931–32935 (1999).
12. Lomonosov, M., Anand, S., Sangrithi, M., Davies, R. & Venkitaraman, A. R. Stabilization of stalled DNA replication forks by the BRCA2 breast cancer susceptibility protein. *Genes Dev.* **17,** 3017–3022 (2003).
13. Yuan, S. S. *et al.* BRCA2 is required for ionizing radiation-induced assembly of Rad51 complex *in vivo*. *Cancer Res.* **59,** 3547–3551 (1999).
14. Godthelp, B. C., Artwert, F., Joenje, H. & Zdzienicka, M. Z. Impaired DNA damage-induced nuclear Rad51 foci formation uniquely characterizes Fanconi anemia group D1. *Oncogene* **21,** 5002–5005 (2002).
15. San Filippo, J. *et al.* Recombination mediator and Rad51 targeting activities of a human BRCA2 polypeptide. *J. Biol. Chem.* **281,** 11649–11657 (2006).
16. Yang, H. *et al.* BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science* **297,** 1837–1848 (2002).
17. Yang, H., Li, Q., Fan, J., Holloman, W. K. & Pavletich, N. P. The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA-ssDNA junction. *Nature* **433,** 653–657 (2005).
18. Davies, A. A. *et al.* Role of BRCA2 in control of the RAD51 recombination and DNA repair protein. *Mol. Cell* **7,** 273–282 (2001).
19. Esashi, F., Galkin, V. E., Yu, X., Egelman, E. H. & West, S. C. Stabilization of RAD51 nucleoprotein filaments by the C-terminal region of BRCA2. *Nature Struct. Mol. Biol.* **14,** 468–474 (2007).
20. Saeki, H. *et al.* Suppression of the DNA repair defects of BRCA2-deficient cells with heterologous protein fusions. *Proc. Natl Acad. Sci. USA* **103,** 8768–8773 (2006).
21. Carreira, A. *et al.* The BRC repeats of BRCA2 modulate the DNA-binding selectivity of RAD51. *Cell* **136,** 1032–1043 (2009).
22. Carreira, A. & Kowalczykowski, S. C. BRCA2: shining light on the regulation of DNA-binding selectivity by RAD51. *Cell Cycle* **8,** 3445–3447 (2009).
23. Petalcorin, M. I., Sandall, J., Wigley, D. B. & Boulton, S. J. CeBRC-2 stimulates D-loop formation by RAD-51 and promotes DNA single-strand annealing. *J. Mol. Biol.* **361,** 231–242 (2006).
24. Marmorstein, L. Y., Ouchi, T. & Aaronson, S. A. The *BRCA2* gene product functionally interacts with p53 and RAD51. *Proc. Natl Acad. Sci. USA* **95,** 13869–13874 (1998).
25. Sharan, S. K. *et al.* Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking *Brca2*. *Nature* **386,** 804–810 (1997).
26. Chen, P. L. *et al.* The BRC repeats in BRCA2 are critical for RAD51 binding and resistance to methyl methanesulfonate treatment. *Proc. Natl Acad. Sci. USA* **95,** 5287–5292 (1998).
27. Thorslund, T., Esashi, F. & West, S. C. Interactions between human BRCA2 protein and the meiosis-specific recombinase DMC1. *EMBO J.* **26,** 2915–2922 (2007).
28. Shinohara, A. *et al.* Cloning of human, mouse and fission yeast recombination genes homologous to *RAD51* and *recA*. *Nature Genet.* **4,** 239–243 (1993).
29. Yoshimura, Y., Morita, T., Yamamoto, A. & Matsushiro, A. Cloning and sequence of the human RecA-like gene cDNA. *Nucleic Acids Res.* **21,** 1665 (1993).
30. Wong, J. M., Ionescu, D. & Ingles, C. J. Interaction between BRCA2 and replication protein A is compromised by a cancer-predisposing mutation in BRCA2. *Oncogene* **22,** 28–33 (2003).
31. Hartley, C. L. & McCulloch, R. *Trypanosoma brucei* BRCA2 acts in antigenic variation and has undergone a recent expansion in BRC repeat number that is important during homologous recombination. *Mol. Microbiol.* **68,** 1237–1251 (2008).
32. Benson, F. E., Stasiak, A. & West, S. C. Purification and characterization of the human Rad51 protein, an analogue of *E. coli* RecA. *EMBO J.* **13,** 5764–5771 (1994).
33. Sugiyama, T., Zaitseva, E. M. & Kowalczykowski, S. C. A single-stranded DNA-binding protein is needed for efficient presynaptic complex formation by the *Saccharomyces cerevisiae* Rad51 protein. *J. Biol. Chem.* **272,** 7940–7945 (1997).
34. Sigurdsson, S., Trujillo, K., Song, B., Stratton, S. & Sung, P. Basis for avid homologous DNA strand exchange by human Rad51 and RPA. *J. Biol. Chem.* **276,** 8798–8806 (2001).
35. Lio, Y. C., Mazin, A. V., Kowalczykowski, S. C. & Chen, D. J. Complex formation by the human Rad51B and Rad51C DNA repair proteins and their activities *in vitro*. *J. Biol. Chem.* **278,** 2469–2478 (2003).

36. Morimatsu, K. & Kowalczykowski, S. C. RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. *Mol. Cell* **11,** 1337–1347 (2003).
37. New, J. H., Sugiyama, T., Zaitseva, E. & Kowalczykowski, S. C. Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A. *Nature* **391,** 407–410 (1998).
38. Shinohara, A. & Ogawa, T. Stimulation by Rad52 of yeast Rad51-mediated recombination. *Nature* **391,** 404–407 (1998).
39. Sung, P. Function of yeast Rad52 protein as a mediator between replication protein A and the Rad51 recombinase. *J. Biol. Chem.* **272,** 28194–28197 (1997).
40. McIlwraith, M. J. *et al.* Reconstitution of the strand invasion step of double-strand break repair using human Rad51 Rad52 and RPA proteins. *J. Mol. Biol.* **304,** 151–164 (2000).
41. Sugiyama, T., New, J. H. & Kowalczykowski, S. C. DNA annealing by Rad52 protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA. *Proc. Natl Acad. Sci. USA* **95,** 6049–6054 (1998).
42. Kantake, N., Madiraju, M. V., Sugiyama, T. & Kowalczykowski, S. C. *Escherichia coli* RecO protein anneals ssDNA complexed with its cognate ssDNA-binding protein: a common step in genetic recombination. *Proc. Natl Acad. Sci. USA* **99,** 15327–15332 (2002).
43. Mazloum, N., Zhou, Q. & Holloman, W. K. DNA binding, annealing, and strand exchange activities of Brh2 protein from *Ustilago maydis*. *Biochemistry* **46,** 7163–7173 (2007).
44. Stark, J. M., Pierce, A. J., Oh, J., Pastink, A. & Jasin, M. Genetic steps of mammalian homologous repair with distinct mutagenic consequences. *Mol. Cell. Biol.* **24,** 9305–9316 (2004).
45. Liu, J., Doty, T., Gibson, B. & Heyer, W.-D. Human BRCA2 protein promotes RAD51 filament formation on RPA-covered ssDNA. *Nature Struct. Mol. Biol.* (in the press).
46. Hilario, J., Amitani, I., Baskin, R. J. & Kowalczykowski, S. C. Direct imaging of human Rad51 nucleoprotein dynamics on individual DNA molecules. *Proc. Natl Acad. Sci. USA* **106,** 361–368 (2009).
47. van der Heijden, T. *et al.* Real-time assembly and disassembly of human RAD51 filaments on individual DNA molecules. *Nucleic Acids Res.* **35,** 5646–5657 (2007).
48. Miné, J. *et al.* Real-time measurements of the nucleation, growth and dissociation of single Rad51-DNA nucleoprotein filaments. *Nucleic Acids Res.* **35,** 7171–7187 (2007).
49. Baumann, P. & West, S. C. Heteroduplex formation by human Rad51 protein: effects of DNA end-structure, hRP-A and hRad52. *J. Mol. Biol.* **291,** 363–374 (1999).
50. Tutt, A. *et al.* Mutation in *Brca2* stimulates error-prone homology-directed repair of DNA double-strand breaks occurring between repeated sequences. *EMBO J.* **20,** 4704–4716 (2001).

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

Acknowledgements We thank D. Shin for BRCA2 cDNA clones; M. Zdzienicka for VC8 cells; A. Mazin for RAD52; A. Nimonkar for DMC1; W. Heyer and Kowalczykowski laboratory for comments. Supported by grants from NIH (NIH GM 62653) and DOD-Breast Cancer Research Program (BC085223) to S.C.K., American Cancer Society Postdoctoral Fellowship to R.B.J. (PF-05-225-01-GMC), and Postdoctoral Fellowship from Ministerio de Educación y Ciencia (Spain) to A.C.; R.B.J. acknowledges financial support from C. Hornung, who funded this work through the American Cancer Society and who passed away last year.

Author Contributions R.B.J. and S.C.K. conceived the general ideas for this study. R.B.J., A.C. and S.C.K. planned experiments and interpreted data; R.B.J. and A.C. performed the experiments. R.B.J. and S.C.K. wrote the manuscript and all authors provided editorial input.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.C.K. (sckowalczykowski@ucdavis.edu).

METHODS

Expression and purification of full length BRCA2. The full-length cDNA (10.3 kb) of human BRCA2 was cloned into phCMV1 (Genlantis) along with two tandem repeats of the maltose binding protein (MBP) tag located at the N terminus of BRCA2. All cloning steps were sequence-verified (MCLab) and the final construct was verified by using 15 primers spanning 700 bp regions of the full length BRCA2 cDNA. A PreScission Protease (GE Life Sciences) site was engineered in between the second MBP sequence and the start of the BRCA2 ORF such that both MBP tags could be cleaved by incubation with the PreScission Protease enzyme.

To express this construct, human 293TD cells (a gift from R. Litman) were transiently transfected using TurboFect (Fermentas) and cells were collected 31 h post-transfection. Typically, twenty 15-cm plates containing 70% confluent 293T cells were used for purification. Cells were re-fed with fresh media, Dulbecco's modified Eagle medium + 10% FBS (Invitrogen), before transfection and 16 h post-transfection. Cells were collected in buffer H: 50 mM HEPES (pH 7.5), 250 mM NaCl, 5 mM EDTA and 1 mM dithiothreitol (DTT) with the addition of 1% Igepal CA-630, 3 mM $MgCl_2$, 1 mM ATP, 1 mM PMSF and Protease Inhibitor Cocktail (Roche). The cell suspension was rotated for 20 min and then spun down at 10,000*g* for 15 min in a Sorval centrifuge (Oakridge tubes) and the supernatant was incubated overnight with 1 ml of amylose resin (New England Biolabs) per 50 ml of cell lysate (washed extensively with buffer H before addition to the supernatant). The amylose resin was then spun down at 2,000*g* in a swinging bucket rotor (JS 5.3 Beckman), washed once with lysis buffer, and then poured into a disposable plastic column (Pierce) and washed extensively with buffer H. The protein was then eluted with 10 mM maltose in buffer HG: 50 mM HEPES (pH 7.5), 250 mM NaCl, 0.5 mM EDTA, 10% glycerol and 1 mM DTT. These fractions were then pooled and loaded onto a HiTrap Q (GE Life Sciences), washed with buffer HG, and eluted with buffer HG containing 450 mM NaCl (final storage buffer for BRCA2). The full-length BRCA2 protein was verified by western blot with antibodies to both the C terminus of BRCA2 (Ab-2, EMD) and to the N-terminal MBP tag (anti-MBP, Zymed). The concentration of 2XMBP–BRCA2 was determined using an extinction coefficient at 280 nm of 365,160 M^{-1} cm^{-1}. The final concentration was adjusted by subtracting the contributions from contaminants, β-tubulin (which ranged from ~10–25% in multiple preparations) and truncated BRCA2 polypeptide (see Fig. 1a, which ranged from ~5–15%), based on SYPRO Orange quantification. The proteins in the Coomassie-stained gel shown in Fig. 1a comprised full-length protein (~85%), truncated BRCA2 (~5%) and tubulin (~10%). Some preparations contained a trace amount of HSP70 as determined by mass spectrometry; however, this contaminant was minimized by the inclusion of 3 mM $MgCl_2$ and 1 mM ATP in the lysis buffer. We were unable to detect DSS1 in our preparation; based on comparison to known amounts of purified human DSS1 (a gift from W. Heyer), we estimate by Coomassie quantification that the amount of endogenous DSS1 bound to our purified BRCA2 protein is less than 2% (mole DSS1/mole BRCA2). Typical purification yields from twenty 15-cm plates ranged from 50–100 μg. Contaminant bands were cut out as gel slices from Coomassie-stained gradient (4–15%) SDS–polyacrylamide gels and analysed by mass spectrometry (UC Davis Proteomics Core Facility).

Immunodetection of BRCA2. Cell lysates or purified fractions generated from 293TD cells transfected with 2XMBP–FL BRCA2 were run on 6% SDS–polyacrylamide gels, transferred to PVDF membranes overnight, blocked in 5% milk with 1× TBS-T, and incubated with the primary antibody, Ab-2 (EMD), overnight. For immunoprecipitations, lysates were quantified for protein content by Bradford method and 1 mg total protein was used in an immunoprecipitation reaction containing 20 μl anti-BRCA2 (Ab-1, EMD) antibody and 40 μl protein G+ agarose (Santa Cruz Biotechnology). Immunoprecipitations were rocked for 2 h at 4 °C and then washed with buffer H followed by resuspension in 15 μl sample buffer. The samples were heated at 54 °C for 4 min, loaded onto 6% SDS–polyacrylamide gels, and processed for western blotting as described above. A secondary antibody, anti-mouse or anti-rabbit horse radish peroxidase-conjugated (Santa Cruz Biotechnology) was incubated on the membranes for 40 min. Blots were then washed four times and incubated with ECL Plus (Amersham GE healthcare) for 5 min before visualization on a Storm PhosphorImager.

Generation of stable cell lines and clonogenic survival assay. The MBP–BRCA2 and 2XMBP–BRCA2 constructs were stably transfected into VC8 (gift from M. Zdienicka) BRCA2 mutant hamster cells, using FuGene6 (Roche) transfection followed by selection in HAM's F10 media (Invitrogen) plus 10% FBS (Invitrogen) containing 1 mg ml^{-1} G418. To verify expression of BRCA2, total RNA was isolated from VC8 stable cell clones using TRIzol (Invitrogen). Total RNA (1 μg) was used in each RT–PCR (Titanium One-step RT–PCR, Clontech) reaction containing either MBP primer set (RJ-5′MBP2XN/RJ-3′MBP2XN) to amplify the MBP tag (1.1 kb) or C-terminal primer set (RJ-5′8269/RJ-3′AGEIBRCA2) to amplify the last 0.9 kb of the BRCA2 open reading frame. RT–PCR reactions were run on 1% agarose gel and visualized with ethidium bromide staining on an Alpha Innotech UV imager. To confirm expression of 2XMBP– or MBP–BRCA2 in VC8 cells at the protein level, an immunoprecipitation/western using Ab-1 (EMD) as the immunoprecipitation antibody and Ab-2 (EMD) as the western antibody was performed. We consistently observed higher expression of the 2XMBP–BRCA2 protein compared to MBP–BRCA2 both in stable VC8 clones, as well as in transient transfections of both VC8 and human 293T cells. Clones positive for expression were tested for complementation by clonogenic survival response to mitomycin C. Cells were seeded at plating density of 5×10^5 cells in 6-cm dishes. Cells were 50–70% confluent at time of drug treatment. Mitomycin C (Sigma) stock was a 1.5 mM stock solution. For treatment of cells, mitomycin C was diluted in 2 ml HAM's F10 media with no serum at the following concentrations: 0.1 μM, 0.25 μM and 0.5 μM and placed on cells for 1 h. After 1 h incubation, the media was aspirated off, cells were washed with PBS, then trypsinized and resuspended in 2 ml of HAM's F10 + 10% FBS. Cells were counted using a haemocytometer, serially diluted, and plated out in triplicate into 6-well plates. Cells were re-fed with media containing penicillin/streptomycin (Hyclone) to prevent any contamination during 8 days of cell growth. After 8 days, the cells were removed from the incubator, washed with 0.9% NaCl (saline solution), and stained with crystal violet. Plates were dried overnight and colonies containing 50 or more cells were counted on each plate and the surviving fraction was determined for each drug treatment.

Protein affinity pull-downs. Prior to pull-down assays, amylose resin (NEB) was equilibrated with binding buffer B: 50 mM HEPES (pH 7.5), 250 mM NaCl, 0.5 mM EDTA and 1 mM DTT. Purified 2XMBP–BRCA2 (2.4 μg) was incubated with 1 μg purified RAD51, RPA, SSB, RecA, yRad51, DMC1 or RAD52 for 30 min at 37 °C and then batch-bound to 30 μl of amylose resin for 1 h at 4 °C. RAD51 and RPA were purified as described previously[21]. SSB, RecA and yRad51 were purified as described[51–53], respectively. RAD52 was a gift from A. Mazin, and the purification of DMC1 (A. Nimonkar) will be described elsewhere. As controls for non-specific binding to the amylose resin, candidate proteins (1 μg) were incubated with amylose resin in the absence of 2XMBP–BRCA2. The complexes were then washed with buffer B containing 0.1% Igepal CA-630 and resuspended in protein sample buffer, heated at 54 °C for 4 min, and loaded onto a 4–15% gradient SDS–polyacrylamide gel (Bio-Rad TGX gel). The gel was run for 1 h at 100 V and stained with SYPRO Orange (Invitrogen) or Coomassie (Bio-safe, Bio-Rad). The protein bands were quantified by ImageQuant software on a Storm 860 PhosphorImager (Molecular Dynamics). The amount of RAD51 pulled down with 2XMBP–BRCA2 in Fig. 1e was determined using standard curves generated from known concentrations of RAD51 and 2XMBP–BRCA2 run in parallel in the same gel. The total input amount of 2XMBP–BRCA2 in each pull-down reaction was 64 nM and the total input amount for RAD51 ranged from 85 nM to 2 μM. The analyses to determine the ratio of RAD51 to 2XMBP–BRCA2 was derived from a fit to a segmented linear regression (GraphPad Prism 5.0b).

Electrophoretic mobility shift assays. Oligonucleotide substrates were obtained from either Sigma or IDT (Ultramers) and were purified by polyacrylamide gel electrophoresis (PAGE). The following oligonucleotides were used: RJ-167-mer (5′-CTGCTTTATCAAGATAATTTTTCGACTCATCAGAAATATCCGTTTC CTATATTTATTCCTATTATGTTTTATTCATTTACTTATTCTTTATGTTCA TTTTTTATATCCTTTACTTTATTTTCTCTGTTTATTCATTTACTTATTTTG TATTATCCTTATCTTATTTA-3′); RJ-5′TAIL-167-mer (5′-ATTTATTCTAT TCCCTTTATTTTCTCTGTTTATTCATTTACTTATTTTGTATTAATTTCCTA TATTTTTTACTTGTATTTCTTATTCATTTACTTATTTTGTATTATCCTTAT TTATATCCTTTCTGCTTTATCAAGATAATTTTTCGACTCATCAGAAATAT CCG-3′); RJ-PHIX-42-1 (5′-CGGATATTTCTGATGAGTCGAAAAATTATC TTGATAAAGCAG-3′); RJ-Oligo1 (5′-TAATACAAAATAAGTAAATGAAT AAACAGAGAAAATAAAG-3′); RJ-Oligo2 (5′-CTTTATTTTCTCTGTTTATT CATTTACTTATTTTGTATTA-3′). To generate the 3′ tailed DNA substrate, RJ-167-mer was radio-labelled with ^{32}P at the 5′-end and then annealed at a 1:1 molar ratio to RJ-PHIX-42-1. To generate the 5′ tailed DNA substrate, RJ-5′TAIL-167-mer was radio-labelled with ^{32}P at the 5′-end and annealed at 1:1 molar ratio to RJ3′PHIX-42-1. The dsDNA was generated by radio-labelling RJ-Oligo1 with ^{32}P at the 5′-end and annealing it to RJ-Oligo2. The ssDNA substrate was RJ-167-mer radio-labelled with ^{32}P at the 5′-end.

2XMBP–BRCA2, at the indicated concentrations, was incubated with 0.2 nM (molecules) radio-labelled DNA substrate for 30 min at 37 °C in DNA strand exchange buffer (25 mM Tris acetate (pH 7.5), 1 mM $MgCl_2$, 2 mM $CaCl_2$, 0.1 μg μl^{-1} BSA, 2 mM ATP and 1 mM DTT). The reactions were resolved by electrophoresis on a 6% polyacrylamide gel in TAE (40 mM Tris-acetate (pH 7.5), 0.5 mM EDTA) buffer for 70 min at 60 V. The gel was then dried and exposed to a PhosphorImager screen overnight. The screen was scanned on a Molecular Dynamics Storm 840 PhosphorImager and bands quantified using ImageQuant software. The percentage of protein–DNA complexes was calculated as the free

radio-labelled DNA remaining in a given lane relative to the protein-free lane, which defined the value of 0% complex (100% free DNA). Where BRCA2 was preincubated with RAD51, the two proteins were mixed at the indicated concentrations at 37 °C for 15 min before addition of the radiolabelled DNA substrate, which was followed by an additional 30 min incubation at 37 °C.

DNA strand exchange assays. DNA substrates were generated as described above for the EMSA analysis except that RJ-167-mer and RJ-5′TAIL-167-mer were not radio-labelled. The dsDNA donor was generated by first radio-labelling RJ-Oligo1 with ^{32}P on the 5′-end and annealing it to RJ-Oligo2 at a 1:1 molar ratio. The assay buffer contained: 25 mM Tris acetate (pH 7.5), 1 mM $MgCl_2$, 2 mM $CaCl_2$, 0.1 μg μl^{-1} BSA, 2 mM ATP and 1 mM DTT. All pre-incubations and reactions were at 37 °C. The DNA substrates and proteins were at the following concentrations unless otherwise indicated in the figure legend: RPA (0.1 μM); RAD51 (0.22 μM); 3′ tail, 5′ tail or ssDNA (4 nM molecules); and dsDNA (4 nM molecules). Unless a time course was shown, the reaction time was 30 min. Where proteins were omitted, storage buffer was substituted. The protein storage buffer contributed an additional 68 mM NaCl and 4% glycerol to the assay buffer used in the reactions. RecA (0.22 μM) reactions contained 3 mM ATPγS instead of ATP and were performed in 10 mM $MgCl_2$ in the absence of $CaCl_2$. Reactions using SSB (0.1 μM) contained 5 mM $MgCl_2$. The product yield using RecA alone or hRAD51 alone was 59% and 37%, respectively. The reaction was terminated with Proteinase K in 0.5% SDS for 10 min. The reactions were loaded on a 6% polyacrylamide gel in TAE buffer and electrophoresis was at 60 V for 70 min. The gel was then dried and exposed to PhosphorImager screen overnight. The percentage of DNA strand exchange product was calculated as labelled product divided by total labelled input DNA in each lane.

ATP hydrolysis assays. The assay was carried out essentially as described[21]. Briefly, BRCA2 at concentrations 0–100 nM was preincubated with 3′ tail DNA (0.9 μM nucleotides) in 10 μL of buffer containing 20 mM Tris-HCl (pH 7.5), 4 mM $MgCl_2$, 1 mM DTT, 0.5 mM ATP and 20 μCi ml^{-1} [γ^{32}P]-ATP. The reaction was started by adding RAD51 (0.3 μM) or storage buffer, and further incubated at 37 °C for 90 min. Aliquots (2 μl) were spotted onto a polyethyleneimine (PEI) thin layer chromatography (TLC) plate (EMD Chemicals) at each time point. The spots were air-dried and the plates were developed in 1 M formic acid and 0.5 M LiCl. The amount of ATP hydrolyzed was determined from dried plates using a Molecular Dynamics Storm 840 PhosphorImager. The percentage of ATP hydrolysis was quantified using ImageQuant software and any residual signal from the BRCA2-only lanes was subtracted from the RAD51 + BRCA2 lanes. The results were plotted using GraphPad Prism 5.0b.

Single-stranded DNA annealing assays. Cold 167-mer (RJ-167mer, IDT Ultramer, PAGE-purified) at 8 nM (molecules) and 5′ radio-labelled 40-mer (RJ-Oligo1, Sigma, PAGE-purified) at 4 nM (molecules) were each incubated separately in 10 μl reactions containing 25 mM Tris acetate (pH 7.5), 1 mM $MgCl_2$, and 1 mM DTT for 5 min with or without RPA (100 nM). The 40-mer is complementary to the 167-mer at the 3′ end. All incubations were at 37 °C. The oligonucleotides were then incubated with either BRCA2 (40 nM), RAD52 (100 nM) or protein storage buffer for 5 min. The two separate reactions were then mixed and incubated for 1, 5, 15 or 30 min to allow for annealing. At the indicated time points aliquots were removed and added to stop buffer (4 mg ml^{-1} proteinase K, 1% SDS and 0.4 μM unlabelled 40-mer (RJ-Oligo2)) complementary to RJ-Oligo1 for 15 min. Loading dye was then added to the samples and they were run on 6% polyacrylamide gels in TAE buffer for 1 h at 60 V. The gels were dried onto DEAE (Whatman) paper and exposed to a PhosphorImager screen overnight. The screens were scanned on a Storm 860 system (Molecular Dynamics) and bands quantified using ImageQuant. The percentage of annealed product was calculated as the radio-labelled product divided by the total radio-labelled input DNA in each lane.

Biotinylated DNA pull-down assay. An oligonucleotide substrate composed of the same sequence as RJ-PHIX-42-1 but containing a 3′ biotin modification (Biotin-TEG) was obtained from Eurofins MWG Operon. The 3′ biotin modified oligonucleotide was purified by polyacrylamide gel electrophoresis (PAGE) and annealed at a 1:1 molar ratio to RJ-167-mer to create the biotinylated 3′ tailed DNA substrate. RAD51 was preincubated with or without BRCA2 for 15 min at 37 °C in buffer S: 25 mM Tris acetate (pH 7.5), 1 mM $MgCl_2$, 0.1 μg μl^{-1} BSA, 2 mM ATP and 1 mM DTT. The proteins were then added to a mixture of both the biotinylated 3′ tailed DNA and a heterologous 90 bp duplex DNA (PAGE-purified Oligo#90: 5′-CGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGT-3′ annealed 1:1 to PAGE-purified Oligo#60: 5′-ACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCG-3′) derived from the pUC19 plasmid sequence. The incubation was continued for 5 min at 37 °C followed by capture of the DNA–protein complexes by adding 2.5 μl of pre-washed MagnaLink Streptavidin magnetic beads (SoluLink) in buffer S supplemented with 0.1% Igepal CA-630. After binding to the beads at 25 °C for 10 min, the bead complexes were washed with buffer S containing 0.1% Igepal CA-630 and resuspended in 15 μl protein sample buffer, heated at 54 °C for 4 min, and loaded onto a 4–15% gradient SDS–polyacrylamide gel. The amount of RAD51 protein bound and eluted from the biotinylated DNA substrate was determined by western blot using a monoclonal antibody specific to human RAD51 (14B4, Novus).

51. LeBowitz, J. H. *Biochemical Mechanism of Strand Initiation in Bacteriophage Lambda DNA Replication.* Ph.D. thesis, Johns Hopkins Univ. (1985).
52. Mirshad, J. K. & Kowalczykowski, S. C. Biochemical characterization of a mutant RecA protein altered in DNA-binding loop 1. *Biochemistry* **42,** 5945–5954 (2003).
53. Zaitseva, E. M., Zaitsev, E. N. & Kowalczykowski, S. C. The DNA binding properties of *Saccharomyces cerevisiae* Rad51 protein. *J. Biol. Chem.* **274,** 2907–2915 (1999).

doi:10.1038/nature09452

High star formation rates as the origin of turbulence in early and modern disk galaxies

Andrew W. Green[1], Karl Glazebrook[1], Peter J. McGregor[2], Roberto G. Abraham[3], Gregory B. Poole[1], Ivana Damjanov[3], Patrick J. McCarthy[4], Matthew Colless[5] & Robert G. Sharp[5]

Observations of star formation and kinematics in early galaxies at high spatial and spectral resolution have shown that two-thirds are massive rotating disk galaxies[1-5], with the remainder being less massive non-rotating objects[2,4,6-8]. The line-of-sight-averaged velocity dispersions are typically five times higher than in today's disk galaxies. This suggests that gravitationally unstable, gas-rich disks in the early Universe are fuelled by cold, dense accreting gas flowing along cosmic filaments and penetrating hot galactic gas halos[9,10]. These accreting flows, however, have not been observed[11], and cosmic accretion cannot power the observed level of turbulence[12]. Here we report observations of a sample of rare, high-velocity-dispersion disk galaxies in the nearby Universe where cold accretion is unlikely to drive their high star formation rates. We find that their velocity dispersions are correlated with their star formation rates, but not their masses or gas fractions, which suggests that star formation is the energetic driver of galaxy disk turbulence at all cosmic epochs.

Understanding how these different kinematic states of star-forming galaxies fit together is complicated by selection, surface brightness and angular resolution effects. Particularly at high redshift, z, resolution is a major limitation. The resolution gain of adaptive optics has allowed kinematic observations of early disks, but not all observations, even within a particular survey, have taken advantage of adaptive optics[2,4,8]. Integral-field spectroscopy (IFS) (mainly of the Hα emission line, which traces star formation and kinematics) has shallower surface brightness limits than integrated spectroscopy. Kinematic observations are generally limited to the brightest (the characteristic Schechter luminosity, L^*, or brighter[5]) high-redshift galaxies. Surface brightness is also a strong function of redshift (proportional to $(1+z)^4$), and optical passbands commonly used for imaging sample the rest-frame ultraviolet. These effects often complicate comparisons between the early and the modern Universe. Finally, local IFS comparison samples are often selected from small volumes and non-uniformly.

To quantify how these difficulties might affect previous results, we used the well-studied Sloan Digital Sky Survey (SDSS) to put kinematic galaxy states in the context of a large, uniformly selected sample. We undertook the first IFS observations of 65 star-forming[13] galaxies at redshift $z \approx 0.1$. As active galactic nuclei interfere with Hα emission as a star formation tracer, they have been excluded. (Further details of the selection criteria are in the Supplementary Information.) Observed galaxies have Hα luminosities of $10^{40.7}$ to $10^{42.6}$ erg s^{-1}, with a median of $10^{41.9}$ erg s^{-1}, and a stellar mass range of $10^{9.1}$ to $10^{10.9}$ solar masses, with a median of $10^{10.3}$ solar masses[14]. Because only bright (which we shall define here as $L_{\mathrm{H}\alpha} > 10^{42}$ erg s^{-1}) objects are typically detected at $z \approx 2$, such objects make up half our observations despite representing only 3.2% of SDSS galaxies that otherwise meet our criteria. A broad selection illustrates the impact of surface brightness and luminosity on our results and those reported for the high-redshift Universe.

Chosen galaxies were observed using the integral-field spectrograph SPIRAL[15] on the 3.9-m Anglo-Australian Telescope or WiFeS[16] on the Australia National University's 2.3-m telescope. Median seeing of 1.3″ corresponds to a median spatial resolution of 2.3 kpc, and the field of view is 10–40 kpc. This is closely matched to high-redshift samples observed with adaptive optics, but with better spectral resolution ($\lambda/\Delta\lambda = 7{,}000$–$11{,}500$). Following standard methods, we fit a Gaussian profile to the Hα emission line spectrum at each spatial position in the reduced data cube. The free parameters of the fit are velocity dispersion (width, corrected for instrumental broadening), flux (height) and velocity (position). The velocity dispersion, σ, is a simple, model-independent measure of the line-of-sight kinematics of a galaxy. Although there are many ways to measure it, we adopt the flux-weighted local mean definition commonly used at high redshift[4,7,8]:

$$\sigma_{\mathrm{m}} = \frac{\sum \sigma_{\mathrm{pix}} f_{\mathrm{pix}}}{\sum f_{\mathrm{pix}}}$$

where σ_{pix} is the standard deviation of the Gaussian fit to the Hα emission line in each spatial pixel and f_{pix} is the Hα flux in that spatial pixel. This quantity measures the intrinsic velocity dispersion of the H II star-forming regions and their random motions independent of large-scale systematic motions, such as rotation or orbital motion in a merger. Apparent velocity dispersion can also peak where systematic motions change abruptly, such as in the bright central regions of a disk galaxy where the velocity curve is steepest. We demonstrate that this does not affect our results in Supplementary Information.

Unexpectedly, we find many disk-like velocity fields (Fig. 1) with $\sigma_{\mathrm{m}} > 50$ km s^{-1} (high dispersion) among the galaxies with $L_{\mathrm{H}\alpha} > 10^{42}$ erg s^{-1} (high luminosity), which are values very similar to those at high redshift. We use two simple, qualitative criteria common to high-redshift analyses[6] to separate disks from non-disks on the basis of their velocity and velocity dispersion fields: the velocity field must have a typical 'spider diagram' shape, and the velocity dispersion must be centrally peaked and fairly symmetric. Using these criteria, all but six of our 17 objects at high dispersion and high luminosity are kinematic disks. The broadband SDSS images also show that most are disk galaxies, often with distinct bulges and disk components. This agrees with the picture that most of the star formation in the local Universe occurs in disks[17] and that those disks forming stars most rapidly are kinematically hot, just as in the early Universe.

We compare the distribution of our galaxies in σ_{m} and Hα luminosity (Fig. 2) with several existing data sets from the early[4,7,8] and the modern[18–20] Universe where σ_{m} has been measured in a comparable way (a summary of these data sets appears in Supplementary Information). To check that the decrease in surface brightness with redshift does not affect this comparison, we applied redshift dimming (to $z = 2.2$) and a surface brightness limit ($>10^{16}$ erg s^{-1} cm^{-2} arcsec^{-2}, half that reported for OSIRIS[7]) to our data and recomputed σ_{m}. We find that this affects our measure of σ_{m} by only 5–10 km s^{-1}.

[1]Centre for Astrophysics and Supercomputing, Swinburne University, PO Box 218, Hawthorn, Victoria 3122, Australia. [2]Research School of Astronomy and Astrophysics, Australian National University, Cotter Road, Weston, Australian Capital Territory 2611, Australia. [3]Department of Astronomy and Astrophysics, University of Toronto, 50 St George Street, Toronto, Ontario M5S3H4, Canada. [4]Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, California 91101, USA. [5]Australian Astronomical Observatory, PO Box 296, Epping, New South Wales 1710, Australia.

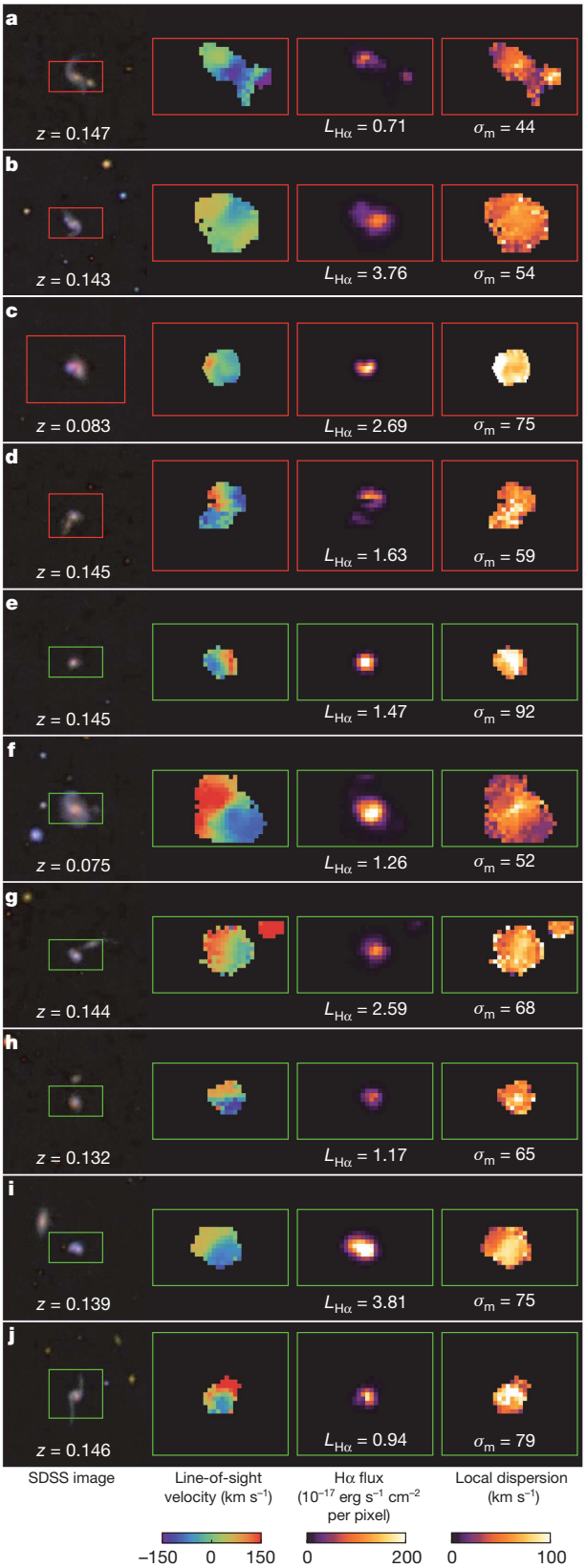


Figure 1 | Kinematic pictures of local galaxies. All panels show the SDSS broadband image (red, ~750 nm; green, ~625 nm; blue, ~475 nm), line-of-sight velocity map, Hα narrowband image and line-of-sight velocity dispersion map. All are scaled identically according to the scale at the bottom of the figure. The maps are masked when the emission line fitting breaks down, typically at a 4σ significance detection threshold. The green and red squares outline the field of view of the instrument in the images; red corresponds to non-disk objects and green to disk-like objects. The broadband imaging in **a**, **b** and **d** shows clear merger activity, whereas **c** may have a merger axis close to the line of sight. Panel **a** shows a relatively small velocity dispersion, despite being a clear merger, suggesting that dispersion alone cannot distinguish mergers. Panels **e**–**j** show objects that have high dispersions ($\sigma > 50\,\mathrm{km\,s^{-1}}$) but are considered disks by our criteria. We emphasize that these criteria are based on those typically used in high-redshift observations[6]. Panel **f** shows a clear disk, with both a blue disk and a red bulge component visible in the broadband image. Panel **g** shows a velocity gradient and a weak central dispersion peak. This galaxy's neighbour is 0.62 mag fainter and has a similar radial velocity. The pair may ultimately become a major merger. The galaxy in **h** also has a companion that may or may not be at the same redshift. Panel **i** shows two distinct clumps in the broadband image, indicative of a major merger, but still shows the clear kinematic characteristics of a disk. Finally, **j** shows the possible signature of a disk in the kinematic data, but is a clear merger, with long tidal tails in the image. With the strong surface brightness fading with redshift, these tails would probably be invisible in a similar galaxy in the early Universe, and a disk interpretation would be made.

Furthermore, all of these comparison data sets fall within the stellar mass range of our data. Remarkably, σ_m seems to correlate with Hα luminosity above $L_{H\alpha} = 10^{42}\,\mathrm{erg\,s^{-1}}$, independent of redshift.

The Hα luminosity is directly correlated with the star formation rate[21]. However, dust extinction in the galaxy can obscure some of this light, reducing the inferred star formation rate. Because dust correction methods differ between studies, we use the observed quantity $L_{H\alpha}$ to provide a simple, direct comparison between samples. In our sample, the mean flux ratio, Hα/Hβ = 4.13 (based on SDSS fibre spectra), corresponds to a typical extinction in Hα of $A_{H\alpha} = 0.78$ mag (ref. 22). This correction has been included in the axis across the top of Fig. 2 to give a rough scale to the star formation rate for these galaxies.

Although primordial galaxies have been interpreted as turbulent, gas-rich disks[2,10], our results suggest that gas density does not drive velocity dispersion. By applying the Kennicutt–Schmidt law[23] to invert our star formation surface densities into gas surface densities, we estimate the gas mass fraction (relative to stars and gas) for these galaxies to range from 0.05 to 0.7 (median, 0.18). The distributions of gas fractions between high- and low-dispersion galaxies do not differ significantly, nor do the gas depletion timescales of 3–6 billion years. The wide range in gas fraction over a small range in luminosity is due to the wide range in stellar mass of these galaxies—stellar mass does not correlate with Hα luminosity. Figure 2 shows that there is no correlation between stellar mass and velocity dispersion. This implies that star formation rate, not mass or gas fraction, is the important variable driving different velocity dispersions at all redshifts.

These high-dispersion galactic disks in the local Universe are an unexpected find. Continued detailed follow-up of these objects will shed more light on similar objects seen at high redshift. Cold-flow accretion is unlikely to be the origin of the high velocity dispersions as this mechanism is expected[24] to shut down rapidly for galaxies at $z < 2$. The high velocity dispersions in low-redshift galaxies then raise the question of whether cold-flow accretion is the appropriate mechanism for high velocity dispersions at high redshifts. Recent absorption line measurements have found no evidence for ubiquitous cold flows in the high-redshift Universe[11]. Our results suggest that star formation itself powers the turbulence through energetic feedback. Indeed, simulations show that supernovae resulting from star formation can drive high velocity dispersions in the interstellar medium[25].

It remains necessary to provide a mechanism to fuel the high star formation rates of these galaxies. Although the kinematics of our sample are not merger dominated, there is evidence in the optical morphologies for minor-merger features such as small, close companions or tidal tails, many of which would be missed by current observations at high redshift (see, for example, Fig. 1f, j). Fresh gas brought in by these minor mergers may drive the high star formation rates. These mergers could also increase the velocity dispersions of the disks. The situation could be similar in the early Universe; the minor-merger mechanism has been commonly invoked to drive the drastic size

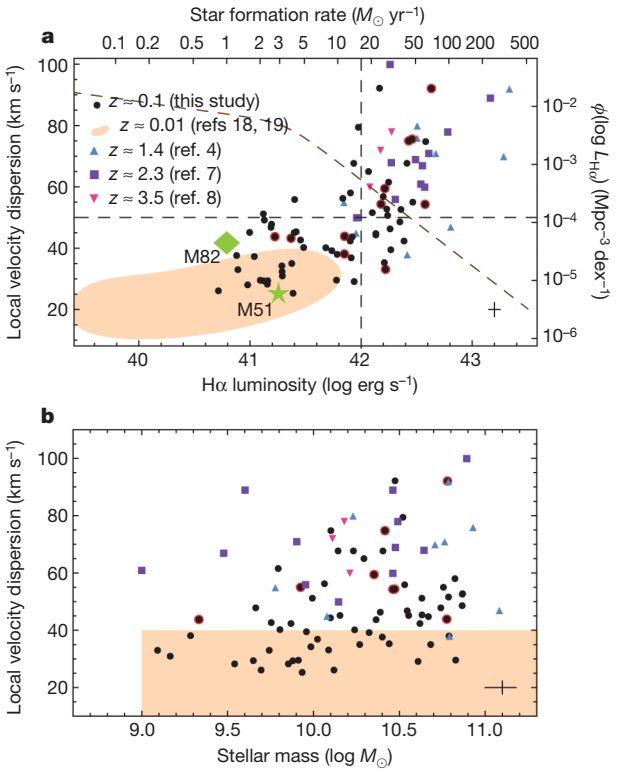


Figure 2 | Distribution of galaxy velocity dispersion against Hα luminosity and stellar mass. **a**, Distribution of flux-weighted mean velocity dispersion and Hα luminosity or star formation rate of galaxies in our sample and comparison samples at $z \approx 2$ (refs 4, 7, 8) and $z < 0.01$ (refs 18, 19). The error bars in the lower right show the combined median errors. Velocity dispersion errors are typically ±5–10% (systematic) and ±1–2 km s^{-1} (statistical), and luminosity errors are typically ±10% (statistical). The variance and error in velocity dispersion are discussed in Supplementary Information. $M_\odot$, solar mass. **b**, Same as in **a**, but plotted against stellar mass instead of Hα luminosity. The comparison samples are described in Supplementary Information. The positions of local galaxies M51[20] and M82[28] are labelled. Dashed lines at $\sigma_m = 50$ km s^{-1} and $L_{H\alpha} = 10^{42}$ erg s^{-1} separate the different regimes described in the text. Velocity dispersion seems to correlate with Hα luminosity but not with mass. Non-disk galaxies are highlighted with red circles and are not distinguishable by velocity dispersion or luminosity. The Hα luminosity is derived from the full aperture of our data cubes and so does not suffer from the large-aperture effects common to SDSS fibre spectra[13]. The brown dashed line in **a** is the Hα luminosity function of SDSS galaxies[29], ϕ (scale on right), for comparison, and shows that the space density of star-forming galaxies declines sharply as luminosity increases. As in **a**, the error bars show combined median errors. Errors in mass are typically ±20% (ref. 14).

evolution of red galaxies[26], requiring minor mergers to occur frequently. Some of our objects also show companions at large distances with luminosity ratios between 1:2 and 1:3. This suggests that gravitational tides could be compressing gas and inducing star formation. Such a 'fly-by' mechanism would be considerably more important in the high-redshift Universe[27] (because of the high merger rate), and such objects could be massive yet faint at optical wavelengths. Further study of the high-dispersion local galaxies is warranted to determine the physical mechanisms in operation. These objects serve as rare reminders of the massive star formation that was so prevalent in the Universe ten billion years ago.

Received 28 April; accepted 20 August 2010.

1. Genzel, R. *et al.* The rapid formation of a large rotating disk galaxy three billion years after the Big Bang. *Nature* **442,** 786–789 (2006).
2. Förster Schreiber, N. M. *et al.* The SINS Survey: SINFONI integral field spectroscopy of $z \sim 2$ star-forming galaxies. *Astrophys. J.* **706,** 1364–1428 (2009).
3. Stark, D. P. *et al.* The formation and assembly of a typical star-forming galaxy at redshift $z \approx 3$. *Nature* **455,** 775–777 (2008).
4. Epinat, B. *et al.* Integral field spectroscopy with SINFONI of VVDS galaxies. I. Galaxy dynamics and mass assembly at $1.2 < z$ 1.6. *Astron. Astrophys.* **504,** 789–805 (2009).
5. Jones, T. A., Swinbank, A. M., Ellis, R. S., Richard, J. & Stark, D. P. Resolved spectroscopy of gravitationally lensed galaxies: recovering coherent velocity fields in subluminous $z \sim 2$–3 galaxies. *Mon. Not. R. Astron. Soc.* **404,** 1247–1262 (2010).
6. Flores, H., Hammer, F., Puech, M., Amram, P. & Balkowski, C. 3D spectroscopy with VLT/GIRAFFE. I. The true Tully Fisher relationship at $z \sim 0.6$. *Astron. Astrophys.* **455,** 107–118 (2006).
7. Law, D. R. *et al.* The kiloparsec-scale kinematics of high-redshift star-forming galaxies. *Astrophys. J.* **697,** 2057–2082 (2009).
8. Lemoine-Busserolle, M., Bunker, A., Lamareille, F. & Kissler-Patig, M. 2D kinematics and physical properties of $z \sim 3$ star-forming galaxies. *Mon. Not. R. Astron. Soc.* **401,** 1657–1669 (2010).
9. Dekel, A. *et al.* Cold streams in early massive hot haloes as the main mode of galaxy formation. *Nature* **457,** 451–454 (2009).
10. Elmegreen, B. G., Bournaud, F. & Elmegreen, D. M. Bulge formation by the coalescence of giant clumps in primordial disk galaxies. *Astrophys. J.* **688,** 67–77 (2008).
11. Steidel, C. C. *et al.* The structure and kinematics of the circum-galactic medium from far-UV spectra of $z \sim 2$–3 galaxies. *Astrophys. J.* **717,** 289–322 (2010).
12. Elmegreen, B. G. & Burkert, A. Accretion-driven turbulence and the transition to global instability in young galaxy disks. *Astrophys. J.* **712,** 294–302 (2010).
13. Brinchmann, J. *et al.* The physical properties of star-forming galaxies in the low-redshift Universe. *Mon. Not. R. Astron. Soc.* **351,** 1151–1179 (2004).
14. Kauffmann, G. *et al.* Stellar masses and star formation histories for 10^5 galaxies from the Sloan Digital Sky Survey. *Mon. Not. R. Astron. Soc.* **341,** 33–53 (2003).
15. Sharp, R. *et al.* Performance of AAOmega: the AAT multi-purpose fiber-fed spectrograph. *Proc. SPIE* **6269,** doi:10.1117/12.671022 (2006).
16. Dopita, M. *et al.* The Wide Field Spectrograph (WiFeS). *Astrophys. Space Sci.* **310,** 255–268 (2007).
17. James, P. A., Knapen, J. H., Shane, N. S., Baldry, I. K. & de Jong, R. S. The Hα galaxy survey. IV. Star formation in the local Universe. *Astron. Astrophys.* **482,** 507–516 (2008).
18. Epinat, B., Amram, P. & Marcelin, M. GHASP: an Hα kinematic survey of 203 spiral and irregular galaxies. VII. Revisiting the analysis of Hα data cubes for 97 galaxies. *Mon. Not. R. Astron. Soc.* **390,** 466–504 (2008).
19. Epinat, B., Amram, P., Balkowski, C. & Marcelin, M. Evidence for strong dynamical evolution in disc galaxies through the last 11 Gyr. GHASP VIII – a local reference sample of rotating disc galaxies for high-redshift studies. *Mon. Not. R. Astron. Soc.* **401,** 2113–2147 (2010).
20. Tully, R. B. The kinematics and dynamics of M51. 1. The observations. *Astrophys. J. Suppl. Ser.* **27,** 415–435 (1974).
21. Kennicutt, R. C. Jr. Star formation in galaxies along the Hubble sequence. *Annu. Rev. Astron. Astrophys.* **36,** 189–232 (1998).
22. Pei, Y. C. Interstellar dust from the Milky Way to the Magellanic Clouds. *Astrophys. J.* **395,** 130–139 (1992).
23. Kennicutt, R. C. Jr. The global Schmidt law in star-forming galaxies. *Astrophys. J.* **498,** 541–552 (1998).
24. Dekel, A. & Birnboim, Y. Galaxy bimodality due to cold flows and shock heating. *Mon. Not. R. Astron. Soc.* **368,** 2–20 (2006).
25. Dib, S., Bell, E. & Burkert, A. The supernova rate-velocity dispersion relation in the interstellar medium. *Astrophys. J.* **638,** 797–810 (2006).
26. Glazebrook, K. Galaxy formation: too small to ignore. *Nature* **460,** 694–695 (2009).
27. Menci, N., Fontana, A., Giallongo, E. & Salimbeni, S. Bimodal color distribution in hierarchical galaxy formation. *Astrophys. J.* **632,** 49–57 (2005).
28. Lehnert, M. D., Heckman, T. M. & Weaver, K. A. Very extended X-ray and Hα emission in M82: Implications for the superwind phenomenon. *Astrophys. J.* **523,** 575–584 (1999).
29. Gilbank, D. G., Baldry, I. K., Balogh, M. L., Glazebrook, K. & Bower, R. G. The local star formation rate density: assessing calibrations using [OII], H and UV luminosities. *Mon. Not. R. Astron. Soc.* **405,** 2594–2614 (2010).

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

Acknowledgements A.W.G. and K.G. acknowledge financial support from the Australian Research Council. A.W.G. acknowledges a special scholarship from the Chancellery of the Swinburne University of Technology. We wish to thank the staff of the Anglo-Australian Observatory and the staff of the ANU 2.3-m telescope for their support of these observations.

Author Contributions K.G. oversaw the project. A.W.G., K.G., I.D., P. J. McGregor, G.B.P. and R.G.S. collected the data at the telescope. A.W.G. completed the data reduction with help from P. J. McGregor and R.G.S. M.C. kindly provided observing time. A.W.G. and K.G. analysed the data and wrote the paper. All authors provided extensive suggestions and comments at each stage of the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.W.G. (agreen@astro.swin.edu.au).

doi:10.1038/nature09392

Single-shot readout of an electron spin in silicon

Andrea Morello[1], Jarryd J. Pla[1], Floris A. Zwanenburg[1], Kok W. Chan[1], Kuan Y. Tan[1], Hans Huebl[1]†, Mikko Möttönen[1,3,4], Christopher D. Nugroho[1]†, Changyi Yang[2], Jessica A. van Donkelaar[2], Andrew D. C. Alves[2], David N. Jamieson[2], Christopher C. Escott[1], Lloyd C. L. Hollenberg[2], Robert G. Clark[1]† & Andrew S. Dzurak[1]

The size of silicon transistors used in microelectronic devices is shrinking to the level at which quantum effects become important[1]. Although this presents a significant challenge for the further scaling of microprocessors, it provides the potential for radical innovations in the form of spin-based quantum computers[2–4] and spintronic devices[5]. An electron spin in silicon can represent a well-isolated quantum bit with long coherence times[6] because of the weak spin–orbit coupling[7] and the possibility of eliminating nuclear spins from the bulk crystal[8]. However, the control of single electrons in silicon has proved challenging, and so far the observation and manipulation of a single spin has been impossible. Here we report the demonstration of single-shot, time-resolved readout of an electron spin in silicon. This has been performed in a device consisting of implanted phosphorus donors[9] coupled to a metal-oxide-semiconductor single-electron transistor[10,11]—compatible with current microelectronic technology. We observed a spin lifetime of ~6 seconds at a magnetic field of 1.5 tesla, and achieved a spin readout fidelity better than 90 per cent. High-fidelity single-shot spin readout in silicon opens the way to the development of a new generation of quantum computing and spintronic devices, built using the most important material in the semiconductor industry.

The projective, single-shot readout of a qubit is a crucial step in both circuit-based and measurement-based quantum computers[12]. For electron spins in the solid state, this has only been achieved in GaAs/AlGaAs quantum dots coupled to charge detectors[13–15]. The spin readout was achieved using spin-dependent tunnelling, in which the electron was displaced to a different location depending on its spin state. The charge detector, electrostatically coupled to the electron site, sensed whether the charge had been displaced, thereby determining the spin state. Here we apply a novel approach to charge sensing, where the detector is not only electrostatically coupled, but also tunnel-coupled to the electron site[11], as shown in Fig. 1a. The strong coupling inherent to this arrangement is responsible for the high charge transfer signals that ultimately allow fast and high-fidelity single-shot spin readout. As a charge detector, we use here the silicon single-electron transistor[10] (SET), a nonlinear nanoelectronic device consisting of a small island of electrons tunnel-coupled to source and drain reservoirs, electrostatically induced beneath an insulating SiO_2 layer. A current can flow from source to drain only when the electrochemical potential of the island assumes specific values[16], resulting in a characteristic pattern of sharp current peaks as a function of gate voltage (Fig. 1e). The shift in electrochemical potential arising from the tunnelling of a single electron from a nearby charge centre into the SET island is large enough to switch the current from zero to its maximum value. This tunnelling event becomes spin-dependent in the presence of a large externally applied magnetic field B, when the spin-up state $|\uparrow\rangle$ has a higher energy than the spin-down state $|\downarrow\rangle$, by an amount $E_Z = g\mu_B B$, where $g \approx 2$ is the spin gyromagnetic ratio and μ_B is the Bohr magneton. The Zeeman splitting E_Z must be larger than the thermal and electromagnetic broadening of electron states in the SET island. Therefore we perform the experiment in high magnetic fields, $B > 1$ T, and with very low electron temperatures, $T_{el} \approx 200$ mK.

The high effective mass of the conduction electrons and the six-fold degeneracy of the conduction band minima (valleys) in silicon[17] imply that very tight confinement is required to isolate a single electron in a non-degenerate state. A phosphorus atom in silicon naturally provides a sharp confining potential for its bound donor electron. The quantum states $|\uparrow\rangle$, $|\downarrow\rangle$ of the electron spin $S = 1/2$ form a natural qubit, with the additional advantage that the ^{31}P nuclear spin can be used as a long-lived quantum memory[18]. Therefore we have fabricated a batch of devices where P donors were implanted in a small region (90×90 nm) next to the SET (Fig. 1c). The P^+ ion fluence was chosen to maximize the likelihood that three donors are located at a distance ~30–60 nm from the SET island and can be tunnel-coupled to it, forming a parallel double-quantum-dot system[19]. The SET top gate and a plunger gate overlaying the P-implanted area provide full electrostatic control of the hybrid double-dot. Here we present the complete data for one device (A), together with the spin lifetime measurements on another one (B).

To perform spin readout, we bias the gates so as to tune the electrochemical potentials on the SET (μ_{SET}) and a nearby donor ($\mu_\downarrow$ and $\mu_\uparrow$ for states $|\downarrow\rangle$ and $|\uparrow\rangle$, respectively) such that the SET current, I_{SET}, is zero when the electron resides on the donor, while $I_{SET} \neq 0$ when the donor is ionized. The readout protocol consists of three phases[13], shown in Fig. 1b. (1) A 'load' phase, during which an electron in an unknown spin state tunnels from the SET island to the donor, as $\mu_{SET} > \mu_\downarrow, \mu_\uparrow$. The electron loading is signalled by I_{SET} dropping to zero. (2) A 'read' phase, during which a spin-down electron remains trapped on the donor, leaving $I_{SET} = 0$, but a spin-up electron can tunnel onto the SET island, causing $I_{SET} = I_{max}$. A (different) spin-down electron from the SET island can later tunnel back onto the donor, blocking the current again. Therefore, the signal of a state $|\uparrow\rangle$ is a single current pulse at the beginning of the read phase. (3) An 'empty' phase during which the donor is ionized, to ensure that a new electron with random orientation can be loaded at the next cycle.

Measuring I_{SET} as a function of the plunger (V_{pl}) and SET top (V_{top}) gate voltages yields the map shown in Fig. 1d. Each time a charge centre coupled to the SET changes its charge state, the sequence of SET current peaks breaks and shifts in gate voltage by an amount $\Delta V_{top} = \Delta q / C_{top}$, where Δq is the charge induced on the SET island and C_{top} is the capacitance between island and top gate (Fig. 1e). Figure 1d shows a large number of charge transitions for $-0.7 < V_{pl} < 0$ V. Most of these transitions are irreproducible and hysteretic, and are probably caused by the charging/discharging of shallow traps at the Si/SiO_2 interface. The transitions for $V_{pl} < -0.7$ V, however, are stable and well reproduced even after several thermal cycles. Their number is consistent with the expected number of implanted donors in the active area, and they have been observed in a number of similar devices[20].

[1]Australian Research Council Centre of Excellence for Quantum Computer Technology, School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, New South Wales 2052, Australia. [2]Australian Research Council Centre of Excellence for Quantum Computer Technology, School of Physics, University of Melbourne, Melbourne, Victoria 3010, Australia. [3]Department of Applied Physics/COMP, Aalto University, PO Box 15100, 00076 Aalto, Finland. [4]Low Temperature Laboratory, Aalto University, PO Box 13500, 00076 Aalto, Finland. †Present addresses: Walther-Meissner-Institut, Bayerische Akademie der Wissenschaften, 85748 Garching, Germany (H.H.); Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA (C.D.N.); Department of Defence, Canberra, Australian Capital Territory 2600, Australia (R.G.C.).

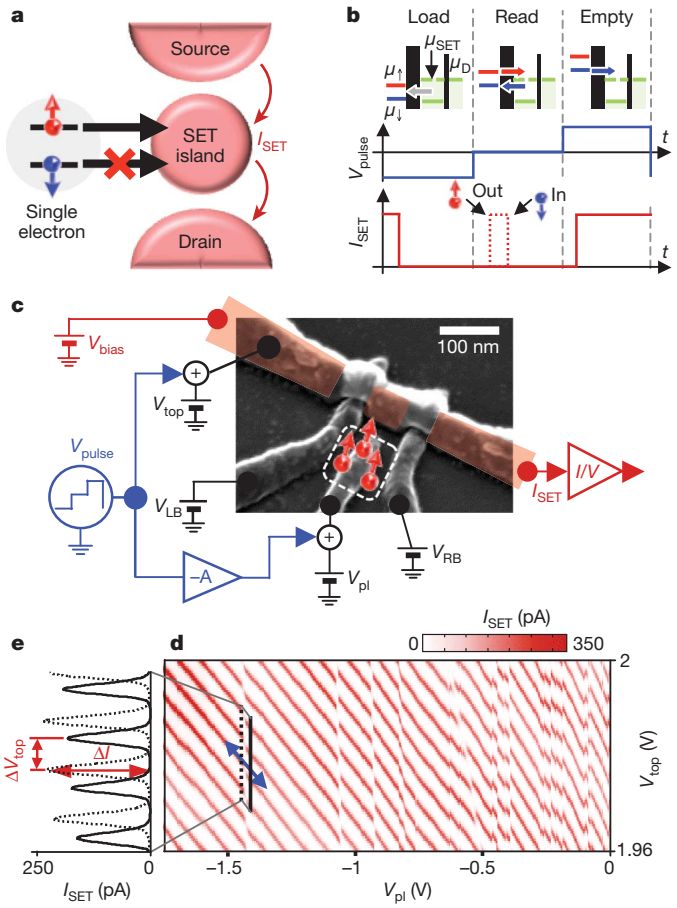


Figure 1 | Spin readout device configuration and charge transitions. **a**, Diagram showing the spin-dependent tunnelling configuration, where a single electron can tunnel onto the island of a SET only when in a spin-up state. **b**, Pulsing sequence for single-shot spin readout (see main text), and SET response, I_{SET}. The dashed peak in I_{SET} is the expected signal from a spin-up electron. The diagrams at the top depict the electrochemical potentials of the electron site ($\mu_{\downarrow,\uparrow}$), of the SET island ($\mu_{SET}$) and of the drain contact (μ_D). **c**, Scanning electron micrograph of a device similar to the one measured. The area where the P donors are implanted is marked by the dashed square. Both d.c. voltages and pulses are applied to the gates as indicated. The red shaded area represents the electron layer induced by the top gate and confined beneath the SiO_2 gate oxide layer. **d**, SET current I_{SET} as a function of the voltages on the top and the plunger gates, V_{top} and V_{pl}, at $B = 0$. The lines of SET Coulomb peaks are broken by charge transfer events. The blue arrow on the transition at $V_{pl} \approx -1.4$ V shows the axis along which V_{top} and V_{pl} are pulsed for compensated time-resolved measurements, ensuring that μ_{SET} remains constant during the pulsing. **e**, Line traces of I_{SET} along the solid and dashed lines in **d**. Ionizing the donor shifts the sequence of SET current peaks by an amount $\Delta V_{top} = \Delta q/C_{top}$, causing a change ΔI in the current. The charging energy of the SET is ~1.5 meV.

Considering the results of the spin lifetime measurements discussed below, it is likely that we are observing transitions between D^+ and D^0 states of implanted P donors[21].

The charge transition at $V_{pl} \approx -1.4$ V in Fig. 1d has a large $\Delta q \approx 0.7e$, where $1e$ is equivalent to the spacing between adjacent current peaks. This indicates a donor very close to the SET island[11]. Accordingly, we find a fast electron tunnelling time between the donor and the SET, of the order of 10 μs. For comparison, the charge transition at $V_{pl} \approx -1.1$ V has a lower $\Delta q \approx 0.3e$ and a much slower tunnel time, ~10 ms, consistent with a donor further away. We chose the donor transition at $V_{pl} \approx -1.4$ V to implement the spin readout protocol. Figure 2b–g illustrates the method we used to find the values of V_{pulse} during the read phase at which spin-dependent tunnelling is achieved. By lowering the read level from too high (Fig. 2c) to too low (Fig. 2g), the time traces of I_{SET} during the read phase show a transition from $I_{SET} = I_{max}$, through random telegraph signal, to $I_{SET} = 0$, passing through a region where I_{SET} can be either zero (Fig. 2e) or show a spin-up signal (Fig. 2f). In this region, the condition $\mu_\downarrow < \mu_{SET} < \mu_\uparrow$ is fulfilled, and a single-shot projective measurement of the electron spin state is performed. When plotting the average of several single-shot traces taken at different read levels, the correct readout range is highlighted by the appearance of a high current region at the beginning of the read phase, spanning a time interval of the order of the electron tunnel time $1/\Gamma$ (Fig. 4). Such a high-current region is absent in measurements performed in zero magnetic field, as expected. With a modified pulse sequence, it is also possible to extract the Zeeman energy splitting, $E_Z = g\mu_B B$, and demonstrate the deterministic loading of a $|\downarrow\rangle$ electron (Supplementary information). Because the loading of a state $|\downarrow\rangle$ is controlled by gate voltages and occurs on ~10-μs timescales as determined by the electron tunnel time, this device already realizes two essential requirements for quantum computation and quantum error correction, namely, single-shot readout and fast preparation of the qubit ground state[22].

Defining $P_\uparrow$ as the probability of observing a spin-up electron, we find that $P_\uparrow$ decreases when increasing the wait time τ_w before the spin is read out (Fig. 3a), because the excited state $|\uparrow\rangle$ relaxes to the ground state $|\downarrow\rangle$. The wait time dependence of $P_\uparrow$ (Fig. 3b, c) is well described by a single exponential decay, $P_\uparrow(\tau_w) = P_\uparrow(0)\exp(-\tau_w/T_1)$, where T_1 is the lifetime of the spin excited state.

The measured spin relaxation rates as a function of magnetic field, $T_1^{-1}(B)$, at phonon temperature $T \approx 40$ mK, are plotted in Fig. 3d. The data on device A for $B \geq 2$ T are well described by the function $T_1^{-1}(B) \approx K_{0A} + K_{5A}B^5$, with $K_{0A} = 1.84 \pm 0.07\ s^{-1}$ and $K_{5A} = 0.0076 \pm 0.0002\ s^{-1}\ T^{-5}$. A fit of the form $T_1^{-1}(B) = K_0 + K_a B^a$, where K_0, K_a and a are free parameters, yields $a = 4.8 \pm 0.2$. The data on device B follow $T_1^{-1}(B) \approx K_{5B}B^5$, with $K_{5B} = 0.015 \pm 0.0005\ s^{-1}\ T^{-5}$ down to $B = 1.5$ T, where the spin lifetime has a value $T_1 = 6 \pm 2$ s. We attribute the B-independent contribution observed in device A to the effect of dipolar coupling between the spin under measurement and those of neighbouring donors (Supplementary Information). This effect depends on the details of the mutual distance between implanted donors, and is therefore strongly sample-dependent. The $T_1^{-1} \propto B^5$ dependence agrees with the low-T limit[23] ($k_B T \ll g\mu_B B$) of a spin–lattice relaxation mechanism arising from valley repopulation[24], that is, the change in the relative weight of the six conduction band minima (valleys) of Si caused by the deformation of the crystal lattice when the state $|\uparrow\rangle$ relaxes to $|\downarrow\rangle$, emitting an acoustic phonon. This is the dominant relaxation channel for donors, where orbital excited states are very high in energy. Conversely, for spins in electrostatically defined quantum dots[23] in silicon, relaxation through low-lying orbital states can lead to $T_1^{-1} \propto B^7$. This dependence has been recently observed in Si/SiGe (ref. 25) and Si/SiO_2 (ref. 26) quantum dots.

Our results are also incompatible with the known relaxation process for interface traps, which is dominated by the coupling to two-level fluctuators[27], yielding $T_1^{-1} \propto B^3$. A recent electron spin resonance experiment on shallow traps at the Si/SiO_2 interface[28] found $T_1 \approx 800$ μs at $T = 350$ mK and $B = 0.32$ T, that is, 2 to 3 orders of magnitude shorter than our result, despite the much lower magnetic field. An experiment on bulk-doped Si:P by conventional electron spin resonance techniques (J. J. L. Morton, personal communication) yielded $T_1 = 0.42$ s at $B = 3.35$ T and $T < 5$ K, that is, in the T-independent regime. This data point is only a factor of ~1.3 below the line $T_1^{-1}(B) \approx K_{5A}B^5$. We conclude that the observation of $T_1^{-1} \propto B^5$ and the quantitative agreement with bulk Si:P data constitute a strong indication that we have measured the spin of a single electron bound to an implanted P donor. The proximity of the donor to electrostatic gates and a Si/SiO_2 interface[29] could be responsible for the slight variability of T_1 (Supplementary Information) but, importantly, does not substantially compromise the long spin lifetime of the donor-bound electron.

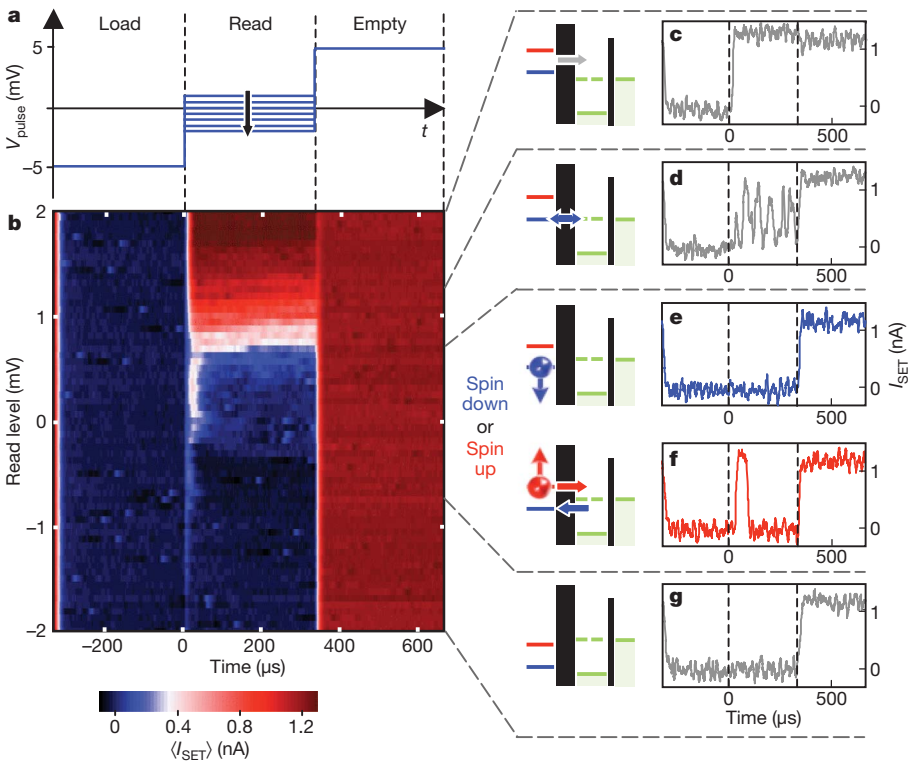


Figure 2 | Single-shot spin readout and calibration of the 'read' level. **a**, Three-level pulsing sequence for spin readout. The 'load' and 'empty' levels are kept constant, while the read level is scanned from high to low. **b**, SET current $\langle I_{SET} \rangle$, averaged over 128 single-shot traces (colour scale), as a function of the V_{pulse} level during the read phase. Data taken with an applied magnetic field $B = 5$ T and a detection bandwidth of 40 kHz (rise time ~10 μs). **c**–**g**, Examples of single-shot traces. **c**, Read level too high, $\mu_{\downarrow} > \mu_{SET}$: the electron always leaves the donor during the read pulse, regardless of its spin. **d**, $\mu_{\downarrow} \approx \mu_{SET}$: random telegraph signal indicates an electron switching between SET island and $|\downarrow\rangle$ state. **e**, **f**, Correct read level, $\mu_{\downarrow} < \mu_{SET} < \mu_{\uparrow}$: $I_{SET} = 0$ during the read phase indicates a $|\downarrow\rangle$ state (**e**). A single current pulse at the beginning of the read phase is the signature of a $|\uparrow\rangle$ state (**f**). The regime of correct read level is recognizable by the isolated increase in $\langle I_{SET} \rangle$ in **b**. **g**, Read level too low, $\mu_{\uparrow} < \mu_{SET}$: the electron never leaves the donor during the read pulse.

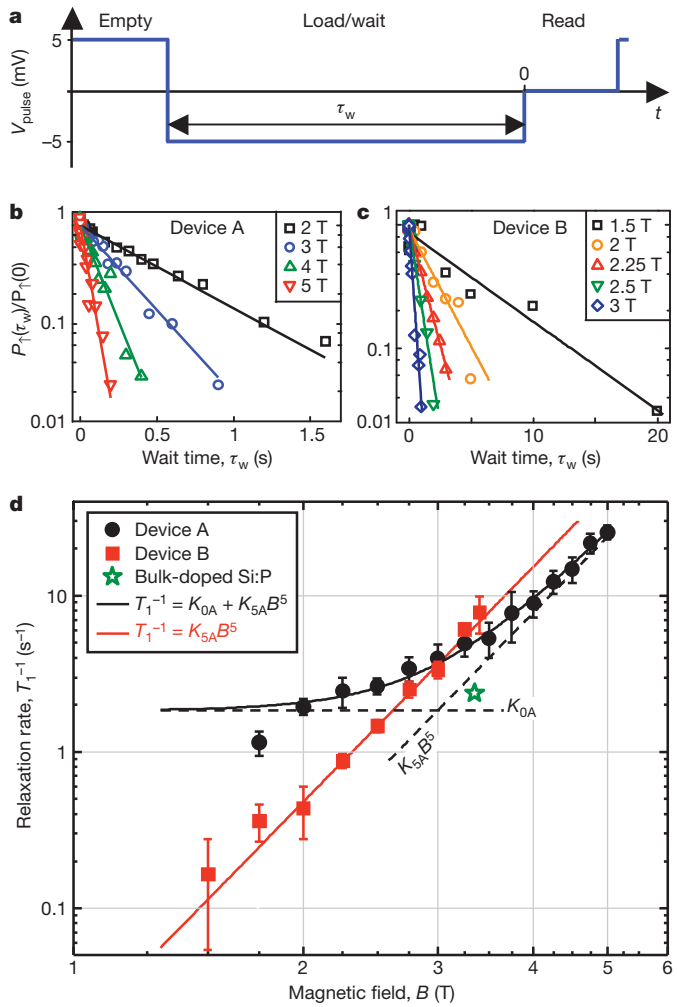


Figure 3 | Spin relaxation rate. **a**, Pulsing sequence for measuring the spin relaxation rate $1/T_1$, identical to Fig. 1b but with a variable load/wait time, τ_w. **b**, **c**, Exponential decays of the normalized spin-up fraction at different magnetic fields, for devices A and B as indicated. **d**, Magnetic field dependence of T_1^{-1}. Error bars, 95% confidence levels. The data for device A follow $T_1^{-1} = 1.84\,\mathrm{s}^{-1} + 0.0076B^5\,\mathrm{s}^{-1}\,\mathrm{T}^{-5}$ (black solid line, sum of the dashed lines). The point at $B = 1.75$ T is not included in the fitted data set. The data for device B follow $T_1^{-1} = 0.015B^5\,\mathrm{s}^{-1}\,\mathrm{T}^{-5}$ (red line). The star is a data point measured on a bulk Si:P crystal at $T < 5$ K (J. J. L. Morton, personal communication).

To assess the effectiveness of the spin readout process for quantum information purposes, it is important to quantify the readout fidelity, that is, the probability that an electron spin state is recognized correctly. In Fig. 4, we show the analysis of the readout fidelity for a set of 10,000 traces. The spin state is declared $|\uparrow\rangle$ if the peak value I_p taken by $I_{SET}(t)$ in the interval $0 < t < 100\,\mu s$ surpasses the threshold current I_T, and $|\downarrow\rangle$ otherwise. The probability distribution of I_p (Fig. 4b) shows well-resolved peaks, indicating that I_p takes two preferential values depending on the electron spin state. We have developed a numerical model that accurately simulates the measurement process and yields two separate histograms of peak current values for the states $|\downarrow\rangle$ and $|\uparrow\rangle$, $N_{\downarrow,\uparrow}(I_p)$, respectively (Supplementary Information). The calculated $N_{\downarrow,\uparrow}(I_p)$ are in excellent agreement with the measured histogram (Fig. 4b).

With the knowledge of $N_{\downarrow,\uparrow}(I_p)$, the readout fidelities[15] are obtained as $F_\downarrow = 1 - \int_{I_T}^{\infty} N_\downarrow(I)\mathrm{d}I$ and $F_\uparrow = 1 - \int_{-\infty}^{I_T} N_\uparrow(I)\mathrm{d}I$ for the states $|\downarrow\rangle$ and $|\uparrow\rangle$, respectively, as a function of the discrimination threshold I_T (Fig. 4c).

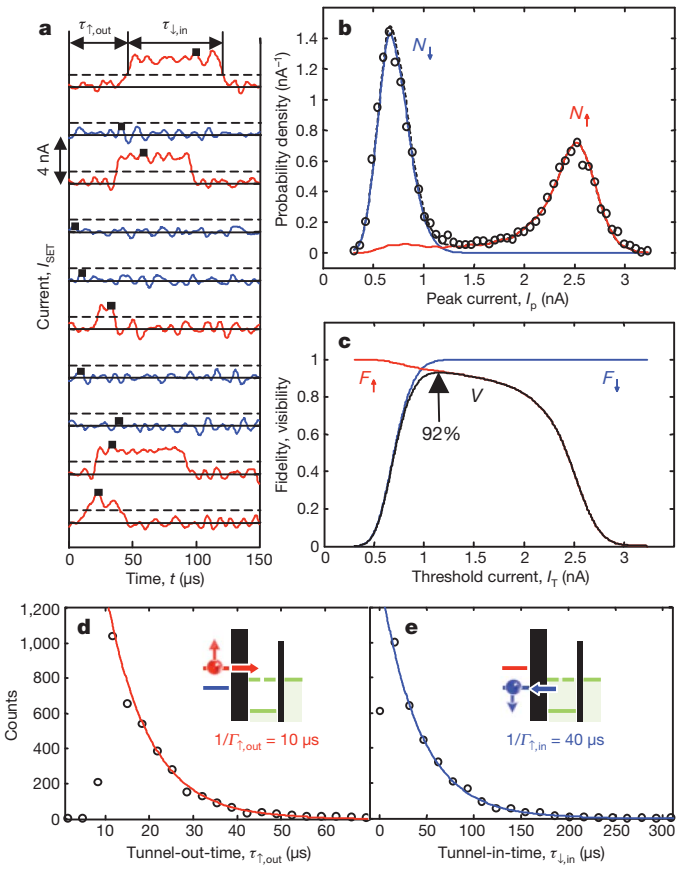


Figure 4 | Readout fidelity and visibility. **a**, Examples of single-shot I_{SET} traces, each shifted by 4 nA for clarity, with $B = 5$ T and 120 kHz bandwidth (~3 μs rise/fall time). The spin is labelled $|\uparrow\rangle$ (red trace) or $|\downarrow\rangle$ (blue trace), depending on whether I_{SET} passes the threshold $I_T = 1.1$ nA (dashed lines). **b**, Histogram (circles) of the maximum values of I_{SET} in the interval $0 < t < 100\,\mu s$ (black squares in **a**), obtained from a 10,000-shots data set. The blue and red lines are simulated histograms for states $|\downarrow\rangle$ and $|\uparrow\rangle$, respectively, and the black dashed line is the sum of the two. The simulated curves are obtained using $P_\uparrow = 0.47$, $\Delta I = 1.9$ nA, $1/\Gamma_{\uparrow,out} = 10\,\mu s$, $1/\Gamma_{\downarrow,in} = 40\,\mu s$. **c**, $|\downarrow\rangle$ (blue) and $|\uparrow\rangle$ (red) readout fidelities, and readout visibility (black) as a function of the discrimination threshold, I_T. The maximum visibility is 92% at $I_T \approx 1.1$ nA. **d**, **e**, Histogram (circles) of the tunnel-out times for spin-up electrons, $\tau_{\uparrow,out}$ (**d**), and subsequent tunnel-in times for spin-down electrons, $\tau_{\downarrow,in}$ (**e**), as defined on the top trace in **a**. In **d**, we note a systematic ~10 μs delay between the beginning of the read phase and the tunnel-out events, due to the response of the amplifier and filter chain. The solid lines are exponential fits to extract the tunnel rates. These values of $1/\Gamma_{\uparrow,out}$ and $1/\Gamma_{\downarrow,in}$ were used to obtain the simulated curves in **b**.

The integrals in $F_{\downarrow,\uparrow}$ represent the probability that the spin state is incorrectly assigned, either because a spin-down trace has a noise spike $>I_T$, or because a spin-up signal does not reach the threshold. The visibility, defined as $V = F_\downarrow + F_\uparrow - 1$, reaches a maximum value of ~92% at $I_T = 1.1$ nA, where the readout fidelities are $F_\downarrow \approx 99\%$ and $F_\uparrow \approx 93\%$.

Combining spin resonance experiments[11,30] with the ability to read out a single spin will provide a promising system in which to demonstrate and exploit coherent quantum control of a donor electron and nuclear spin. The high-fidelity, single-shot electron spin readout, demonstrated here for the first time in silicon, represents an important step towards unlocking the full potential of silicon-based quantum information science and technology.

METHODS SUMMARY

The devices were fabricated on a high-purity, near-intrinsic, natural-isotope (100) silicon substrate, with n^+ ohmic source/drain contacts obtained by phosphorus diffusion. A high-quality SiO_2 gate oxide was grown by dry oxidation at 800 °C, with thicknesses of 5 nm (device A) or 10 nm (device B). Phosphorus ions were implanted through a 90 nm × 90 nm aperture defined by electron-beam lithography in a poly (methyl methacrylate) mask. The chosen fluence maximizes the likelihood of having three P atoms in a 30 nm × 30 nm area, subject to Poisson statistics. The implantation energy was 14 keV, resulting in an average depth of ~15 nm (device A) or ~10 nm (device B) below the Si/SiO_2 interface. A 5 s, 1,000 °C rapid thermal anneal was performed to activate the donors and repair the implantation damage. The aluminium gates to form the SET were defined by electron-beam lithography, followed by Al thermal evaporation and lift-off. A final forming gas anneal (400 °C, 15 min, 95% N_2 / 5% H_2) was performed to reduce the interface trap density to the level of $2 \times 10^{10}\,cm^{-2}\,eV^{-1}$, as measured on devices fabricated with the same process.

The sample was mounted on a high-frequency printed circuit board in a copper enclosure, and thermally anchored to the cold finger of a dilution refrigerator with base temperature ~40 mK. Semirigid coaxial lines, fitted with copper-powder filters at the base temperature plus *RC*-filters on the circuit board, connected the sample to the room-temperature electronics. All d.c. voltages to the gates were provided by optoisolated and battery-powered voltage sources, and added through resistive voltage dividers/combiners to the pulses produced by an arbitrary waveform generator. The SET current was measured by a room-temperature transimpedance amplifier, followed by a voltage post-amplifier, an eighth-order low-pass Bessel filter, and a fast digitizing oscilloscope.

Received 28 April; accepted 2 August 2010.
Published online 26 September 2010.

1. Levi, A. J. F. Towards quantum engineering. *Proc. IEEE* **96,** 335–342 (2008).
2. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57,** 120–126 (1998).
3. Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393,** 133–137 (1998).
4. Hollenberg, L. C. L., Greentree, A. D., Fowler, A. G. & Wellard, C. J. Two-dimensional architectures for donor-based quantum computing. *Phys. Rev. B* **74,** 045311 (2006).
5. Žutić, I., Fabian, J. & Das Sarma, S. Spintronics: fundamentals and applications. *Rev. Mod. Phys.* **76,** 323–410 (2004).
6. Tyryshkin, A. M., Lyon, S. A., Astashkin, A. V. & Raitsimring, A. M. Electron spin relaxation times of phosphorus donors in silicon. *Phys. Rev. B* **68,** 193207 (2003).
7. Feher, G. & Gere, E. A. Electron spin resonance experiments on donors in silicon. II. Electron spin relaxation effects. *Phys. Rev.* **114,** 1245–1256 (1959).
8. Ager, J. W. *et al.* High-purity, isotopically enriched bulk silicon. *J. Electrochem. Soc.* **152,** G448–G451 (2005).
9. Jamieson, D. N. *et al.* Controlled shallow single-ion implantation in silicon using an active substrate for sub-20-keV ions. *Appl. Phys. Lett.* **86,** 202101 (2005).
10. Angus, S. J., Ferguson, A. J., Dzurak, A. S. & Clark, R. G. Gate-defined quantum dots in intrinsic silicon. *Nano Lett.* **7,** 2051–2055 (2007).
11. Morello, A. *et al.* Architecture for high-sensitivity single-shot readout and control of the electron spin of individual donors in silicon. *Phys. Rev. B* **80,** 081307(R) (2009).
12. Ladd, T. D. *et al.* Quantum computers. *Nature* **464,** 45–53 (2010).
13. Elzerman, J. M. *et al.* Single-shot read-out of an individual electron spin in a quantum dot. *Nature* **430,** 431–435 (2004).
14. Hanson, R. *et al.* Single-shot readout of electron spin states in a quantum dot using spin-dependent tunnel rates. *Phys. Rev. Lett.* **94,** 196802 (2005).
15. Barthel, C., Reilly, D. J., Marcus, C. M., Hanson, M. P. & Gossard, A. C. Rapid single-shot measurement of a singlet-triplet qubit. *Phys. Rev. Lett.* **103,** 160503 (2009).
16. Devoret, M. H. & Schoelkopf, R. J. Amplifying quantum signals with the single-electron transistor. *Nature* **406,** 1039–1046 (2000).
17. Goswami, S. *et al.* Controllable valley splitting in silicon quantum devices. *Nature Phys.* **3,** 41–45 (2007); published online 10 December 2006..
18. Morton, J. J. L. *et al.* Solid-state quantum memory using the ^{31}P nuclear spin. *Nature* **455,** 1085–1088 (2008).

19. Hofmann, F. *et al.* Single electron switching in a parallel quantum dot. *Phys. Rev. B* **51,** 13872–13875 (1995).
20. Huebl, H. *et al.* Electron tunnel rates in a donor-silicon single electron transistor hybrid. *Phys. Rev. B* **81,** 235318 (2010).
21. Tan, K. Y. *et al.* Transport spectroscopy of single phosphorus donors in a silicon nanoscale transistor. *Nano Lett.* **10,** 11–15 (2010).
22. DiVincenzo, D. P. The physical implementation of quantum computation. *Fortschr. Phys.* **48,** 771–783 (2000).
23. Hanson, R., Kouwenhoven, L. P., Petta, J. R., Tarucha, S. & Vandersypen, L. K. Spins in few-electron quantum dots. *Rev. Mod. Phys.* **79,** 1217–1265 (2007).
24. Hasegawa, H. Spin-lattice relaxation of shallow donor states in Ge and Si through a direct phonon process. *Phys. Rev.* **118,** 1523–1534 (1960).
25. Hayes, R. R. *et al.* Lifetime measurements (T_1) of electron spins in Si/SiGe quantum dots. Preprint at ⟨http://arxiv.org/abs/0908.0173⟩ (2009).
26. Xiao, M., House, M. G. & Jiang, H. W. Measurement of the spin relaxation time of single electrons in a silicon metal-oxide-semiconductor-based quantum dot. *Phys. Rev. Lett.* **104,** 096801 (2010).
27. de Sousa, R. Dangling-bond spin relaxation and magnetic 1/f noise from the amorphous-semiconductor/oxide interface: theory. *Phys. Rev. B* **76,** 245306 (2007).
28. Shankar, S., Tyryshkin, A. M., He, J. & Lyon, S. A. Spin relaxation and coherence times for electrons at the Si/SiO_2 interface. Preprint at ⟨http://arxiv.org/abs/0912.3037⟩ (2009).
29. Calderón, M. J., Saraiva, A., Koiller, B. & Das Sarma, S. Quantum control and manipulation of donor electrons in Si-based quantum computing. *Appl. Phys. Lett.* **105,** 122410 (2009).
30. Xiao, M., Martin, I., Yablonovitch, E. & Jiang, H. W. Electrical detection of the spin resonance of a single electron in a silicon field-effect transistor. *Nature* **430,** 435–439 (2004).

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

Acknowledgements We thank D. D. Awschalom, C. Tahan, J. J. L. Morton and G. Prawiroatmodjo for comments and suggestions, W. H. Lim for assistance with device fabrication, and R. P. Starrett, D. Barber, A. Cimmino and R. Szymanski for technical assistance. We acknowledge support from the Australian Research Council, the Australian Government, the US National Security Agency and the US Army Research Office under contract number W911NF-08-1-0527. M.M. acknowledges support from the Academy of Finland and the Emil Aaltonen foundation.

Author Contributions A.M., H.H., C.D.N., D.N.J., C.C.E., L.C.L.H., R.G.C. (while at UNSW) and A.S.D. conceived and designed the experiment, K.W.C. and K.Y.T. fabricated the devices, C.Y., J.A.v.D., A.D.C.A. and D.N.J. implanted the P donors, A.M., J.J.P. and F.A.Z. performed and analysed the measurements, A.M., M.M. and J.J.P. analysed the readout fidelity. A.M. wrote the manuscript with input from all coauthors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.M. (a.morello@unsw.edu.au).

doi:10.1038/nature09404

Geometry-induced electrostatic trapping of nanometric objects in a fluid

Madhavi Krishnan[1], Nassiredin Mojarad[1], Philipp Kukura[1]† & Vahid Sandoghdar[1]

The ability to trap an object—whether a single atom or a macroscopic entity—affects fields as diverse as quantum optics[1], soft condensed-matter physics, biophysics and clinical medicine[2]. Many sophisticated methodologies have been developed to counter the randomizing effect of Brownian motion in solution[3–10], but stable trapping of nanometre-sized objects remains challenging[8–10]. Optical tweezers are widely used traps, but require sufficiently polarizable objects and thus are unable to manipulate small macromolecules. Confinement of single molecules has been achieved using electrokinetic feedback guided by tracking of a fluorescent label, but photophysical constraints limit the trap stiffness and lifetime[8]. Here we show that a fluidic slit with appropriately tailored topography has a spatially modulated electrostatic potential that can trap and levitate charged objects in solution for up to several hours. We illustrate this principle with gold particles, polymer beads and lipid vesicles with diameters of tens of nanometres, which are all trapped without external intervention and independently of their mass and dielectric function. The stiffness and stability of our electrostatic trap is easily tuned by adjusting the system geometry and the ionic strength of the solution, and it lends itself to integration with other manipulation mechanisms. We anticipate that these features will allow its use for contact-free confinement of single proteins and macromolecules, and the sorting and fractionation of nanometre-sized objects or their assembly into high-density arrays.

Our trap concept uses topological modulations of the gap between two fluidic slit surfaces[11,12] that acquire a net charge on exposure to water. As sketched in Fig. 1a, one slit surface consists of a silicon dioxide layer topographically structured by standard nanofabrication techniques. The other surface is a cover glass that provides optical access to the interior of the slit (Methods). The loading of the trap uses the capillary effect, which introduces aqueous suspensions of nanometre-sized objects into 200-nm-deep slits at a typical velocity of 100 $\mu m\,s^{-1}$. Particles transported by the flow past the surface indentations become strongly trapped and remain confined after the flow is switched off, as illustrated by optical snapshots of single gold particles 100 nm in diameter trapped by groove- and disc-shaped indentations (Fig. 1b, e).

The motion of confined nanoparticles is mapped with high spatio-temporal resolution on an inverted optical microscope with interferometric scattering detection (iSCAT)[13], using the set-up outlined in Fig. 1d (also see Methods). Briefly, interference of the light fields scattered by the object ($\mathbf{E}_{sca}$) and reflected by the substrate ($\mathbf{E}_{ref}$) yields a signal proportional to $|\mathbf{E}_{ref}||\mathbf{E}_{sca}|\cos(\Delta\varphi)$, where $\Delta\varphi$ denotes the phase difference between the two fields. iSCAT allows fast imaging with high sensitivity to particle motion in the z direction[14], and in contrast to fluorescence imaging allows simultaneous imaging of a particle and the surface nanostructure above it (see, for example, Fig. 1b). Figure 1c, f shows the trajectories of the gold particles seen in Fig. 1b, e, and movies of the dynamics of single trapped objects can be viewed in Supplementary Movies 1, 2 and 3.

The red symbols in Fig. 2a plot the lateral motion of a 100-nm gold particle trapped by a pocket of diameter $D = 500$ nm and depth $d = 100$ nm, in a solution of ionic strength 7×10^{-5} M. The particle, imaged for a period of 1 s at 1 kHz, was localized with a lateral static error less than 4 nm. However, we estimate a dynamic position uncertainty of about 30 nm in view of our measured value of 1.8 $\mu m^2\,s^{-1}$ for the particle diffusion coefficient (which agrees with previous measurements[15]). The radial probability density distribution, $P(r)$, of particle positions is summarized by the red histogram in Fig. 2b, which shows the particle to be confined to within 150 nm of the trap centre. Supplementary Movie 4 is a time-lapse movie of a representative particle trapped and observed for more than 1 h.

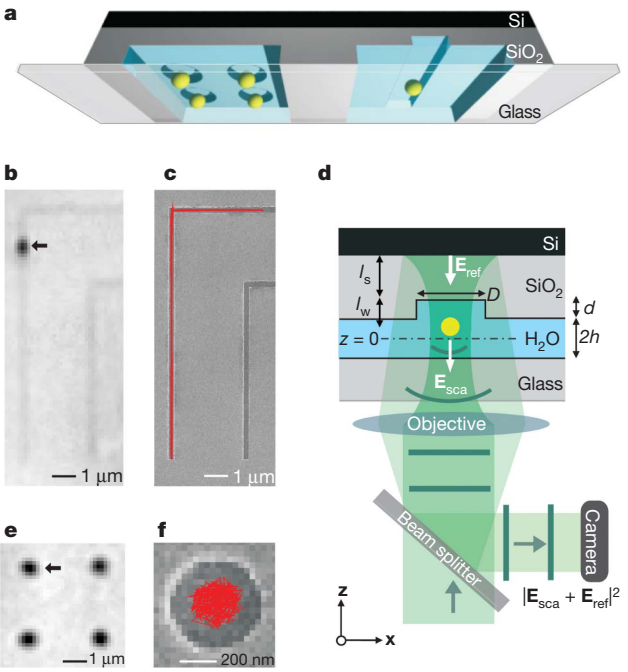


Figure 1 | Pictorial overview of the device and experimental set-up. **a**, Illustration of the fluidic device with nanostructured slit surfaces. **b**, iSCAT image of a gold particle confined by a 'groove' nanostructure. Destructive interference between the light scattered by the particle, $\mathbf{E}_{sca}$, and the reflected field, $\mathbf{E}_{ref}$, causes the particle to appear dark against a light background. **c**, Trajectory of particle motion (red) along a groove overlaid on a scanning electron micrograph (SEM) of the underlying surface topography. **d**, Device and optical set-up for iSCAT (not to scale), highlighting dimensions that are important in the confinement interaction: slit height, $2h = 200$ nm; nanostructure diameter or width, $D = 100$–500 nm; nanostructure depth, $d = 100$ nm. l_s, l_w, physical path lengths in silicon dioxide and water, respectively. **e**, **f**, iSCAT image of single gold particles confined by 'pocket' nanostructures, and the trajectory of a single particle (red) overlaid with an SEM of the underlying surface topography.

[1]Laboratory of Physical Chemistry, ETH Zurich, 8093 Zurich, Switzerland. †Present address: Department of Chemistry, Physical and Theoretical Chemistry Laboratory, University of Oxford, Oxford OX1 3QZ, UK.

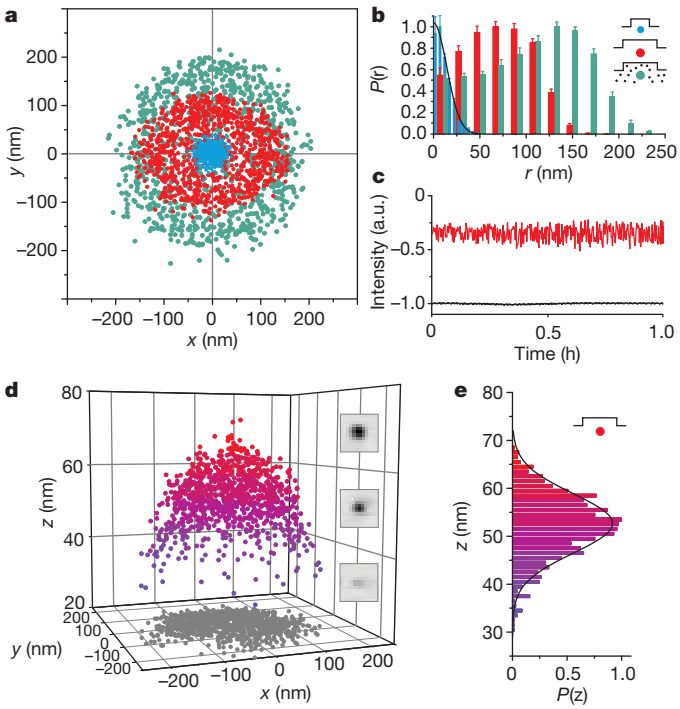


Figure 2 | Transverse and axial confinement of a 100-nm gold particle by pocket nanostructures. **a**, Transverse (x–y) positions of single 100-nm gold particles trapped by pockets of diameter $D = 500$ nm (red) and 200 nm (blue) in 7×10^{-5} M electrolyte and $D = 500$ nm in 0.1 mM NaCl (green). **b**, Radial probability histograms averaged over about ten different trapped particles in each case. $P(r)$ was defined according to $\int P(r) 2\pi r \, dr = 1$. The displayed distribution was rescaled so that $P_{max}(r) = 1$ (error bars, s.e.m.). A Gaussian fit to $P(r)$ for $D = 200$ nm (black curve) yields a radial trap stiffness of $k_r = 0.02$ pN nm^{-1}. **c**, Intensity–time traces for a single particle trapped by a $D = 500$ nm pocket in 7×10^{-5} M electrolyte (red) and for the background (black). The intensity fluctuation of the particle reflects its motion in z (a.u., arbitrary units). **d**, Three-dimensional scatter plot of a particle confined by a $D = 500$ nm pocket in 7×10^{-5} M electrolyte (z scale exaggerated). Insets, iSCAT images of the particle highlighting the modulation of contrast at different axial positions. **e**, Histogram of axial displacement revealing a mean levitation height of $z = 52 \pm 2$ nm and an axial trap stiffness of $k_z = 0.027$ pN nm^{-1}.

In addition to the lateral particle motion, we observe strong fluctuations in the iSCAT signal of a single particle as a function of time (Fig. 2c, red trace) that are orders of magnitude larger than background fluctuations (black trace). These intensity fluctuations are directly correlated with $\Delta\varphi$ and, therefore, with the axial motion of the particle. This correlation allows a complete spatial characterization of particle motion in the x, y and z directions (Methods), as illustrated in Fig. 2d for the 100-nm gold particle indicated by red symbols in Fig. 2a, b. The corresponding histogram of axial displacements, $P(z)$ (Fig. 2e), illustrates the exceptionally strong axial confinement and levitation of the particle within a region of only 20 nm about the $z = 52 \pm 2$ nm plane. A Gaussian fit to $P(z)$ yields an estimate of $k_z = 0.027$ pN nm^{-1} for the trap stiffness in the z direction.

The blue data in Fig. 2a, b illustrate the confinement of a 100-nm gold particle after reducing the diameter of the pocket-shaped trap from $D = 500$ nm to 200 nm, and verify that the trap range scales down with D (Supplementary Movie 5). The nanoparticle is now laterally confined to within 20 nm of the trap centre, with a radial stiffness of $k_r = 0.02$ pN nm^{-1}, which is comparable to the best attainable value in an optical trap[16].

The trapping may be understood by considering a fluid element of rectangular cross-section surrounded on all sides by negatively charged surfaces, which in our experiment arise from the spontaneous ionization of surface chemical groups on the slit walls and the particles on contact with water[17]. Zeta-potential measurements by means of light scattering and electro-osmotic flow experiments give estimated charge densities for the gold particles and slit surfaces of -0.06 μC cm^{-2} and -0.16 μC cm^{-2}, respectively (which for the particle translates into an effective charge of about $-120e$, where e is the elementary charge). The electrostatic potential near a charged surface in contact with a solution of ionic strength C decays roughly exponentially, with a characteristic Debye length $\kappa^{-1} \propto C^{-1/2}$ (ref. 17). If two such surfaces are placed close to each other, a negatively charged particle between them experiences an energy minimum at the slit midplane, and its motion is thereby confined in the depth dimension (see region I in Fig. 3a). If the separation, $2h$, between the two surfaces is locally increased by a depth, d, which is of the order of κ^{-1}, the absolute value of the potential minimum drops in that region (see region II in Fig. 3a). This results in a local potential energy modulation of $q\Delta U$ in the x–y plane, where q is the particle charge and ΔU is the potential depth. If $q\Delta U$ is substantially larger than the thermal energy, $k_B T$ (k_B, Boltzmann constant), the particle is stably trapped for a period given by the Kramers time, $\tau = \tau_0 \exp(q\Delta U/k_B T)$, in the harmonic potential approximation[18]. Here τ_0 denotes the time that an untrapped particle would take to diffuse across a distance corresponding to the width of the potential well[19], which may be estimated as the radius of the pocket. We note that values of $q\Delta U \approx 10 k_B T$ over a range of 200 nm are equivalent to trapping forces of about 0.2 pN, suggesting that a trapped 100-nm particle can withstand Stokes drag forces exerted by fluid flow velocities of up to about 200 μm s^{-1}.

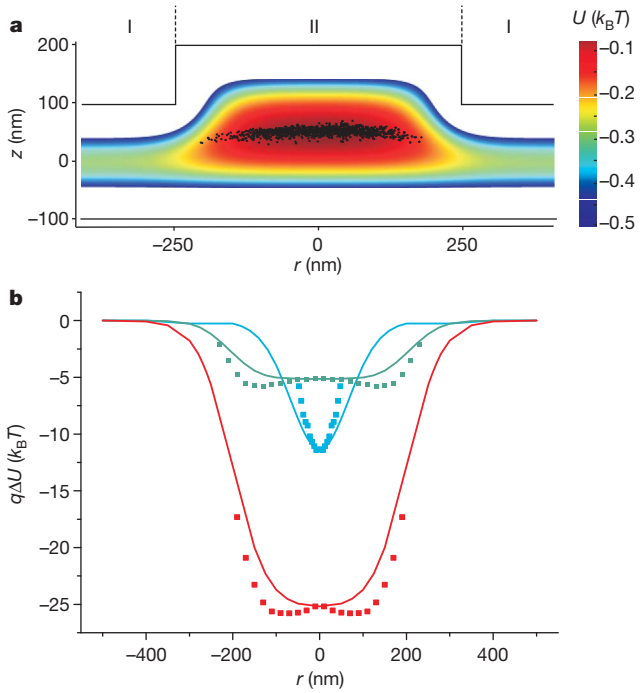


Figure 3 | Electrostatic potential in a topographically structured fluidic nanoslit. **a**, Two-dimensional electrostatic potential obtained by solving the Poisson–Boltzmann equation for an axisymmetric pocket of diameter $D = 500$ nm and depth $d = 100$ nm, for $C = 7 \times 10^{-5}$ M electrolyte in a slit of height $2h = 200$ nm. The experimental data of Fig. 2d is presented to scale as black symbols overlaid on the calculated potential distribution. **b**, Electrostatic energy of a point charge of $q = -120e$ as a function of r along a contour of minimum axial (z) electrostatic energy. Red curve: $D = 500$ nm, $C = 0.07$ mM; blue curve: $D = 200$ nm, $C = 0.07$ mM; green curve: $D = 500$ nm in 0.1 mM NaCl (total ionic strength, $C = 0.17$ mM). Overlaid squares represent the local potentials $u(r)$ derived from the experimentally measured $P(r)$ from Fig. 2b. Experimental data for large r are not provided because the probability of finding the trapped particle far from the potential minimum falls off exponentially, leading to rare occurrences.

Figure 3a shows the electrostatic potential calculated in the mean-field approach by numerically solving the two-dimensional Poisson–Boltzmann equation[20,21] (Methods) for a pocket trap with $D = 500$ nm, $d = 100$ nm, $2h = 200$ nm and $C = 7 \times 10^{-5}$ M. An overlay of the experimental data from Fig. 2d on the calculated potential distribution reveals an excellent agreement between the mean levitation height of the object and the theoretically expected plane of the potential minimum at $z = 50$ nm. We point out that in our experiments the locally non-zero space charge density in the fluid allows for the realization of a stable potential minimum in a static electric field. This is in contrast to the situation of ion traps in vacuum, where the Earnshaw theorem forbids the existence of a local minimum[22].

The theoretical considerations discussed above predict that the range and residence time of the trap can be tuned by adjusting the pocket diameter (D) and depth (d), the slit height ($2h$) and the solution's ionic strength (C). The solid traces in Fig. 3b are the radial electrostatic energy profiles for a point charge $q = -120e$ (refs 15, 23) calculated for three different systems. The red and blue curves respectively deal with pockets of diameter $D = 500$ and 200 nm in $C = 7 \times 10^{-5}$ M solution, and confirm the observations of Fig. 2b that a smaller pocket results in a stiffer trap. We also note that the trap becomes shallower, implying shorter trapping times for a given object charge. In the experiment, trapping times for a 100-nm gold particle decreased from several hours for $D = 500$ nm pockets to the order of one minute for $D = 200$ nm. To compare experiment and theory more quantitatively, we deduce local potentials, $u(r)$, from the measured $P(r)$ values of Fig. 2b using the Boltzmann relation $u(r) = -\ln(P(r))$, and represent these as symbols in Fig. 3b. Note that the potentials deduced from experiment were overlaid on the corresponding theoretical curves by superimposing their values at $r = 0$ without any fitting, with the fair agreement between calculation and experiment emphasizing that our simple theoretical model effectively captures the critical features of the trapping phenomenology. We remark that previous observations on confinement of charged nanometre-sized objects in an unstructured fluidic slit[24], although apparently similar, cannot be explained by consideration of the local potential due to the structure alone[25].

In addition to the trap geometry, the ionic strength of the fluidic environment serves as a further parameter that modulates trap performance. A solution with higher ionic strength screens charges on the slit walls more strongly and renders the trap not only shallower but also more compliant. The scatter plot of particle positions displayed by the green data in Fig. 2a and the corresponding histogram of $P(r)$ in Fig. 2b illustrate this prediction for a 0.1 mM NaCl solution. Here, the particle's residence time in the trap was on the order of minutes, compared with several hours in the absence of NaCl. As before, the potential deduced from the experimental observations agrees with our model prediction (green symbols and curve in Fig. 3b). However, the slightly radially skewed nature of $P(r)$ (see red and green series in Fig. 3b) reveals some deviation between the calculated electrostatic potential and the apparent potential experienced by the particles in these experiments, which could stem from additional geometry-related effects such as correlated fluctuations in the counterion distributions[26]. Contributions from the object's electric field and its finite volume, as well as variable surface charge densities, would need to be considered for a complete quantitative description of the trap behaviour.

In addition to gold nanoparticles, pocket traps with $D = 500$ nm also retained single polystyrene nanospheres with diameters as small as 20 nm and charges of about $-26e$ for periods of about one minute. We then examined trapping of lipid vesicles, which could act as nanometre-sized chambers encapsulating proteins or other macromolecules under physiological conditions. Figure 4a shows a fluorescence image of single vesicles with a diameter of 50 nm and charge of $-85e$ in an array of pocket traps (Supplementary Movie 6), and Fig. 4b and Fig. 4c respectively present scatter plots and histograms of $P(r)$ for vesicles in traps with $D = 200$ and 500 nm. A Gaussian fit to the $D = 200$ nm trap data yields a radial trap stiffness of $k_r = 0.01$ pN nm^{-1}, which is

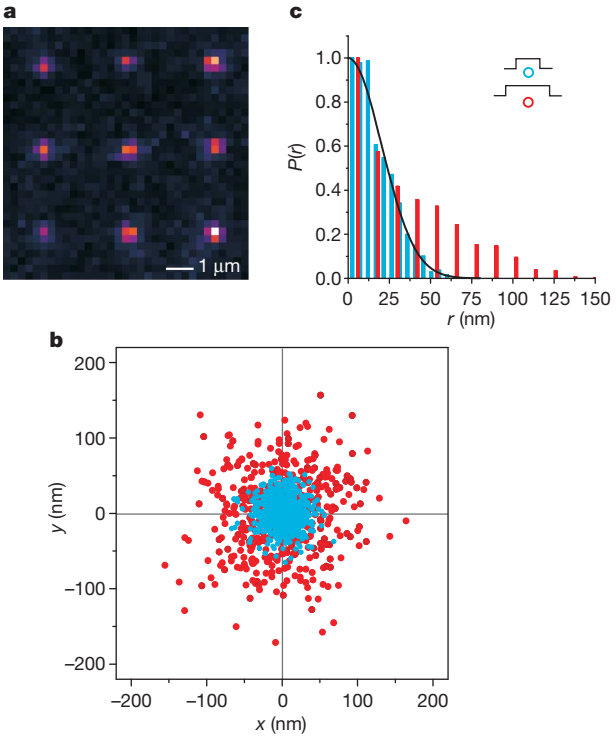


Figure 4 | Trapping single lipid vesicles. **a**, Image of single fluorescent negatively charged vesicles trapped in arrays. **b**, Transverse spatial sampling of single vesicles trapped by $D = 500$ nm (red) and $D = 200$ nm (blue) pockets in 4×10^{-5} M electrolyte (data from five different vesicles overlaid in each case). **c**, Radial probability density histograms $P(r)$ for the data shown in **b**. A Gaussian fit to $P(r)$ for $D = 200$ nm (black curve) yields a radial trap stiffness of $k_r = 0.01$ pN nm^{-1}.

comparable to the value obtained with a 100-nm gold nanoparticle under similar conditions. Owing to their small size and low polarizability in water, stable trapping of aqueous lipid vesicles has not been feasible using optical forces.

Our ability to stably trap vesicles, polymer spheres and gold nanoparticles carrying comparable total charges confirms that the key parameter that determines the performance of a given trap is the total electric charge of the object and not its mass, volume or dielectric function. In principle, our trapping concept should therefore allow contact-free trapping of objects as small as single proteins, provided that they are sufficiently charged. Many proteins such as calsequestrin ($q \approx -80e$) and GroEL complex ($q \approx -250e$) carry a net negative charge at neutral pH. However, this leaves the challenge that solutions with higher ionic strengths (that is, large κ), as required for biological studies, reduce the trap depth (Fig. 3b). In this regard, we note that the dependence of the potential depth on the dimensionless parameter κh (see equation (3) in Methods) implies that reducing h can recover effective trapping. For example, slits of height $2h \approx 10$ nm and a solution with $C = 30$ mM give $\kappa h \approx 3$ (the range investigated here); this should allow the trapping of a protein of charge $q = -50e$ for about 1 min, which could open doors to new well-controlled investigations at the single-molecule level[27].

The pocket and groove structures can be used to create a range of trap morphologies with different functionalities. Elongated trap geometries could confine and align anisotropic objects such as ellipsoids and nanorods[28,29], and dilute suspensions of particles could be assembled into high-density lattices by flowing them past trap arrays (as illustrated in Supplementary Fig. 1 with rewriteable arrays of gold nanoparticles at a lattice spacing of 500 nm). Such a passive trapping platform could be further combined with local or global external stimuli—such as fluid flow[30], electric or magnetic fields, or optical forces—for targeting and addressing the trapped objects on command. In this

fashion, it might even be possible to fractionate mixtures of charged macromolecules into spatially discrete locations on a chip for further manipulation.

METHODS SUMMARY

We defined trenches 20 μm wide and 200 nm deep in silicon dioxide using optical lithography and wet etching of the oxide. Electron-beam lithography followed by reactive ion etching was used to introduce nanoscale features on the floors of these trenches. Fully functional fluidic slits were created by anodically bonding the processed silicon/silicon dioxide substrates with microscopy-compatible cover glass. We loaded nanoslits with an aqueous suspension of the nanometre-sized object of interest by the capillary effect as previously described[24], and allowed the device to equilibrate at room temperature (298 K) for 1–2 h before making optical measurements. Gold particles were imaged using iSCAT, and fluorescent nanospheres and vesicles were imaged using fluorescence microscopy. Measurements of solution conductivity and particle zeta potential were obtained using commercial light scattering instrumentation (Zetasizer Nano, Malvern Instruments). The value of the zeta potential for the slit surfaces was estimated from electro-osmotic flow measurements in fluidic slits. We calculated the two-dimensional electrostatic potential distribution in the fluidic slit by solving the Poisson–Boltzmann equation $\nabla^2 U = (\kappa h)^2 \sinh(U)$ in cylindrical coordinates for an axisymmetric pocket structure in a slit using constant-charge boundary conditions for the slit walls and verifying overall electroneutrality.

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

Received 23 March; accepted 4 August 2010.

1. Chu, S. Cold atoms and quantum control. *Nature* **416,** 206–210 (2002).
2. Grier, D. G. A revolution in optical manipulation. *Nature* **424,** 810–816 (2003).
3. Ashkin, A. Acceleration and trapping of particles by radiation pressure. *Phys. Rev. Lett.* **24,** 156–159 (1970).
4. Gosse, C. & Croquette, V. Magnetic tweezers: micromanipulation and force measurement at the molecular level. *Biophys. J.* **82,** 3314–3329 (2002).
5. Chiou, P. Y., Ohta, A. T. & Wu, M. C. Massively parallel manipulation of single cells and microparticles using optical images. *Nature* **436,** 370–372 (2005).
6. Soyka, F., Zvyagolskaya, O., Hertlein, C., Helden, L. & Bechinger, C. Critical Casimir forces in colloidal suspensions on chemically patterned surfaces. *Phys. Rev. Lett.* **101,** 208301 (2008).
7. Cordovez, B., Psaltis, D. & Erickson, D. Electroactive micro and nanowells for optofluidic storage. *Opt. Express* **17,** 21134–21148 (2009).
8. Cohen, A. E. & Moerner, W. E. Suppressing Brownian motion of individual biomolecules in solution. *Proc. Natl Acad. Sci. USA* **103,** 4362–4365 (2006).
9. Yang, A. H. J. *et al.* Optical manipulation of nanoparticles and biomolecules in sub-wavelength slot waveguides. *Nature* **457,** 71–75 (2009).
10. Juan, M. L., Gordon, R., Pang, Y. J., Eftekhari, F. & Quidant, R. Self-induced back-action optical trapping of dielectric nanoparticles. *Nature Phys.* **5,** 915–919 (2009).
11. Jo, K. *et al.* A single-molecule barcoding system using nanoslits for DNA analysis. *Proc. Natl Acad. Sci. USA* **104,** 2673–2678 (2007).
12. Reisner, W., Larsen, N. B., Flyvbjerg, H., Tegenfeldt, J. O. & Kristensen, A. Directed self-organization of single DNA molecules in a nanoslit via embedded nanopit arrays. *Proc. Natl Acad. Sci. USA* **106,** 79–84 (2009).
13. Kukura, P. *et al.* High-speed nanoscopic tracking of the position and orientation of a single virus. *Nature Methods* **6,** 923–927 (2009).
14. Jacobsen, V., Klotzsch, E. & Sandoghdar, V. in *Nano Biophotonics* (eds Masuhara, H., Kawata, S. & Tokunaga, F.) 143–160 (Elsevier, 2007).
15. Eichmann, S. L., Anekal, S. G. & Bevan, M. A. Electrostatically confined nanoparticle interactions and dynamics. *Langmuir* **24,** 714–721 (2008).
16. Hansen, P. M., Bhatia, V. K., Harrit, N. & Oddershede, L. Expanding the optical trapping range of gold nanoparticles. *Nano Lett.* **5,** 1937–1942 (2005).
17. Israelachvili, J. *Intermolecular and Surface Forces* (Academic, 1992).
18. Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7,** 284–304 (1940).
19. Kaplan, P. D., Faucheux, L. P. & Libchaber, A. J. Direct observation of the entropic potential in a binary suspension. *Phys. Rev. Lett.* **73,** 2793–2796 (1994).
20. Tessier, F. & Slater, G. W. Effective Debye length in closed nanoscopic systems: a competition between two length scales. *Electrophoresis* **27,** 686–693 (2006).
21. Dubois, M. *et al.* Osmotic pressure and salt exclusion in electrostatically swollen lamellar phases. *J. Chem. Phys.* **96,** 2278–2286 (1992).
22. Maxwell, J. C. *A Treatise on Electricity and Magnetism* Vol. 1, 116 (Dover, 1954).
23. Adamczyk, Z. & Warszynski, P. Role of electrostatic interactions in particle adsorption. *Adv. Colloid Interface Sci.* **63,** 41–149 (1996).
24. Krishnan, M., Monch, I. & Schwille, P. Spontaneous stretching of DNA in a two-dimensional nanoslit. *Nano Lett.* **7,** 1270–1275 (2007).
25. Krishnan, M., Petrasek, Z., Moench, I. & Schwille, P. Electrostatic self-assembly of charged colloids and macromolecules in a fluidic nanoslit. *Small* **4,** 1900–1906 (2008).
26. Polin, M., Grier, D. G. & Han, Y. Colloidal electrostatic interactions near a conducting surface. *Phys. Rev. E* **76,** 041046 (2007).
27. Lu, H. P., Xun, L. Y. & Xie, X. S. Single-molecule enzymatic dynamics. *Science* **282,** 1877–1882 (1998).
28. Stebe, K. J., Lewandowski, E. & Ghosh, M. Oriented assembly of metamaterials. *Science* **325,** 159–160 (2009).
29. Jamshidi, A. *et al.* Dynamic manipulation and separation of individual semiconducting and metallic nanowires. *Nature Photon.* **2,** 85–89 (2008).
30. MacDonald, M. P., Spalding, G. C. & Dholakia, K. Microfluidic sorting in an optical lattice. *Nature* **426,** 421–424 (2003).

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

Acknowledgements We thank A. Alt for expert assistance in electron-beam lithography and the staff of the Center for Micro- and Nanoscience (FIRST), ETH Zurich, for clean-room support; H. Ewers for assistance in producing lipid vesicles; M. Celebrano and A. Renn for experimental input; M. Agio for discussions; and E. Petrov and R. Klemm for reading the manuscript. M.K. acknowledges a Marie Curie Intra-European Fellowship for Career Development of the European Commission, and greatly appreciates the support of B. Eichler, I. Mönch and P. Schwille in the early stages of the project.

Author Contributions M.K. conceived the research, designed the experiments and performed the theoretical analysis. N.M. and M.K. performed the experiments and analysed the data. P.K. and V.S. provided expertise on iSCAT. M.K. and V.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.K. (mkrishnan@ethz.ch).

METHODS

Sample preparation. Samples of 100-nm gold nanospheres (British Biocell International) and 100-nm-diameter fluorescent polystyrene spheres (Invitrogen) were centrifuged and resuspended in deionized H_2O (18 MΩ cm^{-1}) twice to remove traces of salt or other contaminants. Fluorescent polystyrene beads (Invitrogen) 20 nm in diameter were diluted 1:100 in deionized H_2O. Small unilamellar vesicles were prepared by extrusion through 50-nm polycarbonate membranes (Avestin). The lipid mixture containing 86.6 mol% di-oleoyl-phosphatidyl-choline (DOPC, Avanti Polar Lipids), 13 mol% trisialoganglioside (GT1b, Hytest) and 0.4 mol% BODIPY-FL-GM1 (Invitrogen) was dried to a film under clean, dry nitrogen and rehydrated for at least 2 h in deionized H_2O (18 MΩ cm^{-1}). GT1b contains three sialic-acid groups and makes the vesicles strongly negatively charged; DOPC is zwitterionic and does not contribute to the net charge. The final ionic strength of the vesicle suspension used in the experiments was 7×10^{-5} M, determined from conductivity measurements.

Device fabrication. A single device consists of several fluidic slits in parallel, with each slit 20 μm wide and 200 nm deep. The slits were fabricated by lithographically patterning the surface of a 1,000-nm-deep silicon dioxide layer on a p-type silicon substrate and subsequently wet etching the silicon dioxide surface to a depth of 200 nm in buffered HF (ammonium fluoride/HF mixture, Sigma-Aldrich). The surfaces of these 200-nm-deep trenches were then patterned with the required submicrometre-scale features using electron-beam lithography and subsequently etching the exposed silicon dioxide to a depth of 100 nm using reactive ion etching. Fully functional fluidic slits were obtained by irreversibly bonding the processed silicon dioxide/silicon substrates with glass substrates compatible with high-numerical-aperture (high-NA) microscopy (PlanOptik, AG) using field-assisted bonding. Nanoslits loaded with an aqueous suspension of the nanometric object of interest (number density, $\sim 10^{10}$ particles per millilitre for gold particles) as previously described[24] were allowed to equilibrate at room temperature (298 K) for 1–2 h before commencing with optical measurements.

Scattering interferometry (iSCAT). Given the rapid dynamics of nano-objects in solution (diffusion coefficient, $>1\ \mu m^2\ s^{-1}$), tracking their motion with high spatiotemporal resolution implies a static three-dimensional localization accuracy of <5 nm in an exposure time of ≤ 1 ms, which is difficult to achieve with standard fluorescence-based single-particle tracking. We therefore imaged the motion of 100-nm gold particles in our experiments by iSCAT.

Gold particles (100 nm) were imaged by iSCAT. A 30-mW diode-pumped solid-state laser (TECGL-30, WSTech) at $\lambda = 532$ nm, near the plasmon resonance wavelength of the particle, was deflected by a two-axis acousto-optical deflector (DTSXY, AA Opto-Electronic), scanning the beam at a rate of 50 kHz. The deflected beam was focused by a ×60, 0.8 NA air objective (Olympus) onto the sample, illuminating a field of area 30–480 μm^2. The scattered and reflected beams were collected by the objective and imaged on a CMOS camera (MV-D1024-160-CL-12, PhotonFocus) at an acquisition rate of 1 kHz. In iSCAT, interference between light scattered from the illuminated particle, $\mathbf{E}_{sca}$, and the reflected (reference) beam, $\mathbf{E}_{ref}$, results in an intensity at the detector which may be expressed as

$$I_{det} = |\mathbf{E}_{ref} + \mathbf{E}_{sca}|^2 = |\mathbf{E}_{ref}|^2 + |\mathbf{E}_{sca}|^2 + 2|\mathbf{E}_{ref}||\mathbf{E}_{sca}|\cos(\Delta\varphi) \quad (1)$$

The first term is the intensity of the reference beam, the second is the scattered intensity from the particle and the third is the interference term. If $|\mathbf{E}_{sca}| \ll |\mathbf{E}_{ref}|$, equation (1) may be normalized to give the contrast

$$I_p \approx 1 + 2\frac{|\mathbf{E}_{sca}|}{|\mathbf{E}_{ref}|}\cos(\Delta\varphi) \quad (2)$$

where $\Delta\varphi = 4\pi l/\lambda + \varphi_0$ is the phase difference between the scattered and reference fields. Here l, the optical path difference between the object and the reference beam, is given by $l = n_w l_w + n_s l_s$, where n_i and l_i respectively represent the refractive index and physical path length in medium i (w: H_2O; s: SiO_2), and φ_0 is an intrinsic phase that contains the Gouy phase, the retardation phase due to the scatterer and the phase introduced by reflection from a high-index material (silicon in our case). The contrast in the image or signal-to-background ratio, I_p, is thus a function of the axial location of the object relative to the plane of reflection of the laser beam in the sample, the silicon dioxide/silicon interface, as shown in Fig. 1d. Because the image at the detector is a superposition of two components (the image of the particle and that of its trapping pocket), we subtracted the iSCAT image of an empty pocket from the raw time series of images of particle motion. Two-dimensional Gaussian fits to the resulting intensity distributions give the full positional information on the particle. The centre of the Gaussian directly yields the particle position in the x–y plane. The amplitude of the Gaussian I_p was converted to a relative height measurement using equation (2). Knowledge of the maximum expected interferometric contrast, $I_{max} = 2|\mathbf{E}_{sca}|/|\mathbf{E}_{ref}|$, attained by a single particle and the value of ϕ_0 were necessary for this. These values were obtained using a separate calibration procedure. The average contrast of reference particles whose axial position in the slit is known a priori was also necessary to map the height displacements of the particle to absolute z values within the slit. Particles statistically stuck on the silicon dioxide surface in the vicinity of the trapping pockets served as convenient reference objects in the axial tracking process. Static localization accuracies were <4 nm in the x–y plane and ~ 2 nm in the z direction.

Fluorescence imaging. Fluorescent polystyrene particles (20 nm) and fluorescence-labelled lipid vesicles 50 nm in diameter were imaged in a standard wide-field configuration using 473-nm excitation and a 520 LP filter. Images were acquired at the rate of 10–20 Hz using a ×100, 1.2 NA oil immersion objective (Leitz) and an EMCCD camera (iXon^{EM} Model 860, Andor). The total observation time was limited by photobleaching to less than one minute. Localization accuracies were 10 nm for vesicles and 25 nm for 20-nm beads.

Calculation of electrostatic potentials using the Poisson–Boltzmann equation. The normalized Poisson–Boltzmann equation for a two-dimensional system in cylindrical coordinates maybe written as

$$\frac{\partial^2 U}{\partial R^2} + \frac{1}{R}\frac{\partial U}{\partial R} + \frac{\partial^2 U}{\partial Z^2} = (\kappa h)^2 \sinh(U) \quad (3)$$

which is valid for an open system in osmotic equilibrium with electrolyte reservoirs in a symmetric 1:1 electrolyte[20,21]. Here $U = e\psi/k_BT$ is the dimensionless local electrostatic potential (ψ), $2h$ represents the height of the nanoslit, $R = r/h$, $Z = z/h$ and $\kappa^{-1} = (2Ce^2/\varepsilon\varepsilon_0 k_B T)^{-0.5}$, the Debye length, provides a measure of the range of the electrostatic interactions. In the expression for the Debye length, C is the concentration of dissolved salt in the solution, ε is the dielectric constant of the medium, ε_0 is the permittivity of free space, k_B is the Boltzmann constant and T is the absolute temperature. Equation (3) was solved numerically using the COMSOL Multiphysics package (COMSOL) for the case of a single disc-shaped pocket symmetric about the $r = 0$ axis in a nanoslit (Fig. 3a). Constant-charge boundary conditions were used for the slit walls, $\mathbf{n}\bullet\nabla U = (e^2/\varepsilon\varepsilon_0 k_B T)\sigma_s h$, where $\mathbf{n}$ is the inward-pointing normal to the slit surface and $\sigma_s = 10^{16} e\ m^{-2}$ is the number density of negative charges on the slit surfaces, estimated from our zeta-potential measurements and in good agreement with the literature[31]. Electroneutrality in the slit was verified by an integration procedure summing all the charges in the system, which showed that the deviation from the equality expressing electroneutrality in the slit, $\langle C^+\rangle - \langle C^-\rangle = \sigma_s/h$, was less than 0.1%. Here $\langle C^+\rangle$ and $\langle C^-\rangle$ are respectively the mean concentrations of positive and negative ions in solution in the slit[20].

Surface charge and solution conductivity measurements. Solution conductivities as well as particle and vesicle zeta potentials were measured by phase analysis light scattering using commercial instrumentation (Zetasizer Nano, Malvern Instruments). The measured zeta potential, ζ, was used to arrive at an estimate of particle surface charge density, σ_p, in coulombs per square metre using the semi-empirical equation

$$\sigma_p = -\varepsilon\varepsilon_0\kappa\left(\frac{k_BT}{je}\right)\left[2\sinh\left(\frac{jy}{2}\right) + \left(\frac{4}{\kappa a}\right)\tanh\left(\frac{jy}{4}\right)\right]$$

proposed in ref. 32, where $y = \zeta/k_BT$ is the dimensionless zeta potential, $j = 1$ is the valence of the counterions and a is the radius of the particle. Measured zeta-potential values were as follows: 100-nm gold, -18 ± 3 mV; 20-nm polystyrene, -40 ± 2 mV; 50-nm lipid vesicles, -40 ± 5 mV.

31. Behrens, S. H. & Grier, D. G. The charge of glass and silica surfaces. *J. Chem. Phys.* **115,** 6716–6721 (2001).
32. Loeb, A. L., Wiersma, P. H. & Overbeek, J. T. G. *The Electrical Double Layer Around a Spherical Colloidal Particle* (MIT Press, 1961).

doi:10.1038/nature09426

An influence of solar spectral variations on radiative forcing of climate

Joanna D. Haigh[1], Ann R. Winning[1], Ralf Toumi[1] & Jerald W. Harder[2]

The thermal structure and composition of the atmosphere is determined fundamentally by the incoming solar irradiance. Radiation at ultraviolet wavelengths dissociates atmospheric molecules, initiating chains of chemical reactions—specifically those producing stratospheric ozone—and providing the major source of heating for the middle atmosphere, while radiation at visible and near-infrared wavelengths mainly reaches and warms the lower atmosphere and the Earth's surface[1]. Thus the spectral composition of solar radiation is crucial in determining atmospheric structure, as well as surface temperature, and it follows that the response of the atmosphere to variations in solar irradiance depends on the spectrum[2]. Daily measurements of the solar spectrum between 0.2 μm and 2.4 μm, made by the Spectral Irradiance Monitor (SIM) instrument on the Solar Radiation and Climate Experiment (SORCE) satellite[3] since April 2004, have revealed[4] that over this declining phase of the solar cycle there was a four to six times larger decline in ultraviolet than would have been predicted on the basis of our previous understanding. This reduction was partially compensated in the total solar output by an increase in radiation at visible wavelengths. Here we show that these spectral changes appear to have led to a significant decline from 2004 to 2007 in stratospheric ozone below an altitude of 45 km, with an increase above this altitude. Our results, simulated with a radiative-photochemical model, are consistent with contemporaneous measurements of ozone from the Aura-MLS satellite, although the short time period makes precise attribution to solar effects difficult. We also show, using the SIM data, that solar radiative forcing of surface climate is out of phase with solar activity. Currently there is insufficient observational evidence to validate the spectral variations observed by SIM, or to fully characterize other solar cycles, but our findings raise the possibility that the effects of solar variability on temperature throughout the atmosphere may be contrary to current expectations.

The peak of the most recent '11-year' solar cycle (identified as number 23) occurred 2000–2002, and from then until about December 2009 the Sun's activity declined. Figure 1 shows the difference between 2004 and 2007 in solar spectral irradiance measured by SIM. This is quite unlike that predicted by multi-component empirical models, based on activity indicators such as sunspot number and area, as exemplified by that of Lean[5] (also shown in Fig. 1). The SIM data indicate a decline in ultraviolet from 2004 to 2007 that is a factor of 4 to 6 larger than in the Lean data and an increase in visible radiation, compared with a small decline in the Lean data. Other empirical models[6,7] show larger-amplitude variations in the near-ultraviolet than does the Lean model but none reflect the behaviour apparent in the SIM data. Also shown in Fig. 1, for wavelengths 116–290 nm, are independent measurements made by the Solar Stellar Irradiance Comparison Experiment (SOLSTICE) instrument on SORCE. The data from SIM and SOLSTICE both indicate substantially more ultraviolet variability than does the Lean model. SIM calibration, and instrument comparisons, are discussed in detail in ref. 8.

To investigate how these very different spectral changes might affect the stratosphere, experiments have been carried out using a two-dimensional (latitude-height) radiative-chemical-transport model of the atmosphere[9]. This model includes detailed representations of photochemistry and radiative transfer and has been used in many studies involving radiation-chemistry interactions[10,11]. (See Supplementary Information for further details.) This type of model produces realistic simulations of the upper stratosphere (above about 25 km) but is less reliable at lower altitudes where photochemical time constants are longer and a more accurate representation of transport processes is required. The results below come from four model runs using solar spectra derived from the SIM measurements (with SOLSTICE data for wavelengths less than 200 nm) and those produced by the Lean model, each for both 2004 and 2007.

In Fig. 2 we present latitude–height maps of the difference between 2004 and 2007 in December ozone concentrations. The Lean spectral data produce a broad structure of ozone concentrations greater in 2004 than in 2007, with maximum values of around 0.8% near 40 km, whereas the SIM data produce a peak enhancement of over 2% in low latitudes around 35 km, along with significant reductions above 45 km. The predicted temperature differences (Supplementary Fig. 1) are also very different, with the Lean data set showing temperatures 0.3–0.4 K greater in 2004 than in 2007 at the top of the model domain, whereas the SIM data set produces a peak warming of 1.8 K at the summer polar stratopause. These temperature differences are qualitatively similar to, but about 50% larger than, those estimated by ref. 12 with an idealized forcing in a full climate model, possibly owing to the broader spectral

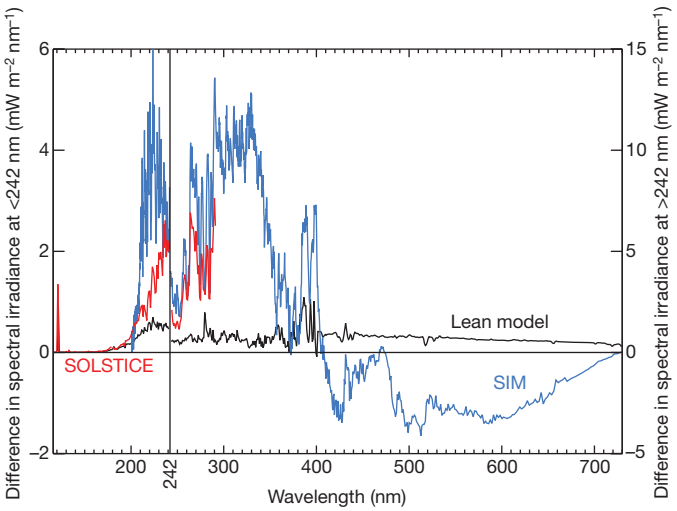


Figure 1 | Difference in solar spectrum between April 2004 and November 2007. The difference (2004–2007) in solar spectral irradiance (W m^{-2} nm^{-1}) derived from SIM data[4] (in blue), SOLSTICE data[8] (in red) and from the Lean model[5] (in black). Different scales are used for values at wavelengths less and more than 242 nm (see left and right axes respectively).

[1]Blackett Laboratory, Imperial College London, London SW7 2AZ, UK. [2]Laboratory for Atmospheric and Space Physics, University of Colorado, 1234 Innovation Drive, Boulder, Colorado 80303-7814, USA.

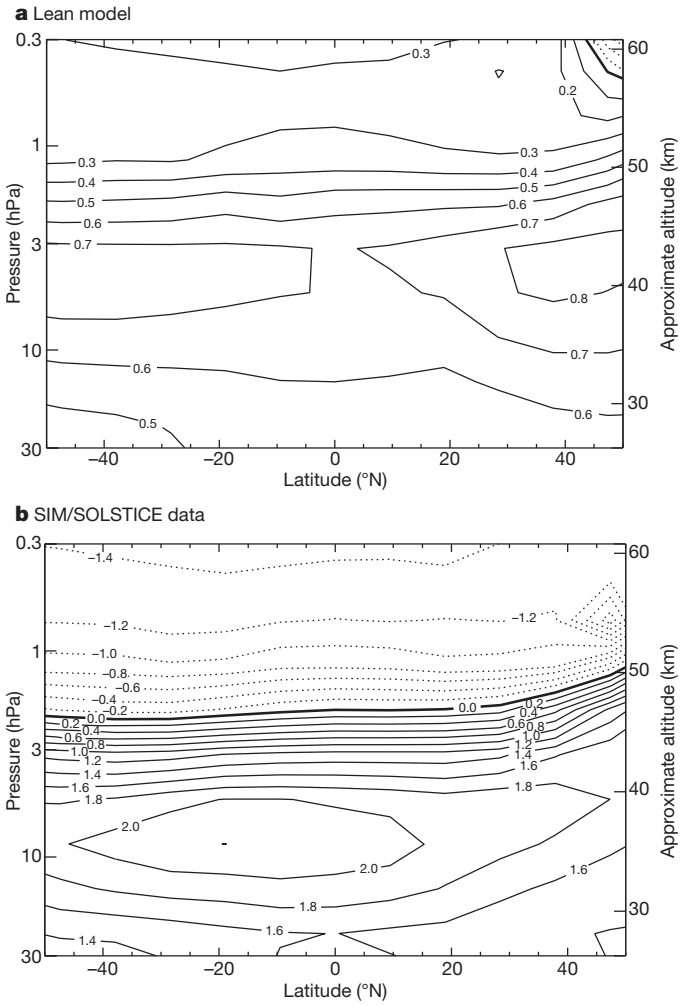


Figure 2 | Modelled difference in ozone between December 2004 and December 2007. Estimates of the percentage difference (2004–2007) in zonal mean ozone concentration (labels on contour lines in per cent) produced by the model using solar spectra from the Lean model (**a**) and SIM/SOLSTICE data (**b**).

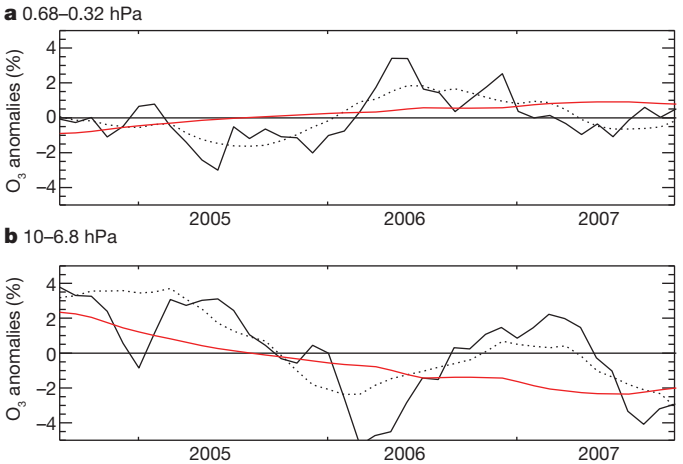


Figure 3 | Time series of AURA-MLS v2.2 ozone concentrations. The data (solid black lines) are percentage anomalies of tropical (22.5 °S–22.5 °N) deseasonalized monthly means from August 2004 to November 2007. The values reconstructed from the 4-component regression model are shown as dashed lines. The solar component of the regression is shown in red. Other components are shown, along with the solar component, in Supplementary Fig. 2. Data were averaged between 0.68 hPa and 0.32 hPa (**a**) and 10 hPa and 6.8 hPa (**b**).

resolution imposed and the lack of ozone–temperature feedback in that model version.

The very different scenarios produced by the two spectral data sets suggest they might be distinguishable in observational records. A multiple regression analysis has been carried out of deseasonalized monthly mean ozone data from the Microwave Limb Sounder (MLS) instrument on the Earth Observing System (EOS) Aura satellite. Four regression indices were used: a constant, two orthogonal indices representing the quasi-biennial oscillation (which dominates ozone variability in the tropical stratosphere)[13] and a solar index constructed from SIM data integrated over 200–400 nm. Motivated by the model results (Fig. 2), we chose two spatial regions, both spanning the tropics, one at altitude 10–6.8 hPa, where the model predicts the largest difference 2004–2007, and one at 0.68–0.32 hPa, where the model shows largest negative values. Figure 3 shows the raw data and the fits reconstructed from the four regression components; it also shows (in red) the derived solar component, which is statistically significant at >95% at the upper levels and >99% at the lower levels (see Supplementary Information).

Over the period from the late 1970s to the late 1990s tropical ozone at altitudes 35–50 km decreased by about 9% (ref. 13) in response to increasing concentrations of active chlorine species. Since about 2000, however, the trend in chlorine has reversed and ozone has stopped declining. Stratospheric cooling by greenhouse gases has probably also contributed to the ozone trend reversal by slowing the chemical reactions that destroy it[14]. Over the short period of the present study it is not possible statistically to differentiate these factors from each other, or from any solar influence. Nevertheless, it seems likely that the Sun is important in the apparent decrease in ozone below 45 km from 2004 to 2007. The change in sign near 45 km is also more consistent with the modelled response to the SIM spectral variations than to the Lean spectra. Previous analyses[13,15] of the solar signal in ozone, averaged over approximately 2.5 solar cycles (1979 to 2005 or 2003), have not shown this structure. This suggests that the declining phase of solar cycle 23 is behaving differently to previous solar cycles or possibly that the solar cycle exhibits different behaviours during its ascending and descending phases.

To understand the different spatial structures, and magnitudes, of the modelled ozone responses we consider photochemical processes. The sharp decrease in ozone above 45 km with the SIM spectra (Fig. 2b) is consistent with it being in photochemical steady state with the dominant sinks, that is, increased levels of HO_x and O. These losses are compensated by the greater production of O_x through photodissociation of O_2 in the Huggins band and this dominates the loss lower down. Furthermore, the ozone decreases produce a self-healing effect whereby more ultraviolet radiation is transmitted to lower levels, resulting in greater O_2 photolysis and thus more O_3. (See Supplementary Information.)

To assess the sensitivity of our results to uncertainty in the measured irradiance values at 200–240 nm (see Fig. 1) we carried out another set of experiments (not shown) in which the switchover from SOLSTICE to SIM was imposed at 240 nm (rather than 200 nm). There are differences in detail in the resulting temperature and ozone fields but the general picture is the same: reduced ozone in the upper stratosphere and mesosphere and a positive peak in the middle stratosphere. Therefore there is uncertainty in the magnitude of the response but this does not affect our conclusions with respect to the impact on the middle atmosphere. It also has little bearing on the radiative forcing estimates now presented.

The response of tropospheric and surface climate to variations in solar activity is an important consideration in the attribution of surface temperature trends to human or natural factors. Radiative forcing of climate is defined by the Intergovernmental Panel on Climate Change as the change in net flux at the tropopause, taking into account the effects of any stratospheric adjustment[16]. It is known that solar radiative forcing is modulated by the ozone response to changes in solar

Table 1 | Difference in global average downward radiative flux

Wavelength	200–310 nm		310–500 nm		500–700 nm		700–1,600 nm		Total solar 200–1,600 nm		Thermal		Net	
Level	TOA	TPS	TOA	TPS	TOA	TPS	TOA	TPS*	TOA	TPS	TOA	TPS	TOA	TPS
Lean data ($W\,m^{-2}$)	0.02	0.00	0.04	0.03	0.03	0.01	0.02	0.02	0.11	0.06	0	0.02	0.11	0.08
SIM data ($W\,m^{-2}$)	0.16	0.00	0.11	0.06	−0.13	−0.17	−0.05	−0.05	0.09	−0.16	0	0.06	0.09	−0.10

The flux difference between 2004 and 2007 at the top of the atmosphere (TOA) and at 105 hPa (representing the tropopause, TPS) was calculated for December in the model using the two spectral data sets.
* The radiation scheme in the 2D model does not calculate flux propagation in this spectral region. The radiation at these wavelengths is absorbed very little by stratospheric gases so the tropopause values here are assumed to be the same as the TOA values.

ultraviolet[2]. The effect of an increase in ozone is twofold: first to reduce the flux of solar radiation reaching the tropopause and second to increase the flux of infrared radiation, mainly through its impact on stratospheric temperatures. The net effect has been assessed to be a small increase in the net downward flux[17].

Simulations of the effect of solar variability on tropospheric climate, such as those reviewed in the IPCC assessments[16], tend to incorporate broad spectral bands that resolve neither details of the spectrum nor the effect of stratospheric ozone. The SIM measurements provide an additional perspective for this. Although the change in total irradiance from 2004 to 2007 is similar in the Lean and SIM data sets their very different spectral compositions, and the resulting impacts on the stratosphere, produce quite different pictures of the transmission of radiation to the tropopause, and thus different modulations of radiative forcing. Table 1 presents the change from 2004 to 2007 in global average downward solar flux at the top of the atmosphere (TOA), and at the tropopause, in four spectral bands, along with the resulting change in downward thermal radiation at the tropopause.

Little of the radiation in the 200–310 nm wavelength band reaches the tropopause, so that the large increase in the SIM irradiance at the TOA is not found here. In the 310–500 nm region, the radiation reaching the tropopause is modulated by the ozone column above, so that the larger increase in ozone produced by the SIM spectra significantly diminishes radiation reaching 105 hPa. In the 500–700 nm region, ozone absorption again plays a part so that the decrease in the TOA value in the SIM experiment becomes an even larger decrease at the tropopause. The change in spectrally integrated solar irradiance at the TOA in the two experiments is very similar: 0.11 $W\,m^{-2}$ with Lean and 0.09 $W\,m^{-2}$ with SIM. However, at the tropopause, whereas the Lean experiment shows an increase of 0.06 $W\,m^{-2}$, the SIM shows a decrease of 0.16 $W\,m^{-2}$. The thermal radiation increases the Lean radiative forcing slightly, and moderates the decrease in SIM, so that the net solar radiative forcing 2004–2007 estimated using the two data sets is $+0.08\,W\,m^{-2}$ with Lean (consistent with previous studies of radiative forcing over a solar cycle[17]) but $-0.10\,W\,m^{-2}$ when the SIM data are used. The latter suggests that radiative forcing of surface climate by the Sun is out of phase with solar activity, at least over this declining phase of solar cycle 23. In their study Cahalan *et al.*[12] did not find this out-of-phase relationship in near-surface air temperature, perhaps because their radiative–convective model did not incorporate the ozone response.

The SIM data provide an entirely different picture from the one currently accepted for the variation of solar irradiance. It is pertinent to ask whether this spectral variability is typical of solar activity cycles and, if so, why it has not been observed previously. It is possible that the Sun has been behaving in an anomalous fashion recently; certainly the current solar minimum is lower and longer than any of those observed over recent decades[18] and perhaps the solar spectrum has different characteristics when the Sun is in a state of very low activity. Gaps in understanding will only be resolved by the acquisition of long-term, well-calibrated, high-vertical-resolution measurements of stratospheric composition and temperature acquired coincidently with essential solar spectral data that have also been properly degradation-corrected and calibrated.

The SORCE observations are, however, consistent with a solar-activity-dependent change in the temperature gradient of the solar photosphere[4], suggesting that the offsetting irradiance trends with wavelength seen in SIM should appear in each solar cycle. If this is the case, then it is necessary to reconsider the current understanding[19] of the mechanisms whereby solar cycle variability influences climate: the impact on the stratosphere is much larger than previously thought and the radiative forcing of surface climate is out of phase with solar activity. At present there is no evidence to ascertain whether this behaviour has occurred before, but if this were the case during previous multi-decadal periods of low solar activity it would be necessary to revisit assessments of the solar influence on climate and to revise the methods whereby these are represented in global models.

METHODS SUMMARY

Solar spectra. The spectra used were ten-day averages of the Lean model and SORCE data centred on 21 April 2004 and 7 November 2007, chosen as the furthest spaced dates of calibrated SORCE data. The data were interpolated onto the 171 wavebands of the two-dimensional (2D) model in the range 116–730 nm.
2D model. The 2D (latitude–log pressure) zonal mean model incorporates interactions between radiative, chemical and dynamical processes. The same solar spectra are used for the calculation of chemical photodissocation and heating rates. The model was run to (seasonally varying) equilibrium with each of the four spectral data sets. All results are presented for December.
Multiple regression analysis. The code (Myles Allen, Oxford University, personal communication) estimates the coefficients of regression indices simultaneously with the parameters of a red noise model, here taken to be of order unity. The fit is iterated until the noise model fits within a pre-defined threshold. This method minimizes the possibility of noise being interpreted as a signal and can produce, using a Student's *t*-test, measures of the confidence intervals of the resultant regression coefficients, taking into account any covariance between the indices.

Received 16 February; accepted 15 August 2010.

1. Brasseur, G. & Solomon, S. *Aeronomy of the Middle Atmosphere: Chemistry and Physics of the Stratosphere* (Springer, 2005).
2. Haigh, J. D. The role of stratospheric ozone in modulating the solar radiative forcing of climate. *Nature* **370,** 544–546 (1994).
3. Harder, J., Lawrence, G., Fontenla, J., Rottman, G. & Woods, T. The Spectral Irradiance Monitor: scientific requirements, instrument design, and operation modes. *Sol. Phys.* **230,** 141–167 (2005).
4. Harder, J. W., Fontenla, J. M., Pilewskie, P., Richard, E. C. & Woods, T. N. Trends in solar spectral irradiance variability in the visible and infrared. *Geophys. Res. Lett.* **36,** L07801 (2009).
5. Lean, J. Evolution of the sun's spectral irradiance since the Maunder Minimum. *Geophys. Res. Lett.* **27,** 2425–2428 (2000).
6. Krivova, N. A., Solanki, S. K. & Floyd, L. Reconstruction of solar UV irradiance in cycle 23. *Astron. Astrophys.* **452,** 631–639 (2006).
7. Krivova, N. A., Solanki, S. K., Fligge, A. & Unruh, Y. C. Reconstruction of solar irradiance variations in cycle 23: is solar surface magnetism the cause? *Astron. Astrophys.* **399,** L1–L4 (2003).
8. Harder, J. W. *et al.* The SORCE SIM Solar Spectrum. Comparison with recent observations. *Sol. Phys.* **263,** 3–24 (2010).
9. Harwood, R. S. & Pyle, J. A. A 2-dimensional mean circulation model for the atmosphere below 80km. *Q. J. R. Meteorol. Soc.* **101,** 723–747 (1975).
10. Bekki, S. *et al.* The role of microphysical and chemical processes in prolonging the climate forcing of the Toba eruption. *Geophys. Res. Lett.* **23,** 2669–2672 (1996).
11. Warwick, N. J., Bekki, S., Nisbet, E. G. & Pyle, J. A. Impact of a hydrogen economy on the stratosphere and troposphere studied in a 2-D model. *Geophys. Res. Lett.* **31,** L05107 (2004).
12. Cahalan, R. F., Wen, G. Y., Harder, J. W. & Pilewskie, P. Temperature responses to spectral solar variability on decadal time scales. *Geophys. Res. Lett.* **37,** L07705 (2010).
13. Randel, W. J. & Wu, F. A stratospheric ozone profile data set for 1979–2005: variability, trends, and comparisons with column ozone data. *J. Geophys. Res. Atmos.* **112,** D06313 (2007).
14. Haigh, J. D. & Pyle, J. A. A two-dimensional calculation including atmospheric carbon dioxide and stratospheric ozone. *Nature* **279,** 222–224 (1979).
15. Soukharev, B. E. & Hood, L. L. Solar cycle variation of stratospheric ozone: Multiple regression analysis of long-term satellite data sets and comparisons with models. *J. Geophys. Res.* **111,** D20314 (2006).

16. Solomon, S. *et al.* (eds) *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2007).
17. Larkin, A., Haigh, J. D. & Djavidnia, S. The effect of solar UV irradiance variations on the Earth's atmosphere. *Space Sci. Rev.* **94,** 199–214 (2000).
18. Fröhlich, C. Evidence of a long-term trend in total solar irradiance. *Astron. Astrophys.* **501,** L27–L30 (2009).
19. Gray, L. J. *et al.* Solar influences on climate. *Rev. Geophys.* (in the press)..

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

Acknowledgements This work was partially funded by the UK Natural Environment Research Council SOLCLI consortium project. We are grateful to J. L. Lean for providing solar spectra, to W. J. Randel for advice on data analysis and to L. Froidevaux for advice on the EOS Aura-MLS data. The MLS data were obtained from the instrument website (http://mls.jpl.nasa.gov/index-eos-mls.php).

Author Contributions J.W.H. provided the SIM and SOLSTICE data, information on its interpretation and on solar variability, R.T. provided input on stratospheric photochemistry, A.R.W. edited the SIM data into a format suitable for the model and carried out preliminary model runs, J.D.H. performed the model experiments and diagnostics, carried out the MLS data analysis and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.D.H. (j.haigh@imperial.ac.uk).

doi:10.1038/nature09417

Melting above the anhydrous solidus controls the location of volcanic arcs

Philip C. England[1] & Richard F. Katz[1]

Segregation of magma from the mantle in subduction zones is one of the principal mechanisms for chemical differentiation of the Earth. Fundamental aspects of this system, in particular the processes by which melt forms and travels to the Earth's surface, remain obscure. Systematics in the location of volcanic arcs, the surface expression of this melting, are widely considered to be a clue to processes taking place at depth, but many mutually incompatible interpretations of this clue exist (for example, see refs 1–6). We discriminate between those interpretations by the use of a simple scaling argument derived from a realistic mathematical model of heat transfer in subduction zones. The locations of the arcs cannot be explained by the release of fluids in reactions taking place near the top of the slab. Instead, the sharpness of the volcanic fronts, together with the systematics of their locations, requires that arcs must be located above the place where the boundary defined by the anhydrous solidus makes its closest approach to the trench. We show that heat carried by magma rising from this region is sufficient to modify the thermal structure of the wedge and determine the pathway through which both wet and dry melts reach the surface.

Volcanic arcs are characterized by sharp fronts whose locations may be described, with misfits of no more than a few kilometres, by small circles on the Earth's surface (Fig. 1 and refs 7 and 8); furthermore, the depth of the top of the slab beneath these fronts falls in a narrow range (~120 ± 40 km; refs 1, 3 and 5). The sharpness of the volcanic fronts implies that a key process in the generation or transport of magma is similarly focused beneath the arcs, but there is no consensus as to what that process may be. A wide range of metamorphic and melting reactions, either in the slab or in the mantle wedge, have been proposed as candidate processes. Some authors suggest that the arcs lie above places where the degree of melting in the mantle wedge becomes high enough for the melt to segregate from the solid[2,4]. Others suggest that the locations of arcs are determined by the release of fluid near the top of the slab in reactions that are either strongly pressure-dependent[1,3,5] or strongly temperature-dependent[6,9]. None of these suggestions has, however, produced a successful quantitative prediction of the location of volcanic arcs. Here we take a different approach: starting with the observed correlation between the descent speed of the slab and its depth beneath the volcanic arcs[8], we use a simple mathematical model to fit the data and reveal the petrological processes responsible for the locations of the arcs.

Although calculations of the full temperature field in subduction zones require numerical models[10], their results can be encapsulated in simple scaling relations that show[11] that temperatures within the mantle wedge and at the top of the slab depend upon a single parameter, $Vr\delta^2/\kappa$. Here V is the convergence rate across the plate boundary, δ is the dip of the slab, r is the radial distance from the wedge corner (Fig. 1), and κ is

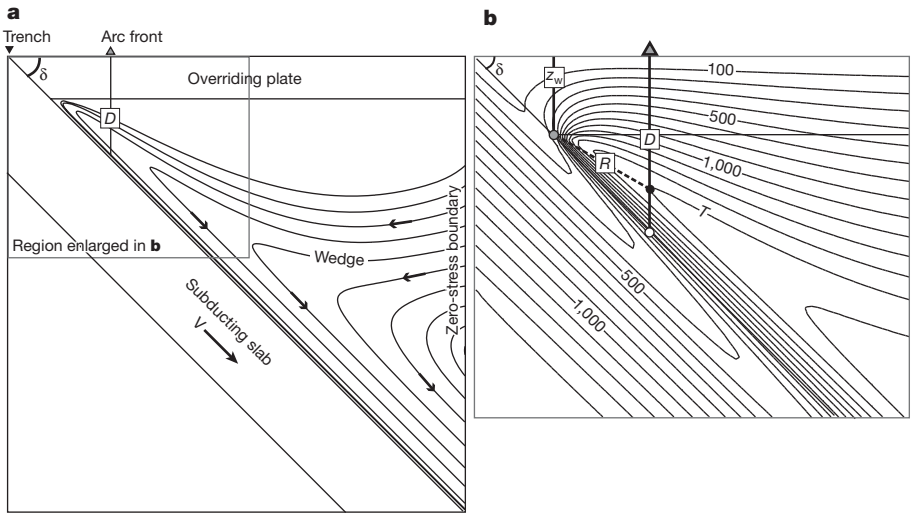


Figure 1 | Idealized cross-sections of a subduction zone, drawn perpendicular to the trench and the island arc. **a**, Two plates converge at a speed V, with the slab of oceanic lithosphere being subducted at an angle δ beneath the overriding plate. The arc front is a zone a few kilometres wide, across which volcanic activity begins as one moves away from the trench; it lies at a distance D above the top of the slab. The relative motion between slab and overriding plate generates a creeping flow in the wedge of mantle between them, which follows stream lines, shown as curved lines. The corner of the wedge is at a depth z_w. **b**, Enlargement of temperature structure of the rectangle in **a**: Isotherms are shown at intervals of 100 °C. A schematic isotherm, labelled T, has its closest approach to the wedge corner (its 'nose') immediately beneath the volcanic front, at a distance R from the corner (black circle). An open circle marks the top of the slab directly below the corner of this isotherm, at a depth D below the surface. The distance R cannot be determined by observation, but is similar to $R_D = (D - z_w)/\sin\delta$, the radial distance of the top of the slab (open circle) from the wedge corner.

[1]Department of Earth Sciences, South Parks Road, Oxford, OX1 3AN, UK.

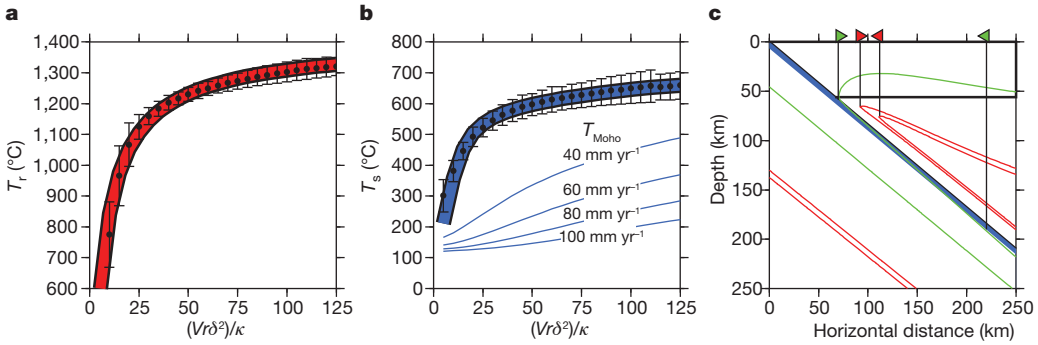


Figure 2 | Scaling relations for temperatures in the core of the mantle wedge, and at the top of the slab. **a**, Maximum temperature in the wedge, T_r, as a function of dimensionless distance from the wedge corner, $Vr\delta^2/\kappa$, where V is convergence speed, δ is dip of the slab, and κ is thermal diffusivity. Dots with error bars indicate the averages and standard deviations of T_r, determined from calculations in which V varies from 10 mm yr^{-1} to 100 mm yr^{-1}, in steps of 10 mm yr^{-1}, while δ varies from 20° to 70° in steps of 10°. The red line corresponds to the theoretical expression for T_r (equation (1)), with $T_0 = 1{,}420$ °C, $B = 3.3$ and $\beta = -0.8$. **b**, As for **a**, but for temperatures within the slab. The thick blue line corresponds to the theoretical expression for T_s, the temperature on the top of the slab (equation (2)), with $C = 1.4$ and $\gamma = -0.06$. Thin blue lines indicate the temperatures at the base of the crust (7 km below the top of the slab) for convergence rates V of 40–100 mm yr^{-1} and a slab dip of 40°. **c**, The temperature structure near the wedge corner for a calculation with a convergence speed V of 80 mm yr^{-1} and a slab dip of 40°. The location of the 500 °C isotherm is shown by green lines and the 1,225 °C and 1,275 °C isotherms are shown by red lines. Green arrowheads show the horizontal extent over which some part of the oceanic crust is at a temperature of 500 °C. Red arrowheads show the range over which the maximum temperature in the mantle wedge lies between 1,225 °C and 1,275 °C (a typical range of temperature represented by the error bars in Fig. 3b). Uncertainties in the temperatures arising from idealizations in the model are discussed in section B2 of the Supplementary Information.

the thermal diffusivity of the mantle. In this scaling, the maximum temperature in the mantle wedge T_r is given by:

$$T_r \approx T_0 \exp\left[-B\left(\frac{Vr\delta^2}{\kappa}\right)^{\beta}\right] \quad (1)$$

while the temperature T_s at the top of the slab is:

$$T_s \approx \frac{T_r}{1 + C\left(Vr\delta^2/\kappa\right)^{\gamma}} \quad (2)$$

where T_0 is a scale temperature, and B, C, β and γ are constants, the values of which depend on the details of the flow near the top of the slab. This scaling was initially derived for a model of subduction zones that treated the mantle as a constant-viscosity fluid[11]; we show here that these relations also hold for the case in which the viscosity has a dependence on temperature that is appropriate for the upper mantle (Fig. 2 and Supplementary Information).

Precise earthquake hypocentral locations[12] reveal that the depth D to the top of the slab beneath the fronts of volcanic arcs is constant, to within a few kilometres, along individual segments of arc, but varies from 80 km to 160 km between different segments[8,13] (Fig. 3). This variation rules out the hypothesis that the arcs are located above the place where the top of the slab reaches a critical pressure corresponding to a single dehydration reaction[1,3,5]. We may also rule out the hypothesis that the release of fluids by temperature-dependent reactions near the top of the slab determines the location of arcs (see, for example, ref. 6). The top of the slab lies within a thermal boundary layer a few tens of kilometres thick, across which there is a temperature difference of ~1,000 °C. Because isotherms within this boundary layer are almost parallel to the slab, any given temperature will be found over a large range of pressure (Fig. 2b, c). Therefore, temperature-dependent processes taking place near the top of the slab cannot be sharply localized, but must occur over a broad range of down-dip distances[14].

In contrast, the steep lateral thermal gradients in the core of the mantle wedge provide a setting in which localization of temperature-dependent processes is likely[2,4]. The maximum temperature in the mantle wedge depends on the dimensionless distance from the wedge corner $Vr\delta^2/\kappa$ (equation (1) and Fig. 2a). Accordingly, we should expect that if a temperature-dependent process is localized beneath the arc, the relevant isotherm will reach its closest approach to the wedge corner (which we refer to below as its 'nose') at a distance R that is inversely proportional to $V\delta^2$. Although R cannot be measured

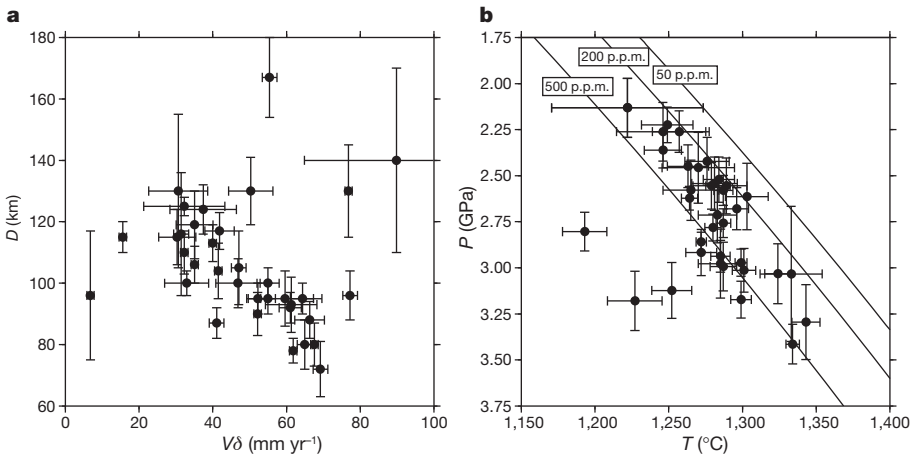


Figure 3 | Systematic variation in depth to the slab beneath volcanic arcs, and its relation to pressure–temperature conditions beneath the arcs. **a**, Depth to the top of the slab beneath volcanic arcs (see Supplementary Information), plotted against $V\delta$ (equation (3)). **b**, Conditions beneath the volcanic fronts estimated from calculations with descent speed, V, and dip, δ, corresponding to the arcs investigated in this paper. Dots show the maximum temperature in the mantle wedge beneath the front, and the pressure at which that temperature is reached. For each calculation, the arc front is taken to lie immediately above the place where the top of the slab reaches the depth D; error bars represent the range in maximum temperature (and the pressure at which it is reached) associated with a ±5 km uncertainty in the horizontal location of the arc front. Lines labelled 50 p.p.m., 200 p.p.m. and 500 p.p.m. correspond to the solidi for peridotite containing these fractions of water[15].

directly, it is very similar to the radial distance $R_D = (D - z_w)/\sin(\delta)$ of the top of the slab from the wedge corner at that location (Fig. 1), thus:

$$\frac{VR_D\delta^2}{\kappa} \approx \frac{V\delta^2(D-z_w)}{\kappa\sin(\delta)} \approx \frac{V\delta(D-z_w)}{\kappa} \qquad (3)$$

Here, z_w is the depth to the wedge corner, and we have made the small-angle approximation $\sin(\delta) \approx \delta$. Hence, for z_w constant or varying little in comparison with D, we should expect D to vary inversely with $V\delta$.

We can make a reliable determination of D for 35 arcs (Supplementary Information); the Spearman rank-order correlation coefficient between D and $V\delta$ for these arcs (Fig. 3b) is -0.34, which is significant at the 95% level of confidence. The significance of the correlation is sensitive to the presence of five arcs with large uncertainties in D (identified in the discussion of Supplementary Fig. 1); the confidence level for the correlation exhibited when these five arcs are excluded is above 99%. Thirty arc segments, with a total length of 22,000 km, exhibit the relationship between D and $V\delta$ that is predicted by equation (3), whereas five arc segments with a total length of ~4,000 km do not lie on the trend. (See Supplementary Material for further discussion.) We therefore conclude that the global systematics[8,13] strongly support the hypothesis that the locations of volcanic arcs are determined by a temperature-dependent process taking place in the wedge[2,4].

To estimate the conditions under which this process takes place, we carried out calculations of steady-state wedge thermal structure for combinations of slab dip and convergence rate corresponding to the 35 individual arcs in Fig. 3a. The maximum temperature in the column of mantle beneath the location of the arc, and the pressure at which that temperature is reached, form an array between about 1,250 °C and 1,325 °C, and 2–3.5 GPa (Fig. 3b) that corresponds closely to the range of pressure–temperature conditions for the melting of peridotite in the presence of between 200 p.p.m. H_2O and 500 p.p.m. H_2O (ref. 15). These water contents are about ten times lower than the concentrations estimated for wet melting beneath arcs, but are representative of the water contents of the mantle beneath backarc basins[16]. In what follows, we refer to such H_2O concentrations as "anhydrous" to distinguish them from the hydrous melts that are much more abundant in the wedge.

It is generally accepted that hydrous melting pervades the mantle wedge, but it has been suggested that the arc front is located somewhere above the region in which the temperature in the wedge exceeds the anhydrous solidus, because the degree of hydrous melting increases rapidly there[2,4]. Melt may also be produced in this temperature range as a result of anhydrous decompression melting of upwelling mantle; indeed, the existence of distinct regimes of melting beneath the arcs has been inferred from the eruption of low-water-content, tholeiitic melt from arc volcanoes that also erupt typical hydrous melts[17–21]. Geodynamic models show, however, that either type of melting occupies a broad region of the wedge core, extending from beneath the arc front towards the back-arc[22,23]. Any explanation for the localization of the arc fronts that invokes melts produced over this broad volume must therefore include a mechanism for focusing the melts to the line beneath the arcs.

The interaction of rising magma with the thermal boundary layer at the top of the wedge provides this mechanism. As the rising magma hits this boundary, it begins to crystallize, forming an impermeable barrier above a sloping, high-porosity channel[23,24]. Such interaction requires sufficient crystallization to seal the pore space against vertical magmatic flow[25]; this condition is met at temperatures above the anhydrous solidus, where isobaric productivity is high and small decreases in temperature result in significant crystallization, but not below it, where isobaric productivity is low[15,26]. We therefore propose, as sketched in Fig. 4a, that the high-porosity channel terminates at the 'nose' of the anhydrous solidus. This 'nose' is controlled by temperature, so its location will depend upon $V\delta$, consistent with the observations (Fig. 3).

If the advective flux of heat carried by the magma is of the same order as the vertical conductive heat flux, the melt arriving at the nose of the solidus will erode the cold boundary layer at the top of the wedge and bow upward the solidi for both wet and dry melting. The rate of heat transfer by the melt is $\sim ML$, where M is the mass flow of magma and L is its latent heat per unit mass. The conductive heat flux is $\sim k\Delta T/h$, where k is thermal conductivity, ΔT is the temperature difference between the solidus and the surface temperature, and h is the depth to the 'nose' of the solidus beneath the arc. This argument, expanded in the Supplementary Information, shows that the ratio of these two quantities controls the efficiency of thermal erosion:

$$\mathcal{M} = \frac{MLh}{k\Delta T} \qquad (4)$$

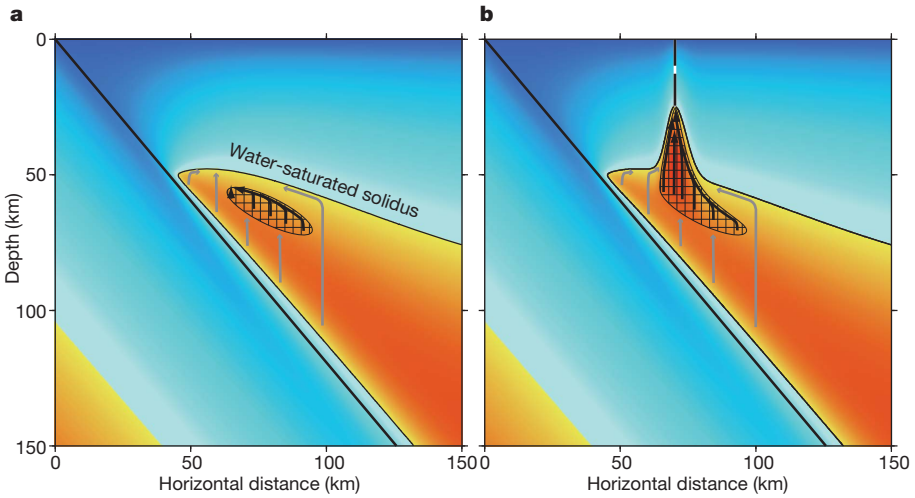


Figure 4 | A sketch of the process that determines the position of volcanoes. The top of the slab is indicated by the diagonal line starting at the origin. The black line separating the blue and yellow-to-red colours within the wedge represents the water-saturated solidus of the mantle, the cross-hatched region is above the nominally anhydrous solidus of ambient mantle. **a**, The distribution of temperature and melting without heat transport by migrating melt. Grey and black arrows show melt formed above the water-saturated and anhydrous solidi, respectively, rising, then travelling through high-porosity channels to the 'noses' of the solidi. **b**, A schematic depiction of how melt transport, indicated by arrows, would modify the distribution in **a**. Magma rising from the 'nose' of the dry solidus heats the region immediately above, deflecting upwards the dry and wet solidi so that they both reach their shallowest depth in the same horizontal location. Each type of melt travels laterally and upward along its respective solidus towards this location. Melts eventually penetrate the lithosphere by hydrofracture and dyking. In the case of low permeability in the wedge, advection of melt by the moving mantle can cause the trajectories to deviate horizontally[23]: this process could affect the details of the sketch.

For reasonable estimates of the rate of mass flux into arc volcanoes[22,27–30] $\mathcal{M} \approx 0.8$, which is sufficient to perturb upward the location of the anhydrous solidus by >40% of its depth (Supplementary Fig. 4). In consequence, all melt migrates towards the place where the solidus is bowed upward—including melt generated closer to the trench (Fig. 4b).

In summary, the observed systematics in the location of the volcanic arcs shows that the mechanisms whereby the melt reaches the Earth's surface are controlled by the dynamics of the wedge, not by any locally concentrated release of fluid in reactions taking place near the top of the slab. In our model, magma formed at temperatures above the anhydrous solidus in the wedge is focused trench-ward, to the 'nose' of the region bounded by that solidus. Above the 'nose', thermal erosion by rising magma establishes the pathway that all melts, hydrous or anhydrous, take to reach the arc volcanoes. Future analysis, including calculations that incorporate the physics of melt migration[31], will help to clarify this hypothesis.

METHODS SUMMARY

Calculations of thermal structure of subduction zones. Our calculation of the thermal structure of subduction zones employs a finite-volume discretization of the incompressible, variable-viscosity Stokes equation and steady-state, advection-diffusion energy equation on a uniformly spaced, staggered Cartesian mesh. It uses a Newton–Krylov solver provided by the Portable, Extensible Toolkit for Scientific Computation (http://www.mcs.anl.gov/petsc). The code has been benchmarked against other numerical solutions for this type of problem. The viscosity η is given by an Arrhenius law for diffusion creep of olivine:

$$\eta = A \exp\left(\frac{E + PV^*}{RT}\right) \quad (5)$$

with activation energy E and activation volume V^* of 375 kJ mol^{-1} and 5×10^{-6} m^3 mol^{-1}, respectively, and pre-exponential $A = 1.8 \times 10^7$ Pa s.

The domain is 600 km by 600 km, oriented as shown in Fig. 1, with a mesh spacing of 1 km. We have checked convergence of our solutions by comparing them, for combinations of parameters that span the range of interest, against solutions with mesh spacing of 0.5 km. We find differences between the two meshes at the level of less than 10 °C between temperatures calculated on the top of the slab or near the wedge corner.

Measurement of depth to top of slab beneath arcs. We apply the methods of ref. 8 to the 45 arc segments studied by ref. 13. For each arc segment we formed cross-sections of the seismicity and estimated D as described by ref. 8. We used the tables given in the supplementary information of ref. 13 to check whether the segment boundaries of the two studies agreed; except where noted in our Supplementary Table, the disagreements were minor and we used the segment boundaries of ref. 8.

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

Received 2 February; accepted 23 July 2010.

1. Gill, J. *Orogenic Andesites and Plate Tectonics* (Springer, 1981).
2. Kushiro, I. in *Magmatic Processes: Physicochemical Principles* (ed. Mysen, B.) Vol. 1, 165–181 (Geochemical Society Special Publication, 1987).
3. Tatsumi, Y. & Eggins, S. *Subduction Zone Magmatism* (Blackwell Science, 1995).
4. Schmidt, M. & Poli, S. Experimentally based water budgets for dehydrating slabs and consequences for arc magma generation. *Earth Planet. Sci. Lett.* **163,** 361–379 (1998).
5. Tatsumi, Y. The subduction factory: how it operates in the evolving Earth. *GSA Today* **15,** 4–10 (2005).
6. Grove, T. L., Till, C. B., Lev, E., Chatterjee, N. & Médard, E. Kinematic variables and water transport control the formation and location of arc volcanoes. *Nature* **459,** 694–697 (2009).
7. Tovish, A. & Schubert, G. Island arc curvature, velocity of convergence and angle of subduction. *Geophys. Res. Lett.* **5,** 329–332 (1978).
8. England, P. C., Engdahl, E. R. & Thatcher, W. Systematic variation in the depths of slabs beneath arc volcanoes. *Geophys. J. Int.* **156,** 377–408 (2004).
9. Ulmer, P. & Trommsdorff, V. Serpentine stability to mantle depths and subduction-related magmatism. *Science* **268,** 858–861 (1995).
10. van Keken, P. *et al.* A community benchmark for subduction zone modeling. *Phys. Earth Planet. Inter.* **171,** 187–197 (2008).
11. England, P. & Wilkins, C. A simple analytical approximation to the temperature structure in subduction zones. *Geophys. J. Int.* **159,** 1138–1154 (2004).
12. Engdahl, E., van der Hilst, R. & Buland, R. Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bull. Seismol. Soc. Am.* **88,** 722–743 (1998).
13. Syracuse, E. M. & Abers, G. A. Global compilation of variations in slab depth beneath arc volcanoes and implications. *Geochem. Geophys. Geosyst.* **7** (5), Q05017, doi: 10.1029/2005GC001045 (2006).
14. England, P. C. & Katz, R. F. Global systematics of arc volcano position. *Nature* doi: 10.1038/nature09154 (in the press).
15. Katz, R., Spiegelman, M. & Langmuir, C. A new parameterization of hydrous mantle melting. *Geochem. Geophys. Geosyst.* **4** (9), 1073, doi: 10.1029/2002GC000433 (2003).
16. Kelley, K. A. *et al.* Mantle melting as a function of water content beneath back-arc basins. *J. Geophys. Res.* **111,** B09208, doi: 10.1029/2005JB003732 (2006).
17. Baker, M., Grove, T. & Price, R. Primitive basalts and andesites from the Mt. Shasta region, N. California—products of varying melt fraction and water content. *Contrib. Mineral. Petrol.* **118,** 111–129 (1994).
18. Sisson, T. & Bronto, S. Evidence for pressure-release melting beneath magmatic arcs from basalt at Galunggung, Indonesia. *Nature* **391,** 883–886 (1998).
19. Elkins-Tanton, L., Grove, T. & Donnelly-Nolan, J. Hot, shallow mantle melting under the Cascades volcanic arc. *Geology* **29,** 631–634 (2001).
20. Cameron, B. *et al.* Flux versus decompression melting at stratovolcanoes in southeastern Guatemala. *J. Volcanol. Geotherm. Res.* **119,** 21–50 (2003).
21. Tatsumi, Y. & Suzuki, T. Tholeiitic *vs* calc-alkalic differentiation and evolution of arc crust: constraints from melting experiments on a basalt from the Izu-Bonin-Mariana arc. *J. Petrol.* **50,** 1575–1603 (2009).
22. Conder, J., Wiens, D. & Morris, J. On the decompression melting structure at volcanic arcs and back-arc spreading centers. *Geophys. Res. Lett.* **29,** 1727, doi: 10.1029/2002GL015390 (2002).
23. Cagnioncle, A.-M., Parmentier, E. M. & Elkins-Tanton, L. T. Effect of solid flow above a subducting slab on water distribution and melting at convergent plate boundaries. *J. Geophys. Res.* **112,** B09402, doi: 10.1029/2007JB004934 (2007).
24. Sparks, D. & Parmentier, E. Melt extraction from the mantle beneath spreading centers. *Earth Planet. Sci. Lett.* **105,** 368–377 (1991).
25. Spiegelman, M. Physics of melt extraction: theory, implications, and applications. *Phil. Trans. R. Soc. Lond. A* **342,** 23–41 (1993).
26. Hirschmann, M. M., Asimow, P. D., Ghiorso, M. S. & Stolper, E. M. Calculation of peridotite partial melting from thermodynamic models of minerals and melts. III. Controls on isobaric melt production and the effect of water on melt production. *J. Petrol.* **40,** 831–851 (1999).
27. Crisp, J. Rates of magma emplacement and volcanic output. *J. Volcanol. Geotherm. Res.* **20,** 177–211 (1984).
28. Reymer, A. & Schubert, G. Phanerozoic addition rates to the continental crust and crustal growth. *Tectonics* **3,** 63–77 (1984).
29. Dimalanta, C., Taira, A., Yumul, G., Tokuyama, H. & Mochizuki, K. New rates of western Pacific island arc magmatism from seismic and gravity data. *Earth Planet. Sci. Lett.* **202,** 105–115 (2002).
30. White, S., Crisp, J. & Spera, F. Long-term volumetric eruption rates and magma budgets. *Geochem. Geophys. Geosyst.* **7,** Q03010 (2006).
31. Katz, R. Magma dynamics with the enthalpy method: benchmark solutions and magmatic focusing at mid-ocean ridges. *J. Petrol.* **49,** doi: 10.1093/petrology/egn058 (2008).

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

Acknowledgements We are grateful to T. Grove and E. Lev for their free exchange of ideas and data concerning their paper (ref. 6), and to C. Langmuir for comments that helped us to improve the manuscript. We thank E. Syracuse for providing hypocentral locations. Numerical models were run on computational clusters at the Oxford Supercomputing Centre.

Author Contributions R.F.K. wrote the code for the numerical experiments; P.C.E. carried out the re-analysis of depth-to-slab (Fig. 3 and Supplementary Information). Both authors participated equally in developing the ideas presented in this paper and in writing it.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.C.E. (philip@earth.ox.ac.uk).

METHODS

Calculations of thermal structure of subduction zones. Our calculation of the thermal structure of subduction zones employs a finite-volume discretization of the incompressible, variable-viscosity Stokes equation and steady-state, advection-diffusion energy equation on a uniformly spaced, staggered Cartesian mesh[32]. It uses a Newton–Krylov solver provided by the Portable, Extensible Toolkit for Scientific Computation[33,34]. The code has been benchmarked against other numerical solutions for this type of problem[10]. The viscosity η is given by an Arrhenius law for diffusion creep of olivine:

$$\eta = A \exp\left(\frac{E + PV^*}{RT}\right) \tag{6}$$

with activation energy E and activation volume V^* of 375 kJ mol^{-1} and 5×10^{-6} m^3 mol^{-1}, respectively, and pre-exponential $A = 1.8 \times 10^7$ Pa s (refs 35 and 36).

The domain is composed of three subdomains, the slab, the wedge, and the overlying lid. Velocity in the slab subdomain is prescribed using the convergence rate and the slab dip; velocity in the lid is forced to equal zero. Within the wedge we solve the incompressible Stokes equation with no buoyancy term. There are no-slip conditions on the wedge–lid and wedge–slab boundaries, and there is a no-stress condition applied at the wedge inflow–outflow boundary. Temperature is fixed at zero on the top of the domain; the slab is given a temperature profile using the one-dimensional conducting slab solution; this solution is also applied to fix the temperatures on the inflow boundary. Complete details are provided in ref. 10.

The domain is 600 km by 600 km, oriented as shown in Fig. 1, with a mesh spacing of 1 km. We have checked convergence of our solutions by comparing them, for combinations of parameters that span the range of interest, against solutions with mesh spacing of 0.5 km. We find differences between the two meshes at the level of less than 10 °C between temperatures calculated on the top of the slab or near the wedge corner.

An important parameter in thermal models of subduction zones is the depth at which the slab couples to the mantle wedge with a no-slip condition[37]. The maximum extent of shallowly dipping thrust faulting on the plate interface in subduction zones is about 45 km (ref. 38) and, allowing for a transitional zone of aseismic sliding below this depth, we fix the depth to the top of full coupling between wedge and slab at 56 km.

Measurement of depth to top of slab beneath arcs. We apply the methods of ref. 8 to the 45 arc segments studied by ref. 13. For each arc segment we formed cross-sections of the seismicity and estimated D as described by ref. 8. We used the tables given in the supplementary information of ref. 13 to check whether the segment boundaries of the two studies agreed; except where noted in our Supplementary Table, the disagreements were minor and we used the segment boundaries of ref. 8.

We could not obtain reliable estimates of D for seven arcs, either because of the sparsity of earthquakes or because the volcanoes are spread out so widely that a clear volcanic front cannot be identified (see remarks on individual arcs, in the Supplementary Information). We combined the remaining volcanoes into 35 arc segments for which D estimated by the two different methods can be reconciled. An example of illustrating the two approaches, and their reconciliation in one of the more problematic arcs, can be seen in Supplementary Fig. 2. The Marianas, Lesser Antilles, and Vanuatu arcs are treated as single arcs here; each is split into northern and southern segments by ref. 41.

32. Knepley, M., Katz, R. & Smith, B. in *Numerical Solution of Partial Differential Equations on Parallel Computers* Vol. 51 *Lecture Notes in Computational Science and Engineering* (eds Bruaset, A. & Tveito, A.) 413–438 (Springer, 2006).
33. Balay, S. *et al. Portable, Extensible Toolkit for Scientific Computation (PETSc)*; ⟨http://www.mcs.anl.gov/petsc⟩ (2001).
34. Katz, R., Knepley, M., Smith, B., Spiegelman, M. & Coon, E. Numerical simulation of geodynamic processes with the Portable Extensible Toolkit for Scientific Computation. *Phys. Earth Planet. Inter.* **163,** 52–68 (2007).
35. Karato, S. & Wu, P. Rheology of the upper mantle—a synthesis. *Science* **260,** 771–778 (1993).
36. Hirth, G. & Kohlstedt, D. in *Inside the Subduction Factory* (ed. Eiler, J.) Vol. 138, 83–105 (AGU Geophysical Monograph, American Geophysical Union, 2003).
37. Conder, J. A case for hot slab surface temperatures in numerical viscous flow models of subduction zones with an improved fault zone parameterization. *Phys. Earth Planet. Inter.* **149,** 155–164 (2005).
38. Tichelaar, B. W. & Ruff, L. Depth of seismic coupling along subduction zones. *J. Geophys. Res.* **98,** 2107–2037 (1993).

doi:10.1038/nature09407

Global metabolic impacts of recent climate warming

Michael E. Dillon[1], George Wang[2]† & Raymond B. Huey[2]

Documented shifts in geographical ranges[1,2], seasonal phenology[3,4], community interactions[5], genetics[3,6] and extinctions[7] have been attributed to recent global warming[8–10]. Many such biotic shifts have been detected at mid- to high latitudes in the Northern Hemisphere[4,9,10]—a latitudinal pattern that is expected[4,8,10,11] because warming is fastest in these regions[8]. In contrast, shifts in tropical regions are expected to be less marked[4,8,10,11] because warming is less pronounced there[8]. However, biotic impacts of warming are mediated through physiology, and metabolic rate, which is a fundamental measure of physiological activity and ecological impact, increases exponentially rather than linearly with temperature in ectotherms[12]. Therefore, tropical ectotherms (with warm baseline temperatures) should experience larger absolute shifts in metabolic rate than the magnitude of tropical temperature change itself would suggest, but the impact of climate warming on metabolic rate has never been quantified on a global scale. Here we show that estimated changes in terrestrial metabolic rates in the tropics are large, are equivalent in magnitude to those in the north temperate-zone regions, and are in fact far greater than those in the Arctic, even though tropical temperature change has been relatively small. Because of temperature's nonlinear effects on metabolism, tropical organisms, which constitute much of Earth's biodiversity, should be profoundly affected by recent and projected climate warming[2,13,14].

Global warming is probably having profound and diverse effects on organisms[1–11]. Organisms living at mid- to high latitudes in the Northern Hemisphere are predicted to be the most affected by climate warming[4,8,10,11], because temperatures have risen most rapidly there[8]. Indeed, the vast majority of biotic impacts of warming have been documented in this region, but few studies have yet searched for impacts in other areas, especially the tropics[2,4,8,10,13,14]. One way to circumvent this geographical sampling bias is to use temperature data with broad geographical coverage to predict global patterns of physiological responses to observed temperature change[13]. Metabolic rate is a heuristic metric here because it is a fundamental physiological index of an organism's energetic and material needs, its processing capacity and its ecological impact[12].

Metabolic rates of ectotherms depend principally on body mass (m) and body temperature (T), as described by a fundamental equation[12]:

$$B(m,T) = b_0 m^{3/4} e^{-E/kT} \quad (1)$$

where B is metabolic rate, b_0 is an empirically derived and taxon-specific normalization constant, m is body mass, E is the average activation energy for biochemical reactions of metabolism, T is body temperature (in Kelvin), and k is the Boltzmann constant. When standardized for mass, this equation enables metabolic comparisons among different sized organisms[12]. These mass-normalized metabolic rates are proportional to the 'Boltzmann factor' ($e^{-E/kT}$; the familiar 'Q_{10}' effect in physiology is an approximation of the Boltzmann factor)[15].

Although the thermodynamic and statistical validity of equation (1) is debated[16–18], it provides a useful approximation of metabolic rates[12,15] for exploratory macrophysiological investigations[16,19]. In the context of climate warming, it predicts that metabolism will shift more in response to a unit change in temperature at high temperature than at low temperature[15], at least over biologically common and non-stressful temperatures (~0 °C to ~40 °C; see Methods)[12].

To estimate geographical patterns of warming-induced changes in metabolic rates of terrestrial ectotherms, we compiled high-frequency temperature data for the period of 1961 to 2009 for 3,186 weather stations across the world (~500 million temperature measurements; Methods and Supplementary Fig. 1)[20]. We derived average values of E (0.69) and of b_0 (23.66) from empirical estimates for diverse ectotherms (Supplementary Table 1)[12]. We substituted these 'average ectotherm' values into equation (1) to estimate mass-normalized metabolic rates ($Bm^{-3/4}$) from global temperature data. Because metabolic rate varies nonlinearly with temperature, calculating mean metabolic rates from mean temperatures is inappropriate (the 'fallacy of the averages'; see Methods and Supplementary Fig. 2)[15]. Therefore we estimated metabolic rate for each temperature measurement and subsequently determined average temperature and average metabolic rate for each station during the Intergovernmental Panel on Climate Change (IPCC) standard reference period (1961–1990) and for all five-year intervals from 1980 to 2009. To account for non-uniform distribution of stations and to enable comparisons among latitudinal regions, we determined averages for all stations within 5° latitude by 5° longitude grid cells and then area-corrected grid-cell means and standard errors of temperature measurements and metabolic rate estimates for each region (Methods).

Temperature changes since 1980 in this data set are consistent with recent findings[8]: temperatures rose fastest in the Arctic, somewhat less quickly in the north temperate zones, and more slowly in the tropics, but remained essentially unchanged in the south temperate zone (Fig. 1a).

Predicted absolute changes in metabolic rates show a markedly different pattern: metabolic rates increased most quickly in the tropics and north temperate zones, and less so in the Arctic (Fig. 1b). In fact, the latitudinal ordering of changes in temperature since 1980 fails to predict the latitudinal ordering of changes in metabolic rate ($P = 0.68$), even when a powerful ordered-heterogeneity test is used[21]. The predicted increase in metabolism in the tropics was large, despite the small rise in temperature there (Fig. 1a), because tropical warming took place in an environment that was initially warm.

Absolute changes in metabolic rates determine an organism's total energy use and thus the impacts of climate change on ecosystem-level processes, but per cent changes in metabolic rates are nonetheless relevant to the impacts of climate change on individual organisms[22]. Such relative changes in metabolic rates (expressed as per cent of the standard reference period on a per-station basis) closely match temperature changes (Fig. 1c), indicating that impacts on individual ectotherms have probably been relatively large in the Arctic and north temperate zones.

To evaluate whether the patterns described earlier (Fig. 1b) are robust to our use of average values of E and b_0, we re-ran analyses using estimates of E and b_0 specific to diverse ectotherm taxa (Supplementary Table 1)[12]. Large effects of recent climate warming on metabolic rates are predicted for invertebrates, amphibians and reptiles in equatorial

[1]Department of Zoology and Physiology, University of Wyoming, Laramie, Wyoming 82071, USA. [2]Department of Biology, Box 351800, University of Washington, Seattle, Washington 98195, USA. †Present address: Max Planck Institute for Developmental Biology, Tübingen 72076, Germany.

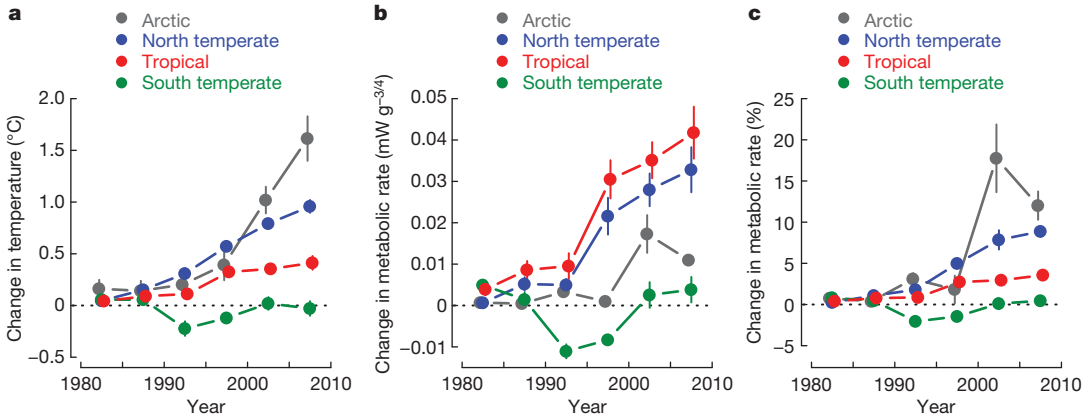


Figure 1 | Global changes in temperature and in metabolic rates since 1980. **a**, Changes in mean temperature (5-year averages) for Arctic ($n = 100$ grid cells), north temperate ($n = 356$), south temperate ($n = 51$) and tropical ($n = 169$) regions. **b**, Predicted absolute changes in mass-normalized metabolic rates by geographical region. **c**, Predicted relative changes in mass-normalized metabolic rates. Both temperature and metabolic rate are expressed as differences from the standard reference period (1961–1990), calculated on a per-station basis, on the basis of E and b_0 for an average ectotherm (Supplementary Table 1). Data points are means ± s.e.m. of area-corrected, gridded weather-station data (Methods).

West Africa, the Caribbean and Central America, Ecuador, eastern equatorial Brazil, and the Persian Gulf region (Fig. 2c–e). However, we emphasize that weather station coverage in some of these regions is sparse and each taxonomic group is not found in all geographical regions. Overall, general patterns in Fig. 1b are robust for different taxa (Fig. 2b–e): the largest predicted absolute shifts in metabolic rate for all taxa are in the tropics. Nevertheless, small differences in the relationship between metabolism and temperature (that is, E and b_0) can alter the magnitude of the effects of climate warming on organism physiology (Fig. 2).

The patterns under discussion are for mass-normalized metabolic rates, but the magnitude of metabolic shift will necessarily differ for small versus large ectotherms. Of course, absolute shifts will be greater for larger ectotherms, but equation (1) indicates that mass-specific metabolic rates of small ectotherms will show larger increases (Supplementary Fig. 3).

Several assumptions underlie the patterns shown in Fig. 1b, c. The exponent (3/4) for metabolic rate as a function of mass is debated[16–18], but reasonable shifts of this exponent for given taxa will only alter the heights of all latitudinal lines, not their relative ordering. We assume that surface air temperatures approximate ectotherm body temperatures; therefore, our metabolic estimates apply to thermoconforming and exposed ectotherms. This is reasonable for small ectotherms living in shaded environments[23], but less so for large ectotherms that live in thermally heterogeneous environments, where behavioural thermoregulation is possible, or for organisms that spend extensive periods in retreats[24]. Also, we assume that the coefficients (E, b_0) of equation (1) are independent of latitude. However, some high-latitude ectotherms have relatively elevated metabolic rates; and this is thought to represent an evolutionary metabolic compensation for the physiologically depressing effects of low body temperature[16,25,26]. With reference to equation (1), metabolic compensation would be indicated[26] by latitudinal increases in b_0 and/or in E. In fact, the patterns of metabolic responses shown in Fig. 1b hold even when we shift these parameters over a large

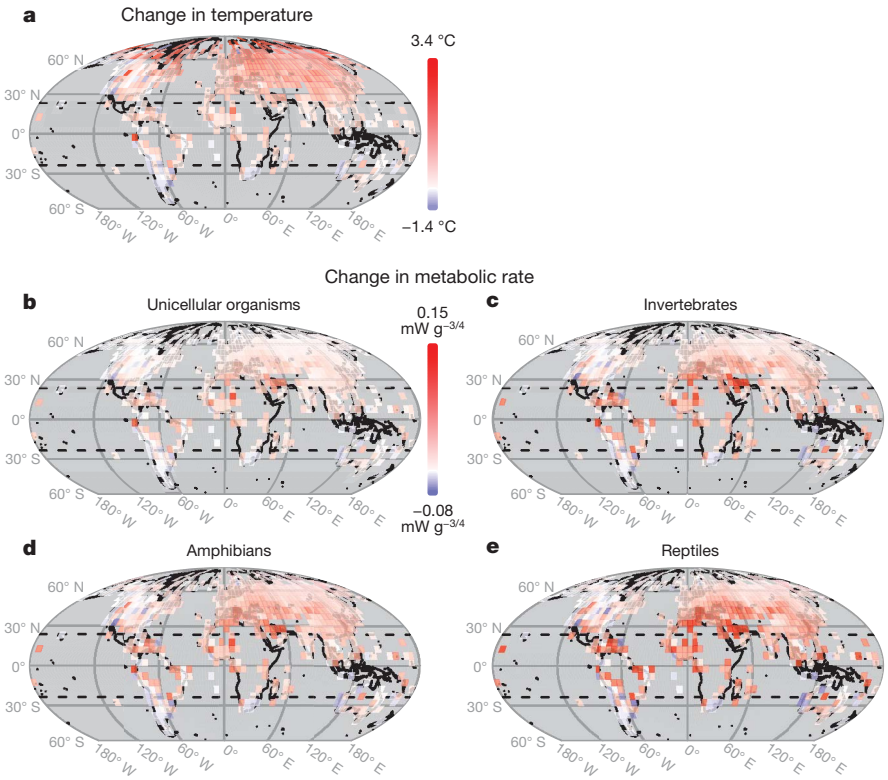


Figure 2 | Predicted changes in metabolic rates of diverse terrestrial ectotherms. **a**, Difference in temperature between 1961–1990 and 2005–2009, with scale bar shown on right. **b–e**, Difference in mass-normalized metabolic rates (predicted) for the same period for four terrestrial ectothermic animal taxa for which empirical estimates of E and b_0 are available (Supplementary Table 1)[12]. Colour bar to right of **b** indicates scale for **b–e**. Grey shading indicates grid cells with no temperature data.

range of biologically reasonable values[26] to simulate extreme metabolic compensation at high latitude (Methods and Supplementary Fig. 4).

Our analyses indicate that warming during the past three decades has had its biggest absolute impacts on metabolic rates in tropical and north temperate zones (Fig. 1b). The outlook for future warming is less clear. Without predictions of future daily and seasonal temperature cycles (not merely of mean annual temperatures), we cannot directly estimate future metabolic changes without violating the fallacy of averages[15]. Nevertheless, our analyses of recent temperature data indicate that even when the temperature shifts in the north temperate region are more than double those in the tropics (Fig. 1a), absolute shifts in metabolic rates are similar in the two regions (Fig. 1b). If this pattern holds, projected increases in median surface air temperature by the end of the twenty-first century for the two regions (3.5–4.0 °C in the tropics, and 4.0–5.5 °C in the north temperate zone)[8] should cause roughly similar absolute increases in metabolic rates of tropical and north temperate organisms.

Recent studies using diverse physiological and biophysical approaches indicate that tropical ectotherms may be particularly vulnerable to climate warming[2,7,13,14,24,27], even though observed and predicted tropical warming is relatively small[8]. Our estimates suggest that tropical ectotherms are also experiencing large increases in metabolic rate (Fig. 1b). Such increases will have physiological and ecological impacts: warmed tropical ectotherms will have an increased need for food and increased vulnerability to starvation unless food resources increase, possible reduced discretionary energy for reproduction[22], increased rates of evaporative water loss in dry environments and altered demographies[13]. Larger increases in metabolic rates of tropical soil biota may explain larger absolute changes in tropical soil respiration[28]. Furthermore, metabolic increases should alter food web dynamics, leading to elevated rates of herbivory and predation, as well as changes in the spread of insect-borne tropical diseases[29]. Because the tropics are the centre of Earth's biodiversity and its chief engine of primary productivity, the relatively large effects of temperature change on the metabolism of tropical ectotherms may have profound local and global consequences.

METHODS SUMMARY

We obtained hourly temperature records from 22,486 weather stations spread across the world[20], but then included only stations that sampled throughout the IPCC standard reference period (1961–1990)[8] as well as 1991–2009, in all seasons and on average at least every six hours. We also excluded five Antarctic stations, such that 3,186 stations remained. Geographical coverage is uneven (Supplementary Fig. 1), but all regions are well represented in this restricted data set (Supplementary Table 2). Furthermore, including data from 5,561 stations with data from 1961 to 2009 (but with no other limitations; Supplementary Table 2), does not alter our conclusions (Supplementary Fig. 5). To correct for the uneven spatial distribution of stations (Supplementary Figs 1 and 5A), we computed mean temperatures and metabolic rates for all stations within 5° latitude by 5° longitude cells. We calculated means and standard errors for latitudinal regions by weighting grid-cell means by the interpolated rectangular mid-cell areas[30]. To estimate whether metabolic compensation at high latitude might alter patterns, we recalculated metabolic rates after substituting extreme values of E and of b_0. Specifically, we used very high values of E (0.76) and b_0 (26.85) for the north temperate and Arctic areas, but very low values of E (0.50) and b_0 (15.68) for tropical areas (see Supplementary Table 1). Such extreme metabolic compensation (these values span most of the known range of E)[26] does not alter our conclusions (Supplementary Fig. 4).

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

Received 5 June; accepted 10 August 2010.

1. Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421,** 37–42 (2003).
2. Colwell, R. K., Brehm, G., Cardelus, C. L., Gilman, A. C. & Longino, J. T. Global warming, elevational range shifts, and lowland biotic attrition in the wet tropics. *Science* **322,** 258–261 (2008).
3. Bradshaw, W. & Holzapfel, C. Genetic shift in photoperiodic response correlated with global warming. *Proc. Natl Acad. Sci. USA* **98,** 14509–14511 (2001).
4. Parmesan, C. Influences of species, latitudes and methodologies on estimates of phenological response to global warming. *Glob. Change Biol.* **13,** 1860–1872 (2007).
5. Both, C., van Asch, M., Bijlsma, R., van den Burg, A. & Visser, M. Climate change and unequal phenological changes across four trophic levels: constraints or adaptations? *J. Anim. Ecol.* **78,** 73–83 (2009).
6. Umina, P. A., Weeks, A. R., Kearney, M. R., McKechnie, S. W. & Hoffmann, A. A. A rapid shift in a classic clinal pattern in *Drosophila* reflecting climate change. *Science* **308,** 691–693 (2005).
7. Sinervo, B. *et al.* Erosion of lizard diversity by climate change and altered thermal niches. *Science* **328,** 894–899 (2010).
8. IPCC. *Climate Change 2007: Impacts, Adaptation, and Vulnerability* (Cambridge Univ. Press, 2007).
9. Walther, G.-R. *et al.* Ecological responses to recent climate change. *Nature* **416,** 389–395 (2002).
10. Rosenzweig, C. *et al.* Attributing physical and biological impacts to anthropogenic climate change. *Nature* **453,** 353–357 (2008).
11. Root, T. L. *et al.* Fingerprints of global warming on wild animals and plants. *Nature* **421,** 57–60 (2003).
12. Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M. & Charnov, E. L. Effects of size and temperature on metabolic rate. *Science* **293,** 2248–2251 (2001).
13. Deutsch, C. A. *et al.* Impacts of climate warming on terrestrial ectotherms across latitude. *Proc. Natl Acad. Sci. USA* **105,** 6668–6672 (2008).
14. Pounds, J., Fogden, M. & Campbell, J. Biological response to climate change on a tropical mountain. *Nature* **398,** 611–615 (1999).
15. Savage, V. M. Improved approximations to scaling relationships for species, populations, and ecosystems across latitudinal and elevational gradients. *J. Theor. Biol.* **227,** 525–534 (2004).
16. Clarke, A. Temperature and the metabolic theory of ecology. *Funct. Ecol.* **20,** 405–412 (2006).
17. Downs, C. J., Hayes, J. P. & Tracy, C. R. Scaling metabolic rate with body mass and inverse body temperature: a test of the Arrhenius fractal supply model. *Funct. Ecol.* **22,** 239–244 (2008).
18. O'Connor, M. P. *et al.* Reconsidering the mechanistic basis of the metabolic theory of ecology. *Oikos* **116,** 1058–1072 (2007).
19. Martínez del Rio, C. Metabolic theory or metabolic models? *Trends Ecol. Evol.* **23,** 256–260 (2008).
20. Lott, N., Baldwin, R. & Jones, P. *The FCC Integrated Surface Hourly Database, A New Resource of Global Climate Data.* ⟨http://www1.ncdc.noaa.gov/pub/data/techrpts/tr200101/tr2001-01.pdf⟩ (National Climatic Data Center, 2001).
21. Rice, W. R. & Gaines, S. D. Extending non-directional heterogeneity tests to evaluate simply ordered alternative hypotheses. *Proc. Natl Acad. Sci. USA* **91,** 225–226 (1994).
22. Dunham, A. E. in *Biotic Interactions and Global Change* (eds Kareiva, P. M., Kingsolver, J. G. & Huey, R. B.) 95–119 (Sinauer, 1993).
23. Hertz, P. E., Huey, R. B. & Stevenson, R. D. Evaluating temperature regulation by field-active ectotherms: the fallacy of the inappropriate question. *Am. Nat.* **142,** 796–818 (1993).
24. Kearney, M., Shine, R. & Porter, W. P. The potential for behavioral thermoregulation to buffer "cold-blooded" animals against climate warming. *Proc. Natl Acad. Sci. USA* **106,** 3835–3840 (2009).
25. Pörtner, H. Physiological basis of temperature-dependent biogeography: tradeoffs in muscle design and performance in polar ectotherms. *J. Exp. Biol.* **205,** 2217–2230 (2002).
26. Irlich, U. M., Terblanche, J. S., Blackburn, T. M. & Chown, S. L. Insect rate–temperature relationships: environmental variation and the metabolic theory of ecology. *Am. Nat.* **174,** 819–835 (2009).
27. Wake, D. B. & Vredenburg, V. T. Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proc. Natl Acad. Sci. USA* **105,** 11466–11473 (2008).
28. Bond-Lamberty, B. & Thomson, A. Temperature-associated increases in the global soil respiration record. *Nature* **464,** 579–582 (2010).
29. Paaijmans, K. P., Read, A. F. & Thomas, M. B. Understanding the link between malaria risk and climate. *Proc. Natl Acad. Sci. USA* **106,** 13844–13849 (2009).
30. Hastings, D. A. & Dunbar, P. K. *Global Land One-Kilometer Base Elevation (GLOBE).* ⟨http://www.ngdc.noaa.gov/mgg/topo/report/globedocumentationmanual.pdf⟩ (National Geophysical Data Center, 1999).

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

Acknowledgements We thank T. L. Daniel, C. Martínez del Rio, W. R. Rice and J. Tewksbury for discussion, and S. L. Chown for sharing his data on latitudinal variation in *E*. Research was funded in part by NSF IOB-041684 to R.B.H. and by an NSF Minority Postdoctoral Fellowship to M.E.D.

Author Contributions M.E.D., G.W. and R.B.H. conceived the project, designed the analyses and wrote the paper; M.E.D. and G.W. collated weather station data and did temperature and metabolic rate calculations.

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

METHODS

Weather station data. We downloaded all available 'isd-lite' weather-station data (http://www.ncdc.noaa.gov/oa/climate/isd/index.php)[20]. From the initial 22,486 weather stations, we extracted a 'restricted' data set (3,186 stations) that sampled throughout the entire IPCC standard reference period (1961–1990)[8] as well as up to 31 December 2009, in all seasons, and at least every six hours. (We did not include Antarctic stations.) This data set had a total of 493,256,415 temperature measurements with an average of 8.8 temperature measurements per station per day; it was used for Figs 1 and 2 and Supplementary Figs 1–4. For the 'unrestricted' data set, we included all stations (other than Antarctic) that had data for the above time period, independent of the seasonality or frequency of sampling. This data set had 5,561 stations (used for Supplementary Fig. 5).

Metabolic rate estimates. To estimate metabolic rates from temperature data, we used empirically derived estimates of the coefficients (E and b_0) of the equation relating metabolic rate to temperature and body mass for unicellular organisms, multicellular invertebrates, amphibians and reptiles[12]. We excluded fish because our data are air (not water) temperatures. We excluded birds and mammals because their body temperatures will not match air temperatures. We excluded plants because the temperature dependence of their metabolic rates differs fundamentally from that of animal ectotherms[31].

Mean temperatures are expedient for analyses of the impacts of climate warming. However, because the relationship between temperature and metabolic rates is inherently nonlinear, the use of mean rather than individual temperatures to predict metabolic rates will induce spurious results[32]—an effect known as the 'fallacy of the averages'[15]. To illustrate this fallacy, we recomputed metabolic rates for geographical regions using mean annual temperatures (Supplementary Fig. 2) for comparison with rates predicted from 'instantaneous' temperatures (Fig. 1b). Note that the use of mean temperatures underestimates the predicted increases in metabolic rates[15], and also de-emphasizes the impact of warming in the north temperate zones relative to the tropics. Consequently, it is imperative to use high-frequency temperature data and to compute metabolic rate separately for each temperature measurement.

Our analysis includes temperatures that fall outside the normal tolerance range of most organisms (that is, below ~0 °C and above ~40 °C). We include these values for analytical transparency, and because their inclusion is conservative for our analyses. Eliminating negative temperatures (for example, substituting metabolic rates at 0 °C for all temperatures below 0 °C) will have little effect because metabolic rates are negligible at these extremely cold temperatures. Substituting metabolic rates at 40 °C for temperatures above 40 °C will tend to reduce metabolic rates at mid-latitudes where these hot temperatures occur in summer; this would induce a downward bias in our predicted metabolic rates for the north temperate zone. In other words, by not truncating metabolic rates to the normal tolerance range (0–40 °C), we avoid a bias that would favour increased metabolic rates in the tropics.

Geographical coverage. Weather stations are not equally spaced across the world (summarized in Supplementary Table 2, Supplementary Figs 1 and 5a). To adjust for the uneven spatial distribution (and non-independence) of stations, we computed mean temperatures and metabolic rates for all stations within 5° latitude by 5° longitude grid cells (see Fig. 2). We then calculated means and standard errors for latitudinal regions (see Fig. 1) by weighting grid-cell means by interpolated rectangular mid-cell areas[30]. We further tested the effects of weather station spatial coverage on our conclusions by comparing analyses using restricted (3,186 stations) and unrestricted (5,561 stations) data sets. Our conclusions are robust and independent of the data set used (Supplementary Fig. 5).

Assumptions. We assume that station temperatures match the temperature of a dry-skinned ectotherm positioned in shade at 2-m height. Of course, mobile ectotherms can often use behaviour (for example, microhabitat selection) to buffer body temperatures against changes in air temperatures[23,24]. Thus our metabolic estimates should be viewed as an estimate for a non-regulating, inert and exposed ectotherm.

We assume that a single metabolic curve (equation (1)) applies to all ectotherms, independent of latitude. However, some high-latitude ectotherms have relatively raised metabolic rates, which may reflect metabolic compensation for temperature[26]. To estimate whether metabolic compensation at high latitude might alter latitudinal patterns (Fig. 1), we recalculated metabolic rates after substituting extreme values of E and b_0. Specifically, we used very high values of E (0.76) and b_0 (26.85) for the north temperate and Arctic areas, but very low E (0.50) and b_0 (15.68) for tropical areas (see Supplementary Table 1). Such extreme metabolic compensation (these values span most of the known range of E)[26] does not alter our conclusions (Supplementary Fig. 4).

Statistics. We used an ordered-heterogeneity test[21] to evaluate whether latitudinal ordering of changes in temperature since 1980 predicts the latitudinal ordering of changes in metabolic rate.

31. Reich, P. B., Tjoelker, M. G., Machado, J.-L. & Oleksyn, J. Universal scaling of respiratory metabolism, size and nitrogen in plants. *Nature* **439,** 457–461 (2006).
32. Ruel, J. J. & Ayres, M. P. Jensen's inequality predicts effects of environmental variation. *Trends Ecol. Evol.* **14,** 361–366 (1999).

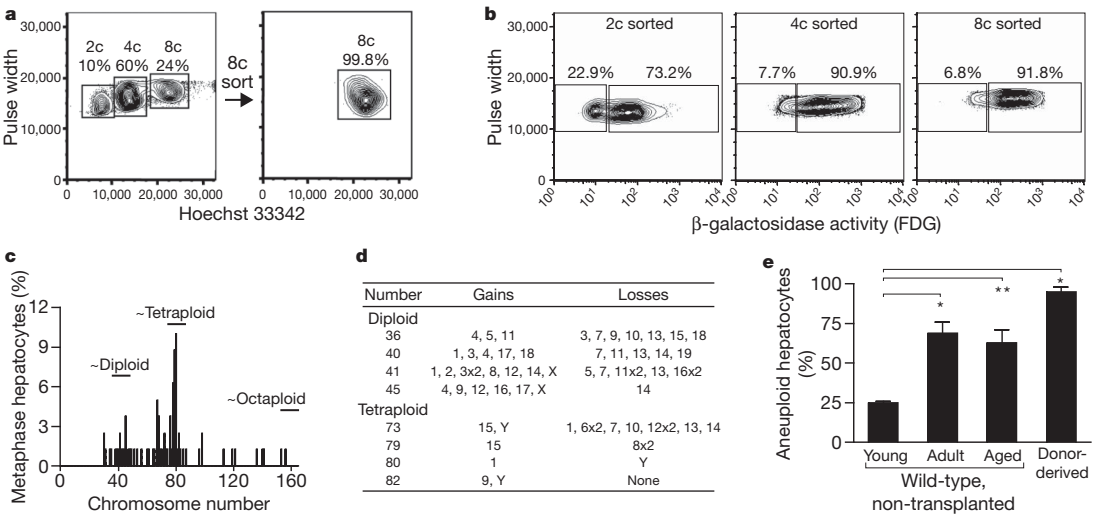
doi:10.1038/nature09414

The ploidy conveyor of mature hepatocytes as a source of genetic variation

Andrew W. Duncan[1], Matthew H. Taylor[2], Raymond D. Hickey[1,3], Amy E. Hanlon Newell[3], Michelle L. Lenzi[3], Susan B. Olson[3], Milton J. Finegold[4] & Markus Grompe[1]

Mononucleated and binucleated polyploid hepatocytes (4n, 8n, 16n and higher) are found in all mammalian species, but the functional significance of this conserved phenomenon remains unknown[1–4]. Polyploidization occurs through failed cytokinesis, begins at weaning in rodents and increases with age[2,5–7]. Previously, we demonstrated that the opposite event, ploidy reversal, also occurs in polyploid hepatocytes generated by artificial cell fusion[8–10]. This raised the possibility that somatic 'reductive mitoses' can also happen in normal hepatocytes. Here we show that multipolar mitotic spindles form frequently in mouse polyploid hepatocytes and can result in one-step ploidy reversal to generate offspring with halved chromosome content. Proliferating hepatocytes produce a highly diverse population of daughter cells with multiple numerical chromosome imbalances as well as uniparental origins. Our findings support a dynamic model of hepatocyte polyploidization, ploidy reversal and aneuploidy, a phenomenon that we term the 'ploidy conveyor'. We propose that this mechanism evolved to generate genetic diversity and permits adaptation of hepatocytes to xenobiotic or nutritional injury.

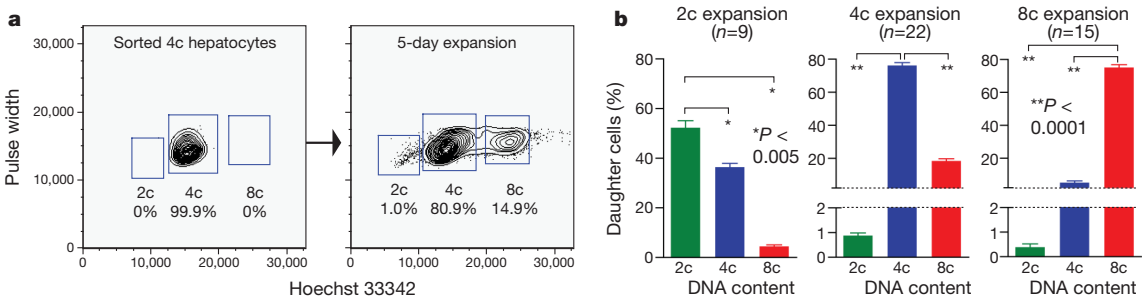
We first tested whether normal polyploid hepatocytes can undergo ploidy reversal *in vivo*. Highly pure (>99%; Fig. 1a and Supplementary Fig. 1) octaploid hepatocytes isolated by fluorescence-activated cell sorting (FACS) from male mice hemizygous for Rosa26 (*lacZ*) were transplanted into female $Fah^{-/-}$ mice, a model for selective liver replacement[11], and the frequency of donor-derived hepatocytes was assessed after extensive (>70%) liver repopulation (Supplementary Fig. 2). Hepatocytes from repopulated recipient mice were loaded with Hoechst and fluorescein di-β-D-galactopyranoside (FDG), a fluorescent substrate of β-galactosidase (β-gal), and analysed by FACS. Donor-derived β-gal was expressed by octaploid (90 ± 2%), tetraploid (83 ± 5%) and diploid hepatocytes (59 ± 5%) (Fig. 1b). Donor-derived Fah expression was also detected in most octaploid (89 ± 3%), tetraploid (86 ± 4%) and diploid (67 ± 11%) hepatocytes (data not shown). The overall ploidy distribution of donor hepatocytes was the same as found in normal liver of an aged mouse. Diploid and octaploid hepatocytes proliferated at equivalent rates (Supplementary Fig. 3), thus eliminating the possibility that the high percentage of near-diploid donor-derived hepatocytes resulted from overgrowth by rare contaminating diploids. Cytogenetics confirmed the presence of reduced-ploidy donor-derived cells (Fig. 1c). Surprisingly, most donor cells were aneuploid, that is, had numerical chromosome gains and/or losses (Fig. 1d, e). In addition to analysing single hepatocytes, liver sections from repopulated livers were stained for donor markers. Whereas most Fah^{+} nodules contained both Y chromosome and β-gal activity, loss of either marker was found in ~5% of Fah^{+} repopulation nodules (Supplementary Fig. 4). Together, these findings indicate



Number	Gains	Losses
Diploid		
36	4, 5, 11	3, 7, 9, 10, 13, 15, 18
40	1, 3, 4, 17, 18	7, 11, 13, 14, 19
41	1, 2, 3x2, 8, 12, 14, X	5, 7, 11x2, 13, 16x2
45	4, 9, 12, 16, 17, X	14
Tetraploid		
73	15, Y	1, 6x2, 7, 10, 12x2, 13, 14
79	15	8x2
80	1	Y
82	9, Y	None

Figure 1 | Purified octaploid hepatocytes generate reduced-ploidy daughters *in vivo*. **a**, Hepatocytes were separated into ploidy populations by FACS (using Hoechst 33342 fluorescence and pulse width parameters) with 2c, 4c and 8c DNA content, corresponding to diploid, tetraploid and octaploid hepatocytes, respectively. Highly pure viable octaploid hepatocytes were collected. **b**, Hepatocytes isolated from mice repopulated by 8c donors were loaded with Hoechst plus FDG. All ploidy classes expressed β-gal ($n = 6$ recipients). **c**, **d**, Chromosome number (**c**) and representative aneuploid karyotypes (**d**) of donor-derived mouse chromosome Y positive hepatocytes in repopulated livers ($n = 3$). Chromosomal gains/losses are described relative to the nearest ploidy. **e**, Percentage of numerical aneuploidy in donor-derived hepatocytes from repopulated mice or wild-type hepatocytes from non-transplanted mice (average ± s.e.m.; see Supplementary Fig. 5 for details). $^{*}P < 0.006$; $^{**}P = 0.01$.

[1]Oregon Stem Cell Center, Papé Family Pediatric Research Institute, Oregon Health & Science University, Portland, Oregon 97239, USA. [2]Division of Hematology and Medical Oncology, Oregon Health & Science University, Portland, Oregon 97239, USA. [3]Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, Oregon 97239, USA. [4]Department of Pathology, Texas Children's Hospital, Houston, Texas 77030, USA.



c

Population	n, cells (experiments)	Expected FISH signals: No. of hCD46, Chr9	Expected FISH signals: Per cent hepatocytes	Skewed FISH signals, per cent hepatocytes
Freshly isolated 2c	997 (6)	1, 2	98.9	1.1
Freshly isolated 4c	1,132 (6)	2, 4	99.8	0.2
Cultured 4c (2c daughters)	465 (6)	1, 2	87.5*	12.5*
Cultured 4c (4c daughters)	565 (3)	2, 4	78.6*	21.4*

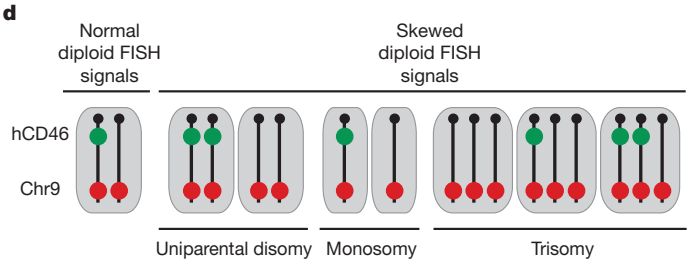


Figure 2 | Polyploid hepatocytes undergo ploidy reversal and unequal marker segregation *in vitro*. **a**, **b**, DNA content of cultured hepatocyte populations was determined by FACS. A representative plot is shown for freshly sorted 4c hepatocytes and cells expanded for 5 days (**a**). Results are summarized over multiple experiments (**b**). Data points represent average values ± s.e.m. **c**, ~99% of freshly isolated 2c or 4c hepatocytes had the expected number of hCD46 and mouse chromosome 9 signals. Cultured hepatocytes displayed chromosome mis-segregation. $*P < 0.0001$ compared to freshly isolated cells. **d**, Schematic of the observed FISH signals in 2c daughters derived from cultured 4c hepatocytes.

that most normal polyploid hepatocytes undergo ploidy reversal and marker segregation when forced to divide extensively.

The transplantation experiments demonstrating ploidy reversal were done using the *Fah* knockout mouse. Therefore, the high degree of aneuploidy observed (Fig. 1d, e) and marker loss (Supplementary Fig. 4) could be attributed to toxic metabolites made by *Fah* deficient hepatocytes[12]. To address this potential artefact, the karyotypes of hepatocytes from wild-type mice were determined. Although chromosome counts clustered around 40, 80 and 160 chromosomes, frequent chromosome gains and losses were detected in adult hepatocytes (Supplementary Fig. 5). At weaning, most hepatocytes were normal diploids, and by adulthood >60% of hepatocytes had numerical abnormalities (Fig. 1e). Thus, normal hepatocytes become aneuploid in adult mice, and this phenomenon is unrelated to *Fah* deficiency.

Owing to the time required for liver repopulation (6–8 weeks) and the dynamic ability of hepatocytes to change ploidy, it was unclear whether the ploidy reversal *in vivo* occurred in a single step or gradually over the course of multiple cell divisions. Therefore, we tracked ploidy changes using a short-term (1–2 mitoses) *in vitro* system. As expected, diploid hepatocytes became binucleated and polyploidized (Supplementary Fig. 6a, b). To investigate ploidy reversal, we followed the fate of polyploid hepatocytes. Concurrent with DNA replication and mitosis (data not shown), the percentage of binucleated cells dropped markedly, generating populations with >80% mononucleated cells (Supplementary Fig. 6a). After 5 days in culture pure tetraploid hepatocytes had produced daughters with 8c and 2c DNA content (18 ± 1% and 0.9 ± 0.1%, respectively) (Fig. 2a, b). Cultured octaploid hepatocytes showed similar ploidy redistribution (Supplementary Fig. 6c and Fig. 2b). The emergence of reduced-ploidy daughter cells over 1–2 cell cycles indicates a single-step mechanism for ploidy reversal.

To test whether polyploid hepatocyte mitosis could also produce aneuploidy, we analysed the frequency and nature of chromosome mis-segregation *in vitro*. Two-month-old mice hemizygous for a yeast artificial chromosome containing the human *CD46* gene (hCD46)[13] on mouse chromosome 9 were used (Supplementary Fig. 7). Using fluorescence *in situ* hybridization (FISH), chromosome signals in hepatocyte nuclei were quantified before and after proliferation (Supplementary Fig. 8a). After 5 days of expansion by 4c hepatocytes, the ploidy distribution shifted to include cells with 2c and 8c DNA content, as illustrated previously (Fig. 2a). Daughter cells with reduced ploidy (2c cells) and equal ploidy (4c cells) were hybridized with probes for hCD46 and mouse chromosome 9. Approximately 99% of hepatocytes analysed directly without prior culture (2c and 4c) contained the appropriate number of FISH signals (Fig. 2c). In contrast, FISH signals were skewed in 2c and 4c daughter cells (12.5% and 21.4%, respectively), indicating a diverse and aneuploid population of daughter cells (Fig. 2c, d and

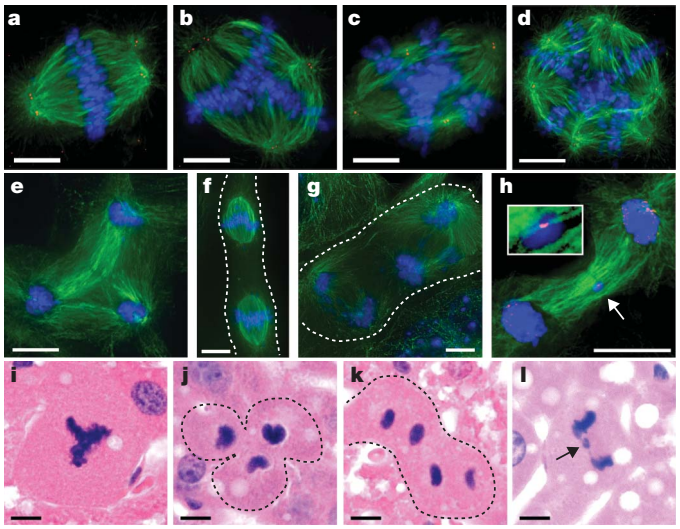


Figure 3 | Polyploid hepatocyte mitoses with multipolar spindles and chromosome segregation defects. **a**–**h**, Mitotic structures were detected in cultured hepatocytes by visualizing DNA (blue) and microtubules (green). Centrioles (red) were detected in **a**–**d**; centromeres (red) were detected in **h**. $n = 5$ experiments. 4c hepatocytes contained bipolar (**a**) or multipolar spindles (**b**, **c**). Multipolar spindles were also seen in 8c hepatocytes (**d**). Three daughter nuclei emerged from a tripolar telophase (**e**). Double mitosis was detected in metaphase (**f**) and anaphase (**g**). Lagging chromosomes (arrow and inset) were also seen (**h**). **i**–**l**, Hepatocytes dividing *in vivo* formed similar mitotic structures: multipolar spindles (**i**), tripolar division (**j**), double mitosis (**k**) and lagging chromosomes (**l**, arrow) ($n = 4$ mice). Dashed lines delineate cell borders. All scale bars are 10 μm.

Supplementary Fig. 8b). Notably, approximately one-third of the 2c cells displayed uniparental disomy for chromosome 9. These data show that proliferating hepatocytes routinely generate a genetically diverse population of daughter cells.

The mechanism by which hepatocytes generated aneuploid or reduced-ploidy daughter cells was unknown. We hypothesized that increased numbers of centrosomes in polyploid hepatocytes could lead to multipolar divisions and/or chromosome mis-segregation. To test this idea, cell divisions by polyploid hepatocytes were analysed *in vitro*. In about half of tetraploid mitoses, bipolar spindles were established and maintained by centrosome clustering (Fig. 3a). The remaining tetraploid hepatocytes contained spindles oriented in a multipolar configuration, with centrosomes oriented on 3–4 distinct poles (reflecting either true multipolar spindles or alignment of prometaphase chromosomes from binucleated hepatocytes) (Fig. 3b, c). Octaploid hepatocytes established multipolar spindles with as many as eight poles (Fig. 3d). However, only ~1% of cells in anaphase or telophase were oriented with tripolar spindles (Fig. 3e). Additionally, hepatocytes with two discrete mitotic spindles synchronized in metaphase (Fig. 3f) and anaphase (Fig. 3g) were identified, an event that we called 'double mitosis'. Mitotic structures nearly identical to those seen in cultured hepatocytes were also observed during hepatocyte proliferation *in vivo* (Fig. 3i–l).

The high percentage of multipolar metaphases and the much lower frequency of multipolar anaphases/telophases was surprising. We hypothesized that multipolar spindles could represent a temporary step in mitosis, and are then reorganized to be bipolar. Similar spindle dynamics were recently documented in cancer cells containing supernumerary centrosomes[14]. A consequence of multipolar spindle reorientation is chromosome mis-segregation. Microtubules from different poles can attach to a single kinetochore, and failure to repair such merotelic attachments can lead to incomplete chromosome segregation[14]. Consistent with spindle reorganization, we identified lagging chromosomes in 25–50% of tetraploid hepatocytes undergoing bipolar anaphase (Fig. 3h), indicating that merotely contributes to marker loss (that is, aneuploidy and loss-of-heterozygosity) observed in proliferating polyploid hepatocytes.

To determine whether multipolar mitoses produced viable offspring, we monitored hepatocyte divisions by time-lapse microscopy. As expected, diploids completed bipolar cell division with successful (89%) or failed cytokinesis (7%) (Fig. 4a). For analysis of polyploid hepatocytes, we focused mostly on tetraploids, but similar findings were seen with octaploids and non-fractionated hepatocytes (that is, a mixture of all ploidy classes that were never exposed to Hoechst or subjected to FACS). Nearly 90% of tetraploid hepatocytes divided in a bipolar manner (Fig. 4a, Supplementary Fig. 9 and Supplementary Movie 1), and in many time-lapse sequences their daughters (14%) divided again. Approximately 7% of tetraploid mitoses failed to complete cytokinesis (Fig. 4a, Supplementary Fig. 10 and Supplementary

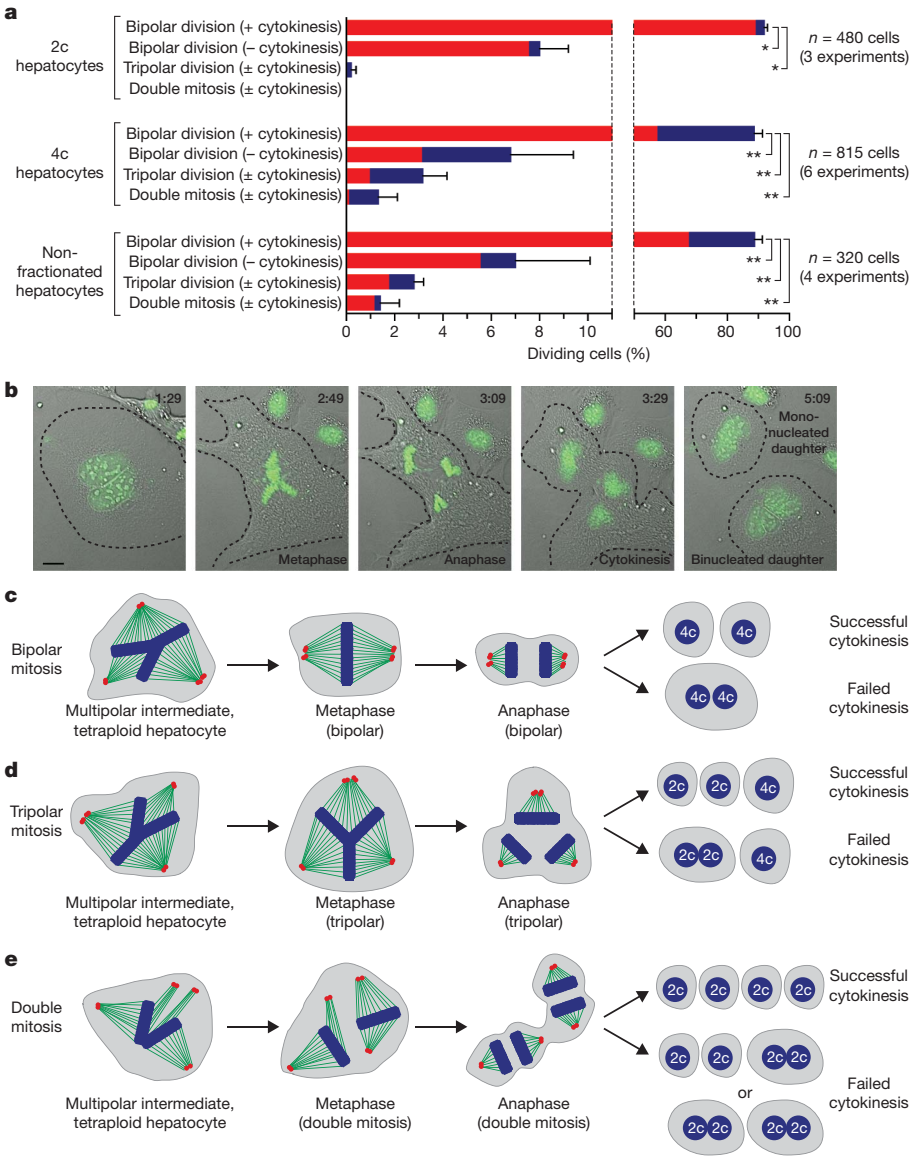


Figure 4 | Live cell imaging of multipolar mitoses in hepatocytes. **a**, Summary of mitotic events captured by time-lapse microscopy. The percentage (average ± s.e.m.) of dividing mononucleated (red) and binucleated (blue) hepatocytes is shown. $*P < 0.001$; $**P < 0.0001$. **b**, Images (related to Supplementary Movie 3) of a binucleated tetraploid hepatocyte undergoing tripolar division. Time (hours:minutes) is indicated. Mitosis was tracked with DIC and Hoechst (pseudocoloured green). Cell boundaries are marked. Scale bar, 10 μm. **c–e**, Schematic depiction of divisions completed by polyploid (mono- or binucleated) hepatocytes. The outcome of each division depends on spindle reorganization and completion/failure of cytokinesis. Metaphase chromosomes (blue), mitotic spindles (green) and centrosomes (red) are indicated.

Movie 2). Half of mononucleated and binucleated tetraploids transitioned from an early multipolar intermediate (as seen by chromosome alignment along multiple axes) to a standard bipolar configuration (Fig. 4c). Mitotic arrest and apoptosis were never seen.

In addition to standard divisions producing two daughters, mitosis along multiple axes was seen. For example, 3.2% of tetraploid hepatocytes (mononucleated and binucleated) were captured undergoing tripolar mitosis (Fig. 4a, b and Supplementary Movie 3). All of the daughter cells were viable for the duration of the imaging session (up to 16 h), and in some cases (15% of daughters) we were able to film subsequent mitoses (Supplementary Fig. 11 and Supplementary Movie 4). Although ~10% of tripolar divisions completed three-way cytokinesis, most divisions (~90%) ended in partial failed cytokinesis (Fig. 4d). Nuclear content frequently segregated in a 4:2:2 ratio, which is consistent with one tetraploid daughter nucleus and two reduced-ploidy diploids (Supplementary Fig. 12). Furthermore, 1.2% of tetraploid hepatocytes completed double mitotic events (Fig. 4a). Double mitoses by either binucleated (Supplementary Fig. 13 and Supplementary Movie 5) or mononucleated tetraploids (Supplementary Fig. 14 and Supplementary Movie 6) generated four distinct nuclei via two synchronized mitoses (Fig. 4e). By definition, generation of four mononucleated cells from a parental cell represents a ploidy-reversal event. Daughters were viable and appeared healthy throughout imaging (as long as 10 h).

Our data demonstrate that hepatocytes can increase (failed cytokinesis) and reduce (multipolar mitosis) their ploidy, thus resulting in the concept of a ploidy conveyor. This dynamic mechanism not only generates numerical chromosome abnormalities, but also uniparental chromosome sets. Given that 5–10% of all genes are thought to be monoallelically expressed[15], this segregation pattern produces tremendous genetic heterogeneity.

The pervasive presence of aneuploid genotypes in the liver raises the question of whether this phenomenon serves a physiological purpose. Studies in yeast showed that aneuploidy can provide a strong selective advantage in response to multiple environmental stressors[16]. Our findings indicate the possibility that hepatocyte polyploidization evolved precisely to result in subsequent ploidy reversal, aneuploidy and genetic diversity. Therefore, hepatic injury, which produces liver regeneration, could result in selection of hepatocytes that are genetically most resistant to the injury from a pre-existing pool of diverse genotypes. Genetic analysis of hepatocytes after liver injury may reveal favourable genotypes that differ from the germ line. Indeed, our own group has already observed an example of 'favourable loss-of-heterozygosity' followed by selection in the *Fah*-deficient mouse[17].

METHODS SUMMARY

Mouse surgery. The Institutional Animal Care and Use Committee of Oregon Health & Science University approved all mouse experiments. Hepatocytes from 3–6-month-old donors were transplanted intrasplenically[9,12]. Two-thirds partial hepatectomy was described previously[18]. Livers were harvested after 44 h, a time point when the maximum number of hepatocyte mitotic structures are found[19].

Cell isolation and flow cytometry. Primary hepatocytes were isolated by two-step collagenase perfusion[11], and cultured hepatocytes were isolated by trypsinization. Detailed protocols are described for isolation/analysis of hepatocyte ploidy populations by flow cytometry in Methods.

Hepatocyte expansion *in vitro*. FACS-purified hepatocytes from 3–6-month-old mice were seeded at 1,000 to 2,000 cells per cm^2 on either Primaria tissue culture plastic (Beckton Dickinson), collagen-coated Lab-Tek II, CC2-treated chamber slides (Nunc) or collagen-coated ibi-Treat 8-well μ-slides (Ibidi). Cells were initially incubated in hepatocyte culture medium containing DMEM with 4.5 g l^{-1} glucose (HyClone), 10% FBS (HyClone), non-essential amino acids (Cellgro) and antibiotic-antimycotic (Cellgro). After 24 h, the culture medium was replaced with SUM3 medium[20] supplemented with 0.5% FBS. Cells were cultured for defined intervals. Detailed protocols for *in vitro* assays are described in Methods: karyotype analysis; interphase FISH[21]; high-resolution imaging of mitotic structures; and time-lapse microscopy.

Histology and immunocytochemistry. Histological analyses[22] and Fah immunocytochemistry[8] were previously reported. The frequency of FACS-isolated hepatocytes in G2/M[23] was determined by cyclin A (Santa Cruz) staining.

Statistical significance. Statistical significance was determined using a 2-sided Student's *t*-test (Figs 1e, 2b and 4a) or Fisher's exact test (Fig. 2c). *P* values less than 0.05 were considered significant.

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

Received 5 March; accepted 11 August 2010.
Published online 22 September 2010.

1. Faktor, V. M. & Uryvaeva, I. V. Progressive polyploidy in mouse liver following repeated hepatectomy. *Tsitologiia* **17,** 909–916 (1975).
2. Guidotti, J. E. *et al.* Liver cell polyploidization: a pivotal role for binuclear hepatocytes. *J. Biol. Chem.* **278,** 19095–19101 (2003).
3. Kudryavtsev, B. N., Kudryavtseva, M. V., Sakuta, G. A. & Stein, G. I. Human hepatocyte polyploidization kinetics in the course of life cycle. *Virchows Arch. B Cell Pathol. Incl. Mol. Pathol.* **64,** 387–393 (1993).
4. Yim, A. P. Some flow-cytofluorimetric studies of the nuclear ploidy of mouse hepatocytes: iii. further observations on early changes in nuclear ploidy of mouse hepatocytes following various experimental procedures. *Br. J. Exp. Pathol.* **63,** 458–461 (1982).
5. Barbason, H., Van Cantfort, J. & Houbrechts, N. Correlation between tissular and division functions in the liver of young rats. *Cell Tissue Kinet.* **7,** 319–326 (1974).
6. Celton-Morizur, S., Merlen, G., Couton, D., Margall-Ducos, G. & Desdouets, C. The insulin/Akt pathway controls a specific cell division program that leads to generation of binucleated tetraploid liver cells in rodents. *J. Clin. Invest.* **119,** 1880–1887 (2009).
7. Margall-Ducos, G., Celton-Morizur, S., Couton, D., Bregerie, O. & Desdouets, C. Liver tetraploidization is controlled by a new process of incomplete cytokinesis. *J. Cell Sci.* **120,** 3633–3639 (2007).
8. Duncan, A. W. *et al.* Ploidy reductions in murine fusion-derived hepatocytes. *PLoS Genet.* **5,** e1000385 (2009).
9. Wang, X. *et al.* Cell fusion is the principal source of bone-marrow-derived hepatocytes. *Nature* **422,** 897–901 (2003).
10. Willenbring, H. *et al.* Myelomonocytic cells are sufficient for therapeutic cell fusion in liver. *Nature Med.* **10,** 744–748 (2004).
11. Overturf, K. *et al.* Hepatocytes corrected by gene therapy are selected *in vivo* in a murine model of hereditary tyrosinaemia type I. *Nature Genet.* **12,** 266–273 (1996).
12. Jorquera, R. & Tanguay, R. M. The mutagenicity of the tyrosine metabolite, fumarylacetoacetate, is enhanced by glutathione depletion. *Biochem. Biophys. Res. Commun.* **232,** 42–48 (1997).
13. Yannoutsos, N. *et al.* A membrane cofactor protein transgenic mouse model for the study of discordant xenograft rejection. *Genes Cells* **1,** 409–419 (1996).
14. Ganem, N. J., Godinho, S. A. & Pellman, D. A mechanism linking extra centrosomes to chromosomal instability. *Nature* **460,** 278–282 (2009).
15. Gimelbrant, A., Hutchinson, J. N., Thompson, B. R. & Chess, A. Widespread monoallelic expression on human autosomes. *Science* **318,** 1136–1140 (2007).
16. Rancati, G. *et al.* Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor. *Cell* **135,** 879–893 (2008).
17. Manning, K., Al-Dhalimy, M., Finegold, M. & Grompe, M. *In vivo* suppressor mutations correct a murine model of hereditary tyrosinemia type I. *Proc. Natl Acad. Sci. USA* **96,** 11928–11933 (1999).
18. Mitchell, C. & Willenbring, H. A reproducible and well-tolerated method for 2/3 partial hepatectomy in mice. *Nature Protocols* **3,** 1167–1170 (2008).
19. Ko, M. A. *et al.* Plk4 haploinsufficiency causes mitotic infidelity and carcinogenesis. *Nature Genet.* **37,** 883–888 (2005).
20. Darlington, G. J., Kelley, J. H. & Buffone, G. J. Growth and hepatospecific gene expression of human hepatoma cells in a defined medium. *In Vitro Cell. Dev. Biol.* **23,** 349–354 (1987).
21. Bayani, J. & Squire, J. A. Fluorescence *in situ* hybridization (FISH). *Curr. Protoc. Cell Biol.*. doi:10.1002/0471143030.cb2204s23 22 (2004).
22. Overturf, K. *et al.* Adenovirus-mediated gene therapy in a mouse model of hereditary tyrosinemia type I. *Hum. Gene Ther.* **8,** 513–521 (1997).
23. Pagano, M., Pepperkok, R., Verde, F., Ansorge, W. & Draetta, G. Cyclin A is required at two points in the human cell cycle. *EMBO J.* **11,** 961–971 (1992).

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

Acknowledgements We thank P. Canaday (Flow Cytometry Resource at OHSU) for cell sorting; A. Snyder and S. Kaech Petrie (Advanced Light Microscopy Core at OHSU, Core grant S10-RR023432) for microscopy assistance; and the Morphology Core of the Texas Medical Center (DK56338) for histology support. We also thank L. Smith and M. Thayer for discussions. This work was supported by grants from the National Institute of Health to M.G. (R01DK067636) and A.W.D. (F32DK076232).

Author Contributions A.W.D. designed and performed most of the experiments, analysed data and wrote the paper. M.H.T. helped with imaging of dividing hepatocytes. R.D.H. assisted with data analysis. A.E.H.N., M.L.L. and S.B.O. performed all of the cytogenetic analyses. Histological analyses were performed by M.J.F. M.G. supervised all aspects of this work. All authors discussed the results and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.W.D. (duncanan@ohsu.edu).

METHODS

Mouse strains and surgery. The Institutional Animal Care and Use Committee of Oregon Health & Science University approved all mouse experiments. The following inbred mouse strains were used: wild type (C57Bl, 129 and BALBc), transgenic Rosa26-*lacZ* (C57Bl and 129)[24], transgenic hCD46[13], *Fah*$^{-/-}$ (C57Bl and 129)[25]. F_1 hybrid mice (C57Bl × SJL) were also used. All mice were obtained from The Jackson Laboratories or bred at Oregon Health & Science University. Hepatocytes from 3–6-month-old donors were transplanted intrasplenically[9,26]. Two-thirds partial hepatectomy was described previously[18]. Livers were harvested after 44 h, a time point when the maximum number of hepatocyte mitotic structures are found[19].

Histology and immunocytochemistry. Histological analyses[22] and Fah immunocytochemistry[8] were previously reported. The frequency of FACS-isolated hepatocytes in G2/M[23] was determined by cyclin A (Santa Cruz) staining.

Isolation/analysis of defined hepatocyte ploidy populations. Primary hepatocytes were isolated by two-step collagenase perfusion[11] and cultured hepatocytes were isolated by trypsinization. For detection of hepatocyte ploidy populations, hepatocytes (2×10^6 per ml) were incubated with 15 μg ml^{-1} Hoechst 33342 (Sigma) and 5 μM reserpine (Invitrogen) for 30 min at 37 °C. Cells were analysed and/or sorted with an InFlux flow cytometer (Beckton Dickinson) using a 150 μm nozzle. Dead cells were excluded on the basis of 5 μg ml^{-1} propidium iodide (Invitrogen) incorporation. Cells adhering to each other (that is, doublets) were eliminated on the basis of pulse width. Blood cells were excluded based on expression of CD45 using an APC-conjugated CD45 antibody (eBioscience). Ploidy populations were identified by DNA content using an ultraviolet 355-nm laser and 425–440-nm bandpass filter. Gating strategy is described (Supplementary Fig. 1a). Sorted hepatocytes were collected in DMEM with 4.5 g l^{-1} glucose (HyClone) containing 50% fetal bovine serum (FBS) (HyClone). The purity of sorted populations was determined at the end of each sort, and only highly purified populations (>99% pure) were used for subsequent assays. β-gal activity in Hoechst-stained hepatocytes was detected using FDG reagent (Invitrogen) as per the manufacturer's instructions. Expression of hCD46 was detected with PE-conjugated hCD46 antibody (eBioscience).

Hepatocyte expansion *in vitro*. FACS-purified hepatocytes from 3–6-month-old mice were seeded at 1,000 to 2,000 cells cm^{-2} on either Primaria tissue culture plastic (Beckton Dickinson), collagen-coated Lab-Tek II, CC2-treated chamber slides (Nunc) or collagen-coated ibi-Treat 8-well μ-slides (Ibidi). Cells were initially incubated in hepatocyte culture medium containing DMEM with 4.5 g l^{-1} glucose (HyClone), 10% FBS (HyClone), non-essential amino acids (Cellgro) and antibiotic-antimycotic (Cellgro). After 24 h, the culture medium was replaced with SUM3 medium supplemented with 0.5% FBS. Cells were cultured for defined intervals.

SUM3 medium. SUM3 was prepared with the following reagents[20]: 75% DMEM with 4.5 g l^{-1} glucose (HyClone), 25% Waymouth's MB 752/1 (Invitrogen), 2 mM L-glutamine (Invitrogen), antibiotic-antimycotic (Cellgro), 10 mM HEPES (Fisher), 50 ng ml^{-1} human epidermal growth factor (Invitrogen), 1.0 μg ml^{-1} insulin (Sigma), 30 nM sodium selenite (Sigma), 10 μg ml^{-1} transferrin (Sigma), 50 ng ml^{-1} somatotropin (Sigma) and 1.0 μM 3,3′,5-triiodo-L-thyronine sodium salt (T_3) (Sigma).

Cytogenetics: karyotypes and FISH. For karyotypes, freshly isolated primary hepatocytes were seeded on Primaria tissue culture plastic (Beckton Dickinson) and incubated in medium containing DMEM with 4.5 g l^{-1} glucose (HyClone), 10% FBS (HyClone), non-essential amino acids (Cellgro) and antibiotic-antimycotic (Cellgro). Non-adherent cells were removed after 4 h, and cultures were switched to SUM3 medium supplemented with 0.5% FBS. Next, after ~40 h, hepatocytes were treated with 150 mg ml^{-1} colcemid (Sigma) for 2–4 h and harvested by trypsinization. Then, after extensive washing, cells were incubated for 10 min in 56 mM KCl with 5% FBS and fixed with methanol:acetic acid (3:1 ratio). Finally, chromosomes from ~20 metaphase-arrested hepatocytes per sample were G-banded with a standard trypsin/Wright's stain protocol. Photographs were taken using CytoVision software from Applied Imaging.

For FISH probe generation, FISH point probes for mouse chromosome Y, mouse chromosome 9 and hCD46 were developed from bacterial artificial chromosomes RP24-225i24 (CHORI), RP23-346E22 (Invitrogen) and RP11-454L1 (Invitrogen), respectively. Bacterial artificial chromosomes were nick translated with the CGH Nick Translation Kit (Abbott Molecular) using SpectrumOrange dUTP and SpectrumGreen dUTP (Abbott Molecular). Each probe was validated using β-inverse calculation to establish its normal cutoff value.

For FISH analysis, FACS-sorted cells were dropped onto slides, coated with 3:1 methanol:acetic acid, and allowed to dry. Further fixation involved incubation in 3:1 methanol:acetic acid for 5 min at room temperature (20–23 °C). For ageing, slides were baked for 5 min at 95 °C. Slides were incubated in 2×SSC for 30 min at 37 °C, followed by 0.005% pepsin (Sigma)/0.01 N HCl (Sigma) for 13 min at 37 °C. Slides were rinsed in 1×PBS (Sigma) for 5 min at room temperature and then fixed in 1% formaldehyde/0.45% $MgCl_2$ (Sigma)/PBS for 5 min at room temperature. Slides were again rinsed in 1×PBS for 5 min at room temperature and then dehydrated by 2 min washes in 70%, 80% and 90% ethanol. Once slides dried, probes were added and coverslips sealed with rubber cement. Slides were incubated at 72 °C for 2 min and 37 °C overnight. After incubation, slides were washed in 0.4×SSC + 0.3% NP-40 (Sigma) at 72 °C for 2 min, placed in 2×SSC + 0.15% NP-40 for 5 s to 1 min at room temperature, and mounted with DAPI II (Abbott). Samples were analysed and scored under a Nikon Eclipse E800 photoscope. For analysis of interphase nuclei, signals were counted for as many nuclei available, up to 200 for each sample. Signals were scored as separate if they were further than a signal's diameter apart in distance, according to established clinical practices of FISH signal interpretation and as described[21]. Photographs were captured using CytoVision software from Applied Imaging.

High-resolution imaging of dividing hepatocytes. Mitotic structures identified *in vitro* were detected in populations of hepatocytes from 2–4-month-old F_1 hybrid mice (SJL × C57Bl) growing on collagen-coated ibi-Treat 8-well μ-slides (Ibidi). After expansion for 2–4 days, wells were fixed and immunostained as described[14]. Briefly, cells were fixed with methanol and incubated with primary antibodies for α-tubulin (clone DM1A, Sigma), Centrin-2 (N-17, Santa Cruz) or Centromere proteins (15-235, Antibodies, Incorporated). Primary antibodies were detected with species-specific secondaries conjugated to Alexafluor 488 or Alexafluor 555 (Invitrogen). Nuclei were stained with 40 μg ml^{-1} Hoechst 33342 (Invitrogen). The images were acquired on a high-resolution wide-field Core DV system (Applied Precision) equipped with a Nikon Coolsnap ES2 HQ camera. This system is an Olympus IX71 inverted microscope with a proprietary *XYZ* stage enclosed in a controlled environment chamber, DIC transmitted light and a short arc 250W Xenon lamp for fluorescence. Each image was acquired as Z-stacks (every 0.2 μm from the bottom to top of each mitotic spindle). Images were deconvolved with the appropriate optical transfer function using an algorithm of ten iterations using SoftWoRx Image Restoration Software (Applied Precision).

Live-cell imaging. Live-cell imaging was performed with a high-resolution wide-field Core DV system with controlled environment chamber (~37 °C, 5% CO_2) (Applied Precision). Hepatocytes from 2–4-month-old F_1 hybrid mice (SJL × C57Bl), growing on collagen-coated ibi-Treat 8-well μ-slides (Ibidi), were imaged between days 2 and 4. Optimal conditions were used. Images were collected every 10 min over sessions lasting 16–24 h. Cellular structures were visualized with DIC (~0.05 s exposure) and chromosomes visualized with Hoechst (0.6 s exposure). Hoechst was originally used to identify defined ploidy populations by FACS; residual Hoechst remained intercalated during cell culture. The focal plane was optimally set at the beginning of each imaging session; however, owing to movement and shape change, cells appeared to drift slightly out-of-focus. Time-lapse movies were formatted using SoftWoRx Explorer (Applied Precision) and exported in QuickTime (Apple). The amount of Hoechst signal (corresponding to nuclear content) was determined for a subset of dividing hepatocytes using Imaris image analysis software (Bitplane).

Statistical significance. Statistical significance was determined using two-sided Student's *t*-test (Figs 1e, 2b and 4a) or Fisher's exact test (Fig. 2c). *P* values less than 0.05 were considered significant.

24. Friedrich, G. & Soriano, P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* **5,** 1513–1523 (1991).
25. Grompe, M. *et al.* Loss of fumarylacetoacetate hydrolase is responsible for the neonatal hepatic dysfunction phenotype of lethal albino mice. *Genes Dev.* **7,** 2298–2307 (1993).
26. Lagasse, E. *et al.* Purified hematopoietic stem cells can differentiate into hepatocytes *in vivo*. *Nature Med.* **6,** 1229–1234 (2000).

doi:10.1038/nature09397

The role of toxin A and toxin B in *Clostridium difficile* infection

Sarah A. Kuehne[1]*, Stephen T. Cartman[1]*, John T. Heap[1], Michelle L. Kelly[1], Alan Cockayne[1] & Nigel P. Minton[1]

***Clostridium difficile* infection is the leading cause of healthcare-associated diarrhoea in Europe and North America[1,2]. During infection, *C. difficile* produces two key virulence determinants, toxin A and toxin B. Experiments with purified toxins have indicated that toxin A alone is able to evoke the symptoms of *C. difficile* infection, but toxin B is unable to do so unless it is mixed with toxin A or there is prior damage to the gut mucosa[3]. However, a recent study indicated that toxin B is essential for *C. difficile* virulence and that a strain producing toxin A alone was avirulent[4]. This creates a paradox over the individual importance of toxin A and toxin B. Here we show that isogenic mutants of *C. difficile* producing either toxin A or toxin B alone can cause fulminant disease in the hamster model of infection. By using a gene knockout system[5,6] to inactivate the toxin genes permanently, we found that *C. difficile* producing either one or both toxins showed cytotoxic activity *in vitro* that translated directly into virulence *in vivo*. Furthermore, by constructing the first ever double-mutant strain of *C. difficile*, in which both toxin genes were inactivated, we were able to completely attenuate virulence. Our findings re-establish the importance of both toxin A and toxin B and highlight the need to continue to consider both toxins in the development of diagnostic tests and effective countermeasures against *C. difficile*.**

Toxin A and toxin B both catalyse the glucosylation, and hence inactivation, of Rho-GTPases: small regulatory proteins of the eukaryotic actin cell cytoskeleton. This leads to disorganization of the cell cytoskeleton and cell death[7]. The toxin genes *tcdA* and *tcdB* are situated on the *C. difficile* chromosome in a 19.6-kilobase (kb) pathogenicity locus (PaLoc), along with the three accessory genes *tcdC*, *tcdR* and *tcdE* (Fig. 1a). To address the individual importance of toxin A and toxin B, we used the ClosTron gene knockout system[6] to inactivate the toxin genes of *C. difficile*. This system inactivates genes by inserting an intron into the protein-encoding DNA sequence of a gene, thus resulting in a truncated and non-functional protein. The intron sequence itself encompasses an erythromycin-resistance determinant that permits selective isolation of mutants. Furthermore, it has been shown experimentally that the insertions are completely stable, meaning that inactivation of a gene is permanent[5].

Using the ClosTron system, we targeted insertions to *tcdA* and *tcdB* at nucleotide positions 1584 and 1511, respectively (Fig. 1a). In both cases, this placed the intron within DNA sequence encoding the toxin catalytic domain. Three separate isogenic mutants of the toxin-A-positive, toxin-B-positive (A^+B^+) *C. difficile* strain 630Δ*erm*[8] were constructed: two 'single mutants', with toxin profiles A^-B^+ and A^+B^-, respectively, and a 'double mutant' with toxin profile A^-B^-. The A^-B^- double mutant was made from the A^+B^- single mutant by targeting *tcdA* with a second intron that carried the chloramphenicol/thiamphenicol-resistance gene *catP* instead of the usual erythromycin-resistance determinant.

The genotype of each toxin mutant was characterized by polymerase chain reaction (PCR) and DNA sequence analysis to confirm the exact location of each intron insertion made (data not shown). Southern blot analysis of EcoRV-digested genomic DNA samples, using an intron-specific probe, confirmed that the A^-B^+ and A^+B^- mutants each had a single insertion, whereas the A^-B^- mutant had a double insertion (Fig. 1b). It is noteworthy that three bands were expected for the A^-B^- double-mutant strain because the *catP* gene harbours an EcoRV site. The phenotype of each strain was confirmed by western blot analysis. Use of a toxin-A-specific antibody probe confirmed that the A^-B^+ and A^-B^- mutants no longer produced toxin A (Fig. 1c). Likewise, use of a toxin-B-specific antibody probe confirmed that the A^+B^- and A^-B^- mutants no longer produced toxin B (Fig. 1d).

Subsequently, *in vitro* cell cytotoxicity assays were carried out using HT29 (human colon carcinoma) cells and Vero (African green monkey kidney) cells. Each of these cell lines is susceptible to both toxin A and toxin B, although HT29 cells are more sensitive to toxin A and Vero cells are more sensitive to toxin B[9]. The action of toxin A and toxin B causes the cells to 'round' (that is, lose morphology) and die; a phenomenon that is clearly visible by light microscopy. We incubated

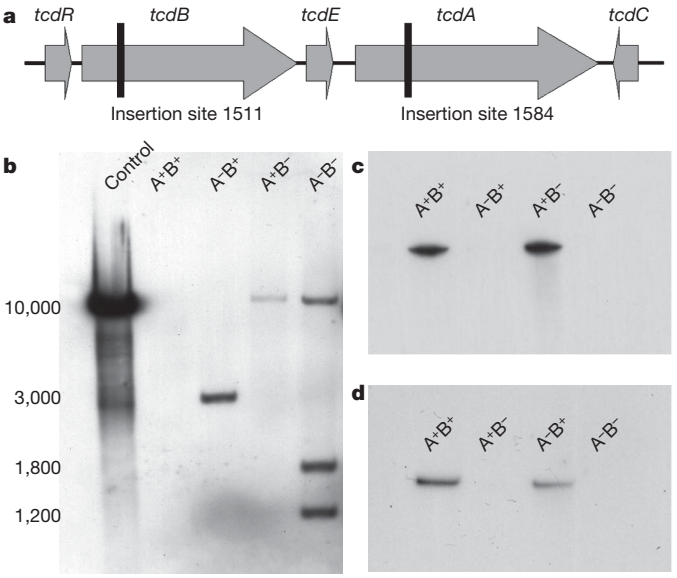


Figure 1 | Characterization of *C. difficile* toxin mutants. **a**, The pathogenicity locus PaLoc of *C. difficile* 630 showing the intron insertion sites for the toxin mutants. **b**, Southern blot using an intron-specific probe. The control plasmid (pMTL007C-E2) and the genomic DNA of the four strains was digested with EcoRV, which resulted in a band of approximately 9 kb for the plasmid, 10 kb for the *tcdB*-mutation (in A^+B^- and A^-B^-), just over 3 kb for the *tcdA*-single mutant (A^-B^+) and 1.8 and 1.2 kb for the *tcdA*-mutation in the double mutant, owing to an additional EcoRV site in the *catP* gene. **c**, Western blot probing culture supernatants with anti-TcdA antibody (tgcBIOMICS). **d**, Western blot probing culture supernatants with anti-TcdB antibody (tgcBIOMICS).

[1]Clostridia Research Group, Centre for Biomolecular Sciences, School of Molecular Medical Sciences, Nottingham Digestive Diseases Centre, NIHR Biomedical Research Unit, University of Nottingham, Nottingham NG7 2RD, UK.
*These authors contributed equally to this work.

cultured cells for 24 h with fourfold dilution series of *C. difficile* culture supernatants. To obtain the most objective data set possible, we determined the end-point titre of each dilution series, rather than implementing a subjective cell-scoring system. End-point titre was defined as the first dilution in a series for which HT29 or Vero cell morphology was indistinguishable from the negative controls (cells that had been incubated with uninoculated *C. difficile* culture medium).

As expected, the A^-B^- double-toxin mutant did not have any cytotoxic activity towards either HT29 or Vero cells (Fig. 2a, b). Compared to the A^+B^+ parental strain, the A^-B^+ mutant showed reduced toxicity towards HT29 cells, although the difference was not statistically significant, and a similar degree of toxicity towards Vero cells (Fig. 2a, b). These findings were anticipated given the respective sensitivities of HT29 cells and Vero cells to toxin A and toxin B[9]. However, unexpectedly, when compared to the A^+B^+ parental strain, the A^+B^- mutant showed increased toxicity towards HT29 cells, although the difference was not statistically significant, and a similar degree of toxicity towards Vero cells (Fig. 2a, b). We reasoned that this may occur owing to increased expression of toxin A by the A^+B^- mutant, a phenomenon that has been reported previously[4]. Indeed, quantitative PCR with reverse transcription (qRT–PCR) analysis confirmed that expression of toxin A was an average of 3.3-fold greater in the A^+B^- mutant than in the A^+B^+ parental strain (data not shown). We do not know the reason for this. However, considering the respective sensitivities of HT29 and Vero cells to toxin A[9], this finding explains our unexpected cytotoxicity results, even accounting for the fact that the A^+B^- mutant does not produce any toxin B.

To confirm that the cytotoxic activity we had observed on HT29 cells and Vero cells was indeed attributable to the respective actions of toxin A and toxin B, we carried out toxin neutralization assays. Culture supernatants of each *C. difficile* toxin mutant and the A^+B^+ parental strain were incubated with either toxin-A-specific or toxin-B-specific neutralizing antibodies before inoculation onto HT29 and Vero cell monolayers. Importantly, culture supernatants were diluted equivalently such that, for the A^+B^+ parental strain, only toxin A activity was detected on HT29 cells and only toxin B activity was detected on Vero cells (that is, toxin B activity towards HT29 cells was diluted out completely and toxin A activity towards Vero cells was diluted out completely) (Fig. 2c, d). As expected, the toxin-A-specific antibody neutralized all toxic activity produced by the A^+B^- mutant and the toxin-B-specific antibody neutralized all toxic activity produced by the A^-B^+ mutant (Fig. 2c, d). Interestingly, the increased production of toxin A by the A^+B^- mutant was clearly visible in this assay, as the cytotoxic activity of this strain towards Vero cells was not diluted out completely, as it was for the A^+B^+ parental strain (Fig. 2d).

Having fully characterized our *C. difficile* toxin mutants *in vitro*, we tested the virulence of each in the hamster model of infection. Hamsters were each challenged with 100 spores of a single *C. difficile* strain, 5 days after an oral dose of clindamycin (30 mg kg^{-1}). Each toxin mutant and the A^+B^+ parental strain were administered to eight hamsters in total. All hamsters became colonized by the *C. difficile* strain administered between 1 and 3 days after challenge, with the exception of one that received the A^+B^- single toxin mutant (Fig. 3a). Following colonization, hamsters that received the A^+B^+ parental strain, the A^-B^+ mutant or the A^+B^- mutant all developed symptoms of *C. difficile* infection, which resulted in a mean time to death of 1.0 day, 1.3 days and 4.0 days, respectively (Fig. 3b). In contrast, none of the hamsters colonized by the toxin-null A^-B^- double mutant developed any symptoms of disease during the 14-day experimental period, indicating that this strain is completely attenuated for virulence. Bacteriological and PCR analysis of caecum samples taken from each hamster post mortem confirmed that the only infecting strain of *C. difficile* was, indeed, the strain administered in every case, thus

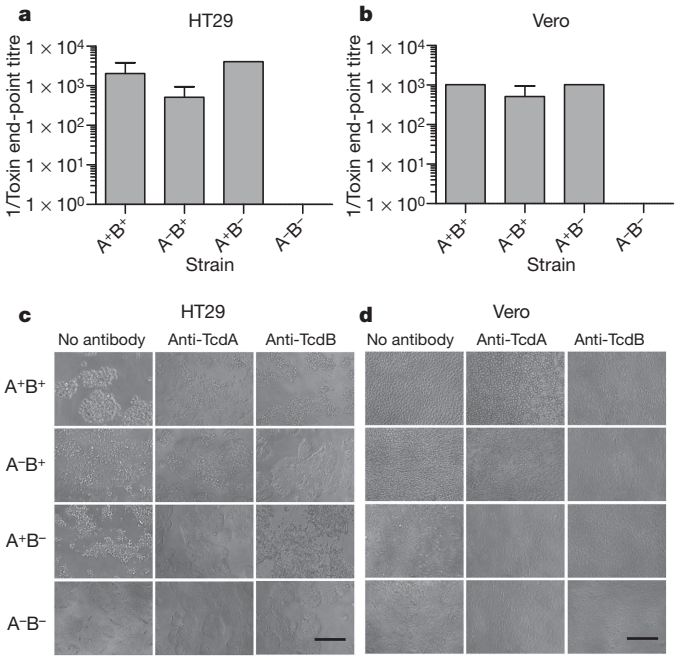


Figure 2 | *In vitro* cytotoxicity. **a**, **b**, Supernatants of the parental strain A^+B^+ and the three mutants A^-B^+, A^+B^- and A^-B^- were used in cell culture assays to measure cytotoxicity. HT29 cells (**a**) and Vero cells (**b**) were cultured to a flat monolayer before adding *C. difficile* supernatants in fourfold dilutions series. After a 24-h incubation, toxin end-point titres were determined. Data represent the mean ± standard deviation; $n = 3$. **c**, **d**, Toxin neutralization assays. Appropriate dilutions of supernatants were pre-incubated with a suitable concentration of anti-TcdA or anti-TcdB serum for 1 h at 37 °C and then added to HT29 cells (**c**) and Vero cells (**d**), which were evaluated after 24 h. Scale bars, 2 mm.

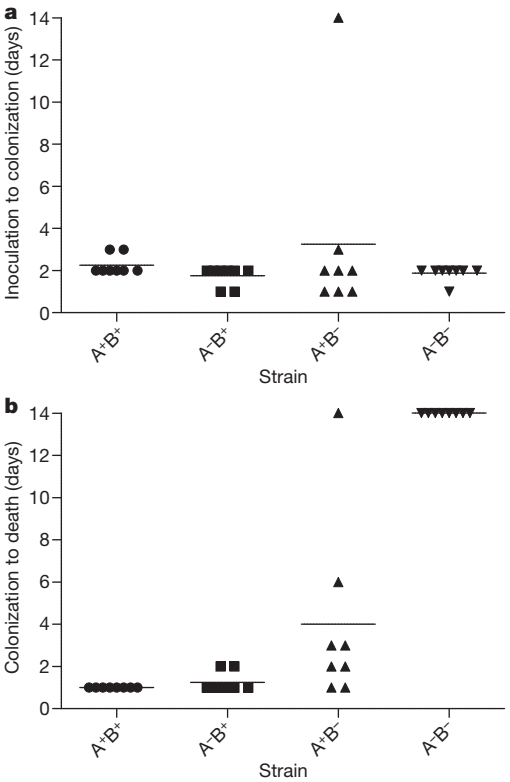


Figure 3 | Virulence of *C. difficile* strains in hamsters. **a**, **b**, Groups of eight hamsters were challenged with *C. difficile* 630Δ*erm* (A^+B^+), or one of the toxin mutant strains, A^+B^-, A^-B^+ or A^-B^-. **a**, Colonization of Golden Syrian hamsters by each strain is presented as time from inoculation to colonization in days ($n = 8$). **b**, Time from colonization to death. The duration of the experiment was set at 14 days ($n = 8$).

ruling out any possibility of cross-contamination between cages or contamination from the environment (Supplementary Fig. 1).

In conclusion, it is clear that both toxin A and toxin B have an important role in *C. difficile* infection, because here we have shown that strains that produce either toxin on its own or both toxins together are virulent. It is pertinent to question why we found that an A^+B^- strain of *C. difficile* was virulent; a result that is in direct contrast with a similar study published recently[4]. This discrepancy may have arisen owing to inherent differences between the hamsters used in each study. However, perhaps more probable is that there is one or more key differences between the strains of *C. difficile* studied. Although both strains are erythromycin-sensitive derivatives of strain 630 (refs 10, 11), they were isolated independently through serial sub-culture[8,12]. Therefore, either strain could have acquired one or more secondary mutations, which may affect the action of either one or both of the toxins. However, it is notable that our findings accord with those of previous studies that have suggested a role for both toxin A and toxin B in *C. difficile* infection[3,13–15]. Moreover, given that the human colon is the principal site of pathology in patients infected with *C. difficile*, it stands to reason that an A^+B^- strain of *C. difficile* is virulent *in vivo*, as it is toxin A that shows the greatest cytotoxicity towards laboratory-cultured HT29 cells. It is important to note that inherent variability exists between the toxins of some *C. difficile* strains, particularly in the case of toxin B[16–19]. In practical terms, this means that the toxins from different strains can vary in enzymatic activity (that is, different GTPase substrates may be glucosylated) and/or host-cell specificity. Consequently, it is not appropriate to over-interpret our findings and make general conclusions about the toxins produced by all toxigenic strains of *C. difficile*. Nonetheless, our results clearly demonstrate that a strain of *C. difficile* producing either toxin A or toxin B alone may be virulent and thus we have re-established the importance of both toxins in *C. difficile* infection.

It is interesting to note that a number of clinical cases of *C. difficile* infection have been attributed to naturally occurring A^-B^+ strains[20,21], but that there have been no reports of naturally occurring A^+B^- isolates until now. This would suggest that A^+B^- strains do not exist, but it may also be an artefact of routine diagnostic testing practices. Either way, our results show that A^+B^- strains may be virulent and even if they do not exist in nature already, they may yet evolve. Consequently, it is imperative that both toxin A and toxin B continue to be considered in routine diagnostic settings and in the development of effective countermeasures against *C. difficile*.

METHODS SUMMARY

Mutants were constructed from the parental strain *C. difficile* 630Δ*erm*[8] using the ClosTron system[5,6]. The retargeted plasmids pMTL007C-E2::Cdi-*tcdA*-1584s, pMTL007C-E2::Cdi-*tcdB*-1511a and, for the double mutant, pMTL007S-C7::Cdi-*tcdA*-1584s were transferred into *C. difficile* via conjugation. The single ClosTron mutants were isolated on erythromycin plates. The double mutant was isolated on thiamphenicol plates.

For cytotoxicity assays, the four strains were grown overnight in 5 ml TY (3% w/v bacto-tryptose, 2% w/v yeast extract and 0.1% w/v thioglycolate, adjusted to pH 7.4) under anaerobic conditions as previously described[22]. The cell densities were standardized before centrifugation and filtration. Supernatants were diluted in a fourfold series and 20 μl of dilutions were added onto monolayers of Vero and HT29 cells preincubated in 96-well plates for 48 h (at 37 °C, 5% CO_2). Cytotoxicity was recorded after 24 h. Statistical analysis was performed using one-way ANOVA tests. For the neutralization assay, appropriate dilutions of supernatants were pre-incubated with a suitable concentration of anti-TcdA or anti-TcdB serum (polyclonal, tgcBIOMICS) for 1 h at 37 °C and then added onto cell monlayers of Vero and HT29 cells as described earlier and evaluated after 24 h. Golden Syrian hamsters were dosed with clindamycin (30 mg kg^{-1}) 5 days before being infected orally with 100 spores each. Hamsters were monitored for signs of infection (including weight loss, behavioural changes and wet tail) and killed when the end point was met. Faecal pellets were collected daily and plated to confirm the presence or absence of *C. difficile*. Caecum samples were homogenized, plated and *C. difficile* counts obtained. PCR was performed to determine the *C. difficile* genotype isolated from all samples. See Supplementary Information for PCR results.

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

Received 14 January 2010; accepted 6 August 2010.
Published online 15 September 2010.

1. Poutanen, S. M. & Simor, A. E. *Clostridium difficile*-associated diarrhea in adults. *Can. Med. Assoc. J.* **171,** 51–58 (2004).
2. Elliott, B., Chang, B. J., Golledge, C. L. & Riley, T. V. *Clostridium difficile*-associated diarrhoea. *Intern. Med. J.* **37,** 561–568 (2007).
3. Lyerly, D. M., Saum, K. E., MacDonald, D. K. & Wilkins, T. D. Effects of *Clostridium difficile* toxins given intragastrically to animals. *Infect. Immun.* **47,** 349–352 (1985).
4. Lyras, D. *et al.* Toxin B is essential for virulence of *Clostridium difficile. Nature* **458,** 1176–1179 (2009).
5. Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P. & Minton, N. P. The ClosTron: a universal gene knock-out system for the genus *Clostridium. J. Microbiol. Methods* **70,** 452–464 (2007).
6. Heap, J. T. *et al.* The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. *J. Microbiol. Methods* **80,** 49–55 (2010).
7. Just, I. & Gerhard, R. Large clostridial cytotoxins. *Rev. Physiol. Biochem. Pharmacol.* **152,** 23–47 (2004).
8. Hussain, H. A., Roberts, A. P. & Mullany, P. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630Δ*erm*) and demonstration that the conjugative transposon Tn*916*ΔE enters the genome of this strain at multiple sites. *J. Med. Microbiol.* **54,** 137–141 (2005).
9. Torres, J., Camorlinga-Ponce, M. & Munoz, O. Sensitivity in culture of epithelial cells from rhesus monkey kidney and human colon carcinoma to toxins A and B from *Clostridium difficile. Toxicon* **30,** 419–426 (1992).
10. Sebaihia, M. *et al.* The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nature Genet.* **38,** 779–786 (2006).
11. Hachler, H., Berger-Bachi, B. & Kayser, F. H. Genetic characterization of a *Clostridium difficile* erythromycin-clindamycin resistance determinant that is transferable to *Staphylococcus aureus. Antimicrob. Agents Chemother.* **31,** 1039–1045 (1987).
12. O'Connor, J. R. *et al.* Construction and analysis of chromosomal *Clostridium difficile* mutants. *Mol. Microbiol.* **61,** 1335–1351 (2006).
13. Kim, P. H., Iaconis, J. P. & Rolfe, R. D. Immunization of adult hamsters against *Clostridium difficile*-associated ileocecitis and transfer of protection to infant hamsters. *Infect. Immun.* **55,** 2984–2992 (1987).
14. Voth, D. E. & Ballard, J. D. *Clostridium difficile* toxins: mechanism of action and role in disease. *Clin. Microbiol. Rev.* **18,** 247–263 (2005).
15. Du, T. & Alfa, M. J. Translocation of *Clostridium difficile* toxin B across polarized Caco-2 cell monolayers is enhanced by toxin A. *Can. J. Infect. Dis.* **15,** 83–88 (2004).
16. Chaves-Olarte, E. *et al.* R-Ras glucosylation and transient RhoA activation determine the cytopathic effect produced by toxin B variants from toxin A-negative strains of *Clostridium difficile. J. Biol. Chem.* **278,** 7956–7963 (2003).
17. Torres, J. F. Purification and characterisation of toxin B from a strain of *Clostridium difficile* that does not produce toxin A. *J. Med. Microbiol.* **35,** 40–44 (1991).
18. Alfa, M. J. *et al.* Characterization of a toxin A-negative, toxin B-positive strain of *Clostridium difficile* responsible for a nosocomial outbreak of *Clostridium difficile*-associated diarrhea. *J. Clin. Microbiol.* **38,** 2706–2714 (2000).
19. Stabler, R. A., Dawson, L. F., Phua, L. T. & Wren, B. W. Comparative analysis of BI/NAP1/027 hypervirulent strains reveals novel toxin B-encoding gene (*tcdB*) sequences. *J. Med. Microbiol.* **57,** 771–775 (2008).
20. Drudy, D., Fanning, S. & Kyne, L. Toxin A-negative, toxin B-positive *Clostridium difficile. Int. J. Infect. Dis.* **11,** 5–10 (2007).
21. Drudy, D., Harnedy, N., Fanning, S., Hannan, M. & Kyne, L. Emergence and control of fluoroquinolone-resistant, toxin A-negative, toxin B-positive *Clostridium difficile. Infect. Control Hosp. Epidemiol.* **28,** 932–940 (2007).
22. Heap, J. T., Pennington, O. J., Cartman, S. T. & Minton, N. P. A modular system for *Clostridium* shuttle plasmids. *J. Microbiol. Methods* **78,** 79–85 (2009).

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

Acknowledgements S.A.K., S.T.C., A.C. and N.P.M. acknowledge the financial support of the UK Medical Research Council (G0601176). Support for M.L.K. was provided by the European Union (HEALTH-F3-2008-223585) and for J.T.H. by the Biotechnology and Biological Sciences Research Council (BBSRC) (BB/F003390/1). We thank C. von-Eichel Streiber for supplying the neutralizing antibody serums and Y. Mahida for the Vero cells. We are also grateful to A. Olling and R. Gerhard for methodological advice on western blots.

Author Contributions The study was conceived by N.P.M. and designed by S.A.K., S.T.C. and J.T.H. Construction of mutants and *in vitro* characterization was carried out by S.A.K. *In vivo* work was carried out by S.T.C., M.L.K. and A.C. Analysis of data was carried out by S.A.K. and M.L.K. with assistance from S.T.C. and J.T.H. The manuscript was written by S.A.K. and S.T.C. with critical input from all other authors. Funding for the study was sourced by N.P.M. and A.C.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.P.M. (nigel.minton@nottingham.ac.uk).

METHODS

Strains and growth conditions. Strains used in this study were *Escherichia coli* TOP10 (Invitrogen) as a cloning host, *E. coli* CA434 (ref. 23) as a conjugal donor and *C. difficile* 630Δ*erm* (ref. 8) and mutants. All strains were stored at −80 °C stocks on arrival and have been maintained as frozen stocks ever since. *E. coli* cultures were grown on Luria Bertani medium, aerobically, at 37 °C and shaking if liquid unless stated otherwise. *C. difficile* cultures were grown in supplemented brain heart infusion medium (BHIS)[24] or TY[25], anaerobically, 37 °C in an anaerobic workstation (D. Whitley). Antibiotics were used at the following concentrations where appropriate: chloramphenicol (25 μg ml^{-1} or 12.5 μg ml^{-}), thiamphenicol (15 μg ml^{-1}), spectinomycin (250 μg ml^{-1} or 750 μg ml^{-1}), erythromycin (2.5 μg ml^{-1}), D-cycloserine (250 μg ml^{-1}) and cefoxitin (8 μg ml^{-1}).

Molecular biology techniques. Qiagen mini prep kits were used to purify plasmids. Genomic DNA was obtained by phenol-chloroform extraction. Digests, PCRs and DNA purification were all done according to general protocols[26]. DNA sequencing was performed by Geneservice.

Construction and characterization of mutants. The *C. difficile* single-mutant strains $A^{-}B^{+}$ and $A^{+}B^{-}$ were made using ClosTron technology as described previously[5,6]. The $A^{-}B^{-}$ double-mutant strain was made from the $A^{+}B^{-}$ mutant using a *catP*-based ClosTron using the 'pseudo-suicide' vector principle as described elsewhere[25]. The following retargeted plasmids, pMTL007C-E2::Cdi-*tcdA*-1584s, pMTL007C-E2::Cdi-*tcdB*-1511a and, for the double mutant, pMTL007S-C7::Cdi-*tcdA*-1584s were used. To verify the correct insertions, the primers used for *tcdA* were: Cdi-tcdA-F2 (5′-TCAATTGACAGAACAAGAAAT AAATAGTCTATGGAGC-3′) and EBS universal[5], and Cdi-tcdA-R2 (5′-TAC CCCATTGTCTTCAGAAAGAGATCCACC-3′) and ErmRAM-R (5′-ACGCG TGCGACTCATAGAATTATTTCCTCCCG-3′); and for *tcdB* they were: Cdi-tcdB-F1 (5′-TGATAGTATAATGGCTGAAGCTAATGCAGATAATGG-3′) and ErmRAM-R, and Cdi-tcdB-R1 (5′-CTTGCATCGTCAAATGACCATAA GCTAGCC-3′) and EBS universal.

Southern blotting. Mutants were verified by Southern blot using an intron-specific probe. Two micrograms of genomic DNA were digested with EcoRV (NEB) overnight. The blot was carried out using a DIG high prime labelling and detection kit (Roche) according to the manufacturer's instructions.

Western blotting. Supernatants from 96-h cultures, grown anaerobically in TY, were concentrated eightfold by chloroform-methanol precipitation. Proteins were standardized and run on Tricine gels 10–20% (Invitrogen) and transferred onto nitrocellulose membrane. The membranes were blocked with milk powder and then incubated with mouse monoclonal anti-TcdA antibody TTC8 and mouse monoclonal anti-TcdB antibody 2CV (tgcBIOMICS) respectively, followed by protein A conjugated with horseradish peroxidase (protein A–HRP) (Sigma). The ECL western blot detection kit from Amersham was used according to the manufacturer's instructions.

Cell toxicity assays. The four strains were grown overnight in 5 ml TY under anaerobic conditions as previously described[22], then the cell density was standardized, the cells centrifuged and supernatants filtered. Supernatants were diluted in a fourfold series and 20 μl of dilutions were added onto monolayers of Vero and HT29 cells preincubated in 96-well plates for 48 h (at 37 °C, 5% CO_2). Cytotoxicity was recorded after 24 h. For the neutralization assay, appropriate dilutions of supernatants were pre-incubated with a suitable concentration of anti-TcdA or anti-TcdB serum (polyclonal, tgcBIOMICS) for 1 h at 37 °C. These were then added to Vero and HT29 cells, which were evaluated after 24 h.

Vero and HT29 cells were grown in DMEM or McCoy's 5A, respectively, with 10% v/v fetal calf serum and 1% v/v penicillin-streptomycin at 37 °C, 5% CO_2 until confluent. Cells were detached using trypsin and seeded into 96-well plates at a density of approximately 2×10^5 cells ml^{-1}. All assays were carried out in triplicate. GraphPad Prism was used for statistical analysis. Significant differences were assessed using one-way ANOVA tests.

qRT–PCR. The qRT–PCR was carried out as described previously[4].

Hamster infection model. We used a block design with final group sizes of eight animals. Female Golden Syrian hamsters (100–130 g) were housed singly in individually ventilated cages. Each hamster was dosed with clindamycin (30 mg kg^{-1}) 5 days before being infected orally with 100 spores each. Hamsters were monitored for signs of infection and killed when the end point was met. The hamsters were handled individually in a microbiological safety cabinet. In line with UK Home Office requirements to reduce animal suffering, an alternative to death was used as the end point. Animals were monitored 3–4 times per day following infection and were assessed for several parameters including presence and severity of diarrhoea, weight loss, level of activity, starey coat, sunken eyes, hunched posture and response to stimulus. A scoring system based on severity of changes observed (ranging from 0–3 for each parameter) was used to quantify changes in the condition of the animals, which were euthanized when a pre-determined cumulative value was reached.

Faecal pellets were collected daily and plated to determine the presence of *C. difficile*. Caecum samples from each hamster were homogenized, plated and *C. difficile* counts obtained. PCR was performed to determine the genotype of each strain recovered from hamsters. Faecal and caecum samples were plated on fructose agar (*C. difficile* agar base, Oxoid) with cycloserine cefoxitin, taurocholate, tetracycline and amphotericin to select for *C. difficile*. The following primer sets were used to authenticate the various strain genotypes: oligonucleotides 3800 and 10050 (ref. 8) to confirm the cells were derived from *C. difficile* 630Δ*erm*, oligonucleotide primers Cdi-tcdA-F2 (5′-TCAATTGACAGAACAAGAAATAAAT AGTCTATGGAGC-3′) and Cdi-tcdA-R2 (5′-TACCCCATTGTCTTCAGAA AGAGATCCACC-3′) to distinguish between the toxin A insertional mutants and wild type, and the primers Cdi-tcdB-F1 (5′-TGATAGTATAATGGCTGAAG CTAATGCAGATAATGG-3′) and Cdi-tcdB-R1 (5′-CTTGCATCGTCAAA TGACCATAAGCTAGCC-3′) to distinguish between the toxin B insertional mutation and wild type. A figure showing annealing of the primers and gel pictures can be found in Supplementary Information.

23. Williams, D. R., Young, D. I. & Young, M. Conjugative plasmid transfer from *Escherichia coli* to *Clostridium acetobutylicum*. *J. Gen. Microbiol.* **136,** 819–826 (1990).
24. Sorg, J. A. & Sonenshein, A. L. Bile salts and glycine as cogerminants for *Clostridium difficile* spores. *J. Bacteriol.* **190,** 2505–2512 (2008).
25. Cartman, S. T. & Minton, N. P. A *mariner*-based transposon system for *in vivo* random mutagenesis of *Clostridium difficile*. *Appl. Environ. Microbiol.* **76,** 1103–1109 (2010).
26. Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 2001).

LETTER

doi:10.1038/nature09445

Asterless is a scaffold for the onset of centriole assembly

Nikola S. Dzhindzhev[1]*, Quan D. Yu[1]*, Kipp Weiskopf[1]†, George Tzolovsky[1], Ines Cunha-Ferreira[1,2], Maria Riparbelli[3], Ana Rodrigues-Martins[1,2], Monica Bettencourt-Dias[2], Giuliano Callaini[3] & David M. Glover[1]

Centrioles are found in the centrosome core and, as basal bodies, at the base of cilia and flagella. Centriole assembly and duplication is controlled by Polo-like-kinase 4 (Plk4): these processes fail if Plk4 is downregulated and are promoted by Plk4 overexpression[1–7]. Here we show that the centriolar protein Asterless (Asl; human orthologue CEP152) provides a conserved molecular platform, the amino terminus of which interacts with the cryptic Polo box of Plk4 whereas the carboxy terminus interacts with the centriolar protein Sas-4 (CPAP in humans). *Drosophila* Asl and human CEP152 are required for the centrosomal loading of Plk4 in *Drosophila* and CPAP in human cells, respectively. Depletion of Asl or CEP152 caused failure of centrosome duplication; their overexpression led to *de novo* centriole formation in *Drosophila* eggs, duplication of free centrosomes in *Drosophila* embryos, and centrosome amplification in cultured *Drosophila* and human cells. Overexpression of a Plk4-binding-deficient mutant of Asl prevented centriole duplication in cultured cells and embryos. However, this mutant protein was able to promote microtubule organizing centre (MTOC) formation in both embryos and oocytes. Such MTOCs had pericentriolar material and the centriolar protein Sas-4, but no centrioles at their core. Formation of such acentriolar MTOCs could be phenocopied by overexpression of Sas-4 in oocytes or embryos. Our findings identify independent functions for Asl as a scaffold for Plk4 and Sas-4 that facilitates self-assembly and duplication of the centriole and organization of pericentriolar material.

To characterize how Plk4 controls centriole duplication we began by identifying its interacting proteins. We first generated a stable *Drosophila* cell line expressing Plk4 fused to Protein A (PrA–Plk4)[2]. Following the affinity purification of PrA–Plk4 we identified co-purifying proteins by mass spectrometry that included components of the SCF ubiquitin protein ligase complex and several centriole/centrosome components of which Asl had the highest Mascott score (Supplementary Fig. 2a). We and others previously showed that the stability of Plk4, and thereby the extent of centriole duplication and centrosome numbers, are under the control of SCF[2,5]. The Asl–Plk4 interaction was of particular interest because Asl was recently found to be a constitutive centriole component important for stabilizing the pericentriolar material (PCM) at the centrosome and for centrosome duplication[8–10]. We first confirmed that Asl and Plk4 could associate by transiently transfecting cells stably expressing either PrA–Plk4 or PrA alone with Asl tagged on either its N or C terminus with GFP (GFP–Asl or Asl–GFP). We pulled down PrA and its associated proteins on immunoglobulin beads and screened for GFP by immunoblotting. This confirmed that PrA–Plk4 could interact with both GFP fusions of Asl but not with GFP alone or the GFP-tagged centromeric protein CID (Supplementary Fig. 2b). To test whether Asl and Plk4 interact directly, we produced ^{35}S-labelled Asl protein by coupled transcription–translation and asked whether it could form a complex with Plk4 produced in bacterial cells. The labelled Asl protein specifically associated with Plk4–GST but not with GST alone, indicative of a direct interaction (Supplementary Fig. 2c). We then expressed fragments of Asl by coupled transcription–translation and used these to narrow the region that could bind Plk4. In this way we mapped the interaction sites to the N-terminal 300 amino acids of Asl (Fig. 1a and Supplementary Fig. 3) and in analogous reciprocal experiments to residues 376–525 of Plk4 (Fig. 1a and Supplementary Fig. 4). We further demonstrated that bacterially expressed residues 376–525 of Plk4 fused to MBP, but not MBP alone, could efficiently bind the N-terminal 400 residues of Asl immobilized on beads (Supplementary Fig. 2d), confirming that the interaction is direct. Thus, Asl interacts with Plk4 within two complete coiled-coil regions from the six predicted within its length. The Plk4 interacting segment encompasses its cryptic Polo box, the motif present in Plk4 orthologues but not other members of the Plk family.

To establish the functional significance of this interaction, we generated cells expressing a Asl–GFP fusion from the constitutive *Act5C* promoter in which GFP-tagged Asl localized to centrioles both in interphase and mitosis (Supplementary Fig. 5a). This Asl–GFP fusion protein was able to rescue Asl depletion by UTR-directed RNA interference (RNAi) (Supplementary Fig. 6a). Notably, cells constitutively overexpressing Asl–GFP (Supplementary Fig. 5c) had supernumerary centrosomes of normal ultrastructure (Fig. 1b and Supplementary Fig. 5b). We also examined centrosome numbers after induction of Asl from the metallothionein promoter in cells arrested in S phase. Control *Drosophila* cells show no increase in centrosome number under such conditions because of a block to reduplication. Notably, induction of Asl expression in S-phase-arrested cells led to an increase in centrosome number (Supplementary Table 1) to a similar extent to downregulation of Slimb F-box protein known to stabilize Plk4 (ref. 5). This suggested that overexpression of Asl can phenocopy Plk4 overexpression in driving centrosome amplification.

We then depleted Asl from cultured *Drosophila* cells by RNAi and asked how this might affect sub-cellular localization of Plk4 and centrosome number. Consecutive treatments with *asl* double-stranded RNA (dsRNA) led to a marked loss of centrosomes comparable in extent to Plk4 depletion (Fig. 1c) but at a lower rate (Supplementary Fig. 5d), consistent with recent reports demonstrating a requirement for Asl for centriole duplication in *Drosophila*[9,10]. We then performed 3–6-day depletions of Plk4 and Asl in cell lines expressing tagged Asl and Plk4, respectively. After such *Plk4* RNAi, all residual single-centrosome-containing cells displayed a robust Asl–GFP signal (data not shown). In contrast, depletion of Asl to 27.5% of wild-type levels resulted in a marked effect of Plk4 delocalization from centrosomes (Fig. 1d) without affecting Myc–Plk4 stability (Fig. 1d). In contrast, depletion of Asl to similar levels did not affect the association of other centriolar proteins—Sas-4, Bld10 or CP110 (Supplementary Fig. 7)—and gave a slight reduction of γ-tubulin levels (Supplementary Fig. 8).

[1]Cancer Research UK Cell Cycle Genetics Group, University of Cambridge, Department of Genetics, Downing Street, Cambridge CB2 3EH, UK. [2]Instituto Gulbenkian de Ciencia, Oeiras P-2780-156, Portugal. [3]Dipartimento di Biologia Evolutiva, Università di Siena, via Aldo Moro, 2-53100 Siena, Italy. †Present Address: Stanford University School of Medicine, Stanford, California 94305, USA.
*These authors contributed equally to this work.

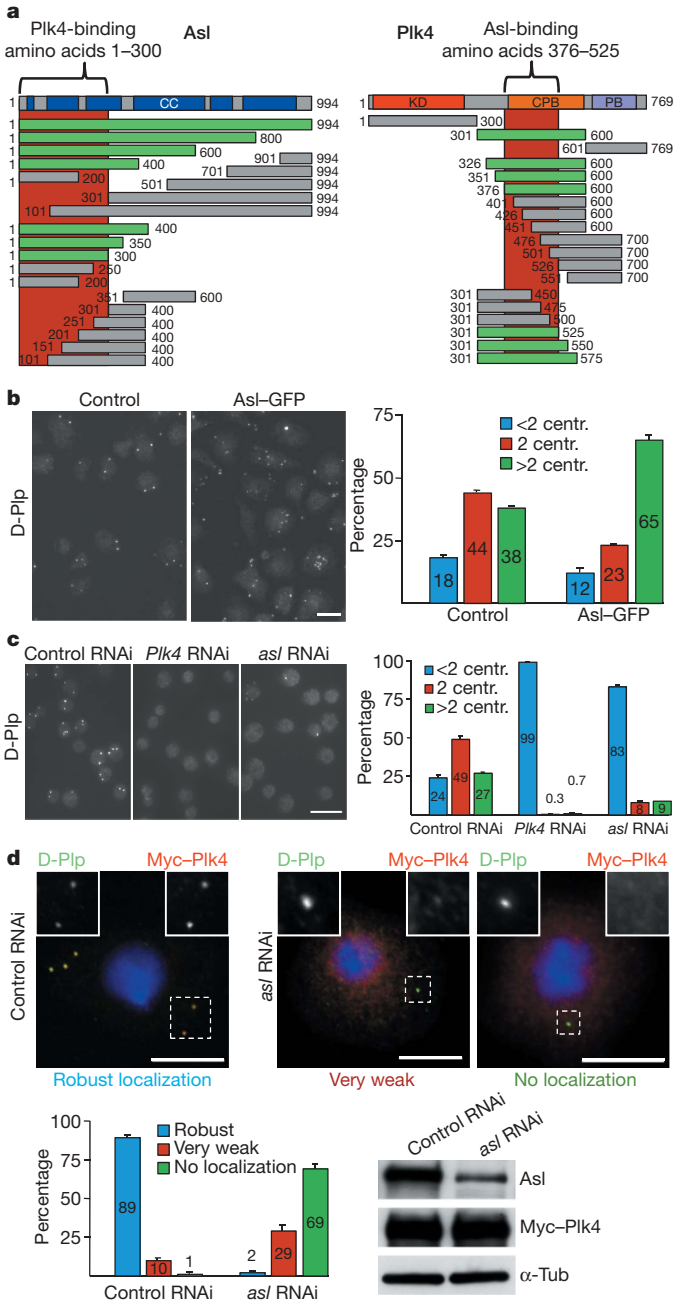


Figure 1 | Asl recruits Plk4 to centrosomes for centriole duplication. **a**, Schematic of interaction sites between Asl and Plk4. (See also Supplementary Figs 2–4.) Fragments able to interact are green; minimal identified binding regions, red. CC, coiled-coil; CPB, cryptic Polo box; KD, kinase domain; PB, polo box. **b**, Asl overexpression leads to centrosome amplification. Left: control cells and cells stably expressing Asl–GFP stained with anti-D-Plp antibody. Right: percentage of cells with <2, 2 and >2 centrosomes. Means with s.d. bars for two experiments (n = 300–500) are shown. (See also Supplementary Fig. 5a–c.) **c**, Depletion of Asl leads to centrosome loss. Left: D.Mel-2 cells treated with control, *Plk4* and *asl* dsRNA and stained with anti-D-Plp. Right: cells scored as in **b**. (See also Supplementary Fig. 5d.) **d**, Asl recruits Plk4 to centrioles. Top: Myc–Plk4 (red) is delocalized from centrosomes (D-Plp, green) after 4-day depletion of Asl. Bottom left: percentage of cells with robust, very weak or absent Plk4 on centrosomes after control or *asl* RNAi. Mean values of three experiments with s.d. bars ($n \geq 200$ cells) are shown. Bottom right: immunoblot showing that Myc–Plk4 protein stability is not affected by Asl depletion. Scale bars, 10 μm.

These results indicate a specific role of Asl in recruiting/stabilizing Plk4 at *Drosophila* centrosomes, thus ensuring their correct duplication.

We previously showed that Plk4 overexpression leads to autonomous centrosome amplification at the expense of spindle formation in *Drosophila* embryos and *de novo* centriole formation in unfertilized eggs[4]. We therefore asked whether Asl overexpression would have similar consequences by triggering expression of an inducible *asl* transgene in the female germ line. In contrast to the geometric increase in spindles in wild-type embryos, Asl overexpression led to reduced numbers of spindles and increasing numbers of free centrosomes with γ-tubulin at their core (Fig. 2a). As was also seen after Sas-6 overexpression in embryos[11], the few spindles that developed often had broad poles and multiple centrosomes. The proportion of embryos with multiple (more than 50) spindles initially increased but then diminished, whereas free centrosomes continued to increase (Fig. 2b and Supplementary Fig. 9b). Supernumerary free microtubule asters also formed in unfertilized Asl-overexpressing eggs, but their appearance and duplication was delayed relative to embryos (Fig. 2b and

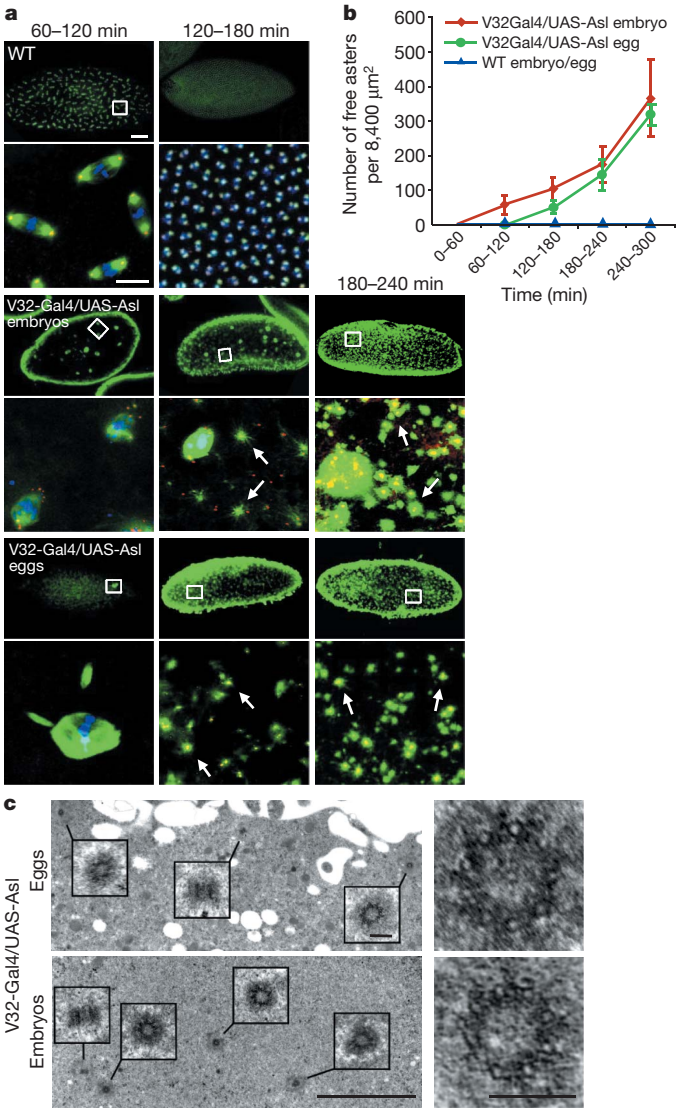


Figure 2 | Asl overexpression causes *de novo* formation and amplification of centrioles. **a**, Wild-type (WT) embryos (top) show a geometric increase in mitotic figures (α-tubulin, green; DNA, blue; γ-tubulin, red). Embryos (middle) or unfertilized eggs (bottom) overexpressing Asl show restricted spindles and increasing numbers of free centrosomes (arrows in magnified images). Scale bars, 50 μm; 10 μm in magnified images. **b**, Numbers of free asters in a surface area of 8,400 μm^2 in wild-type and Asl overexpressing eggs and embryos. Values are means of 15 eggs/embryos scored ± s.d. (See also Supplementary Fig. 9.) **c**, Free centrioles from Asl overexpressing eggs and embryos have normal ultrastructure by EM. Scale bar, 2.5 μm; in the insets, 0.15 μm; in magnified images, 0.1 μm. For eggs, 69 centrioles were analysed (23 in cross-section); for embryos, 93 centrioles were analysed (36 in cross-section).

Supplementary Fig. 9a). Such free asters are never observed in wild-type eggs or embryos. The kinetics of centrosome production were slower than in embryos or eggs overexpressing Plk4 but similar final numbers were attained. Centrioles in Asl-overexpressing eggs or embryos had normal morphology (Fig. 2c). Thus, like Plk4, Asl overexpression can drive both *de novo* formation and duplication of centrioles.

To determine whether Asl is also required for *de novo* centriole formation in cultured cells, we first carried out 15 days (5 rounds) of *Plk4* RNAi treatment so that virtually all cells have lost their centrosomes[4] (Supplementary Fig. 10). Subsequent overexpression of GFP–Plk4 (but not GFP alone) generated multiple centrioles in many cells within 24–48 h. This Plk4-induced centriole biogenesis was entirely suppressed when Asl was pre-depleted. Thus, although Asl seems to function upstream of Plk4 in centriole biogenesis, the consequences of its depletion cannot be rescued by increasing the amount of the kinase. This suggests that Asl may provide a platform for centriole *de novo* formation, onto which Plk4 and possibly its substrates are loaded to trigger downstream events.

To understand the function of the interaction between Asl and Plk4, we designed an Asl mutation to render it incapable of binding the kinase. This involved deleting residues 24–55 to generate a variant protein, Asl-M1, that was not pulled down with Plk4–GST (Supplementary Fig. 11). Loss of Plk4 binding was further confirmed *in vitro* (Supplementary Fig. 12a) and *in vivo* (Supplementary Fig. 12b). Transiently expressed Asl-M1–GFP could associate with centrosomes in the presence or absence of endogenous Asl (Supplementary Fig. 12c), indicating that the mutation did not affect its normal localization. However, in contrast to the robust centriolar localization of Myc–Plk4 when expressed alone or together with exogenous wild-type Asl–GFP, Myc–Plk4 showed little or no centriolar localization when co-expressed with Asl-M1–GFP (Fig. 3a). This was enhanced upon depletion of endogenous Asl. Accordingly, stable expression of Asl-M1–GFP from the *Act5C* promoter led to significant centrosome loss (Fig. 3b). Thus, expression of an Asl variant no longer able to bind Plk4 has a dominant-negative effect upon centrosome duplication in cultured *Drosophila* cells. We were therefore surprised that overexpression of Asl-M1 in embryos and unfertilized eggs led to multiple free MTOCs (Fig. 3c). However, these MTOCs were irregular in shape, size and spatial distribution, frequently showing 'strings' of γ-tubulin (arrows in Fig. 3c). Electron microscopy revealed that these MTOCs did not have centrioles at their cores (Fig. 3c) but instead amorphous, electron-dense material reminiscent of PCM (Supplementary Fig. 12d).

The above observations suggested that although it was unable to bind Plk4, Asl-M1 was still able to bind other centriolar/pericentriolar components. We therefore expressed and purified PrA-tagged Asl with its associated proteins and, in addition to Plk4, identified Sas-4 as a member of the resulting complex. Using a similar strategy to the one we used to examine Asl–Plk4 interactions (Supplementary Fig. 13), we found that residues 648–995 of Asl could interact directly with residues 101–650 of Sas-4 (Fig. 3d). We then found that Sas-4 was associated with the MTOCs resulting from overexpression of Asl-M1 (Supplementary Fig. 14). It has already been shown that Sas-4 overexpression in unfertilized eggs leads to multiple MTOCs, but their ultrastructure was never examined[7]. Electron microscopy revealed that such supernumerary MTOCs were similar to those formed upon Asl-M1 overexpression and had no centrioles at their core (Fig. 3e). Thus, either overexpression of Sas-4 or an Asl variant able to bind Sas-4 but not Plk4 both promote assembly of acentriolar MTOCs.

Finally, we examined the interactions of the human counterpart of Asl, CEP152, and found the same partners. The N-terminal part of CEP152 could bind the cryptic Polo box of human PLK4 and its C-terminal part could bind to the N-terminal part of CPAP, orthologue of Sas-4 (Fig. 4a and Supplementary Fig. 15). We tested the functional requirements for CEP152 in U2OS cells where centrioles continue to replicate after treatment with the DNA polymerase α-inhibitor

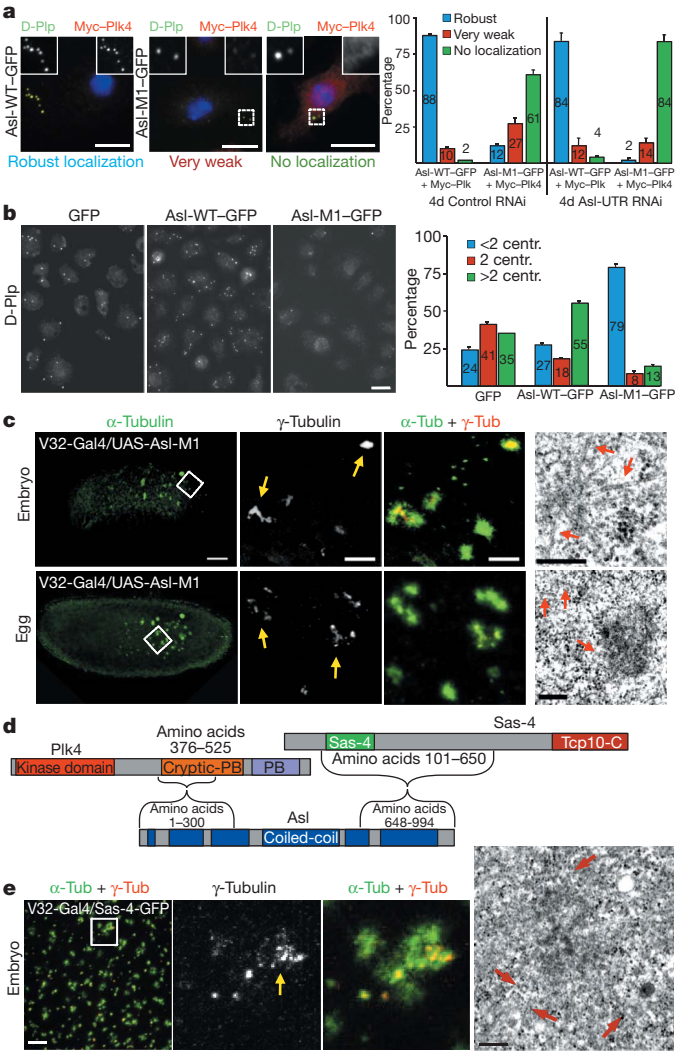


Figure 3 | Overexpression of Asl-M1 mutant or the Asl interactor Sas-4 leads to loss of centrioles. **a**, Asl-M1 (see Supplementary Figs 11 and 12) prevents recruitment of Plk4 to centrioles. Left: cells stably expressing both Asl-WT–GFP and Myc–Plk4 or Asl-M1–GFP and Myc–Plk4 stained to reveal Myc (red) and D-Plp (green). Scale bar, 10 μm. Right: cells scored as in Fig. 1d. Means of two experiments with s.d. bars ($n \geq 200$ cells) are shown. **b**, Expression of Asl-M1 leads to centrosome loss. Left: cells from GFP, Asl-WT–GFP or Asl-M1–GFP cell lines stained for D-Plp. Scale bar, 10 μm. Right: cells scored as in Fig. 1c. Means of two experiments with s.d. bars ($n \geq 200$ cells) are shown. **c**, Asl-M1 overexpression in embryos and eggs induces formation of centriole-free irregular microtubule asters with abnormal γ-tubulin aggregates (arrows), left. Scale bar, 50 μm; high magnification, 10 μm. Right: EM sections of asters from embryo (top) and egg (bottom) expressing Asl-M1 showing nucleation of microtubules (arrows) from amorphous, PCM-like masses. Scale bars, 1 μm. **d**, Schematic of interacting regions in Asl and Sas-4 (see Supplementary Fig. 13). **e**, Left: Sas-4 overexpression in embryos and eggs (latter not shown) causes formation of supernumerary acentriolar disorganized microtubule asters (green) with γ-tubulin in their centres (red, arrow in monochrome). Scale bar, 10 μm. Right: electron micrograph of such an acentriolar aster (microtubules indicated by arrows). Scale bar, 1 μm.

aphidicolin (Fig. 4b). After knockdown of either CEP152 or PLK4 by RNAi, the number of cells accumulating supernumerary centrosomes in an S-phase block was less than half that of control cells (Fig. 4b). Conversely, when we overexpressed either CEP152 or PLK4, this led to an increased number of centrosomes per cell (Fig. 4c). After *CEP152* RNAi treatment, the great majority of cells showed markedly reduced or absent CEP152 and CPAP at centrosomes (Fig. 4d). Notably, centrosomes persisted and contained γ-tubulin and centrin 3, even in the apparent absence of CEP152 and CPAP. There was little

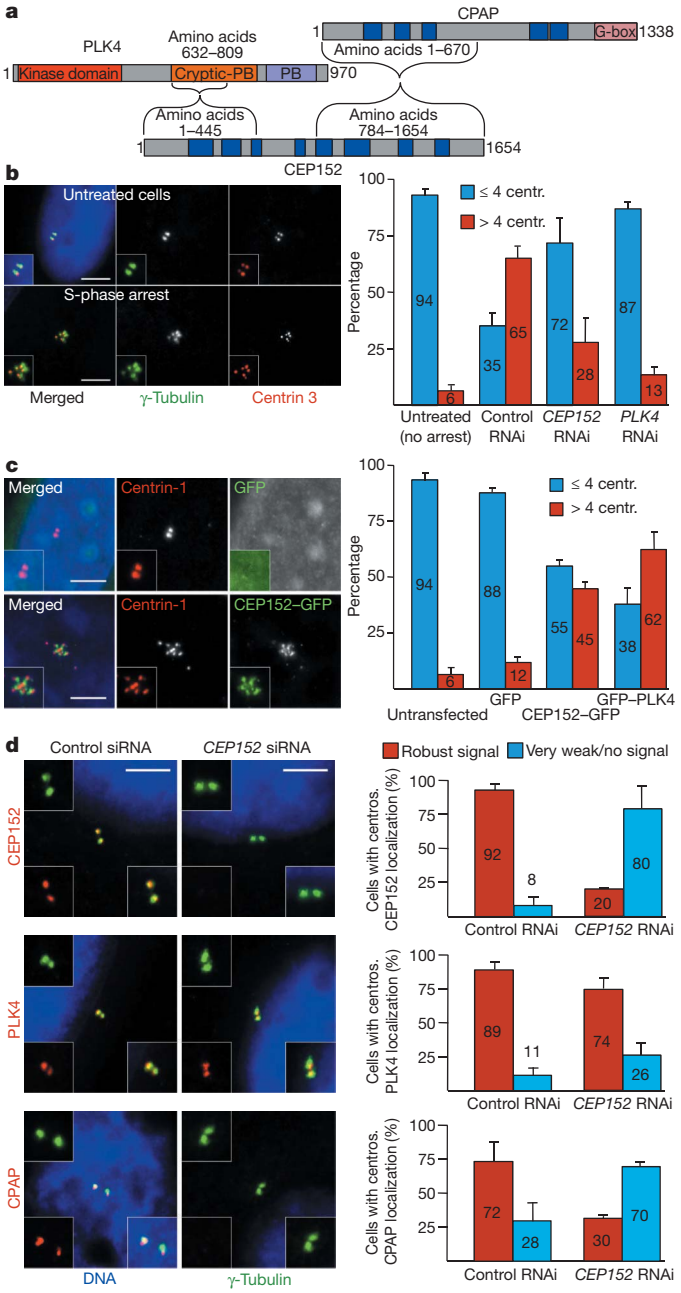


Figure 4 | CEP152 interacts with human PLK4 and CPAP and controls centrosome duplication in human cells. a, Schematic of interactions between CEP152, human PLK4 and CPAP (see Supplementary Fig. 15). **b**, CEP152 depletion blocks centriole reduplication in S-phase-arrested U2OS cells. Left: untreated and aphidicolin-treated cells stained for γ-tubulin (green) and centrin 3 (red). Right: cells scored as having either ≤4 centrioles or >4 centrioles per cell. Values are means of two experiments with s.d. bars ($n > 200$ cells). **c**, Transient overexpression of CEP152 in U2OS cells leads to centrosome amplification. Left: DNA, blue; GFP, green; centrin 1, red. Right: cells scored as in **b**. **d**, CEP152 depletion significantly decreases CPAP localization at centrioles, but has no effect on PLK4. Left: representative images of control and CEP152-depleted cells. DNA, blue; γ-tubulin, green; and either CEP152, PLK4, or CPAP, red. Right: in each case cells scored as having either robust signal at centrosomes, red; or strongly reduced/absent signal, blue. Values are means of two experiments with s.d. bars ($n > 200$ cells). All scale bars, 5 μm.

effect, however, on the centrosomal levels of PLK4 (Fig. 4d and Supplementary Fig. 16). Thus, although the physical interactions of CEP152/Asl with these two partners have been strongly conserved, the recruitment dependencies of Plk4 and Sas-4/CPAP differ between human and *Drosophila* cells.

Our study provides an example of an interaction with Plk4's cryptic Polo box. The three-dimensional structure of the Polo box (PB) domain of Plk4 (ref. 12) indicated that its mode of binding to partners was likely to differ from that of Polo/Plk1. The double PB of the latter clamps onto its targets, the majority of which first undergo priming phosphorylation[13–16]. The interaction of Plk4's single cryptic PB with Asl seems unlikely to require a clamping mechanism and our data indicate that it occurs independently of phosphorylation, although it remains possible that phosphorylation might modulate the interaction. In *Drosophila* cells, Plk4's interaction with Asl is responsible for its recruitment to the centrosome. In human cells, PLK4 still localizes to centrosomes in the absence of CEP152, indicating that in these cells it has alternative mechanisms of associating with either centrioles or PCM. Although the recruitment dependencies may have diverged, we suggest that Asl/CEP152 is a molecular scaffold facilitating association of two molecules that regulate centriole formation: Plk4 and Sas-4 (Supplementary Fig. 1). Such an association might be required for Plk4 to phosphorylate Sas-4/CPAP as was recently reported to occur[17]. Although it seems likely that there are other Plk4 substrates in the centriole or PCM, their identity is unclear. We have not been successful in demonstrating robust phosphorylation of *Drosophila* Asl by Plk4 *in vitro*, although we cannot rule out the possibility that Asl is, under certain conditions, a Plk4 substrate. Our study focuses on *Drosophila* cells where overexpression of Plk4 is insufficient for *de novo* centriole formation in the absence of Asl and where expression of an Asl variant unable to bind Plk4 prevents formation of centrioles. Although unable to bind Plk4, Asl-M1 can still bind Sas-4, and can drive formation of multiple MTOCs in the absence of centrioles. This phenocopies the consequences of Sas-4 overexpression (Supplementary Fig. 1). Thus, either directly or indirectly, Asl and Sas-4 seem to be able to recruit pericentriolar material components and so nucleate cytoplasmic microtubules, consistent with recent reports[8,10]. Our results demonstrate that Asl is a scaffold protein with mechanistically separable roles in both centriole formation and PCM recruitment.

METHODS SUMMARY

pUASp-Asl and pUASp-Asl-M1 transgenic flies were generated by BestGene Inc. Expression in the female germ line was driven by V32-gal4 or poly-ubiquitin (pUb) promoters. All cloning used Gateway technology (Invitrogen). *Drosophila* D.Mel-2 cells (Invitrogen) were transfected using Cellfectin (Invitrogen) and stable cell lines were generated by Blasticidin selection (20 μg ml^{-1}). For RNAi, 1×10^6 cells were transfected with 20 μg of dsRNA using Transfast (Promega). For mutagenesis the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene) was used. S-phase arrest of D.Mel-2 cells was as described[8]. Asl antibody was raised in rabbits against entire Asl protein. Other antibodies used are as follows: YL1/2 (Oxford Bioscences); GTU88 (Sigma); AK-15 (Sigma); centrin 3 (Santa Cruz); centrin 1 (Sigma, C7736); CEP152 (Bethyl Laboratories, A302-480A); rat anti-mouse-Plk4 (P. Coelho, unpublished reagent); rabbit anti-CPAP (from Pierre Gönczy); D-PLP[7]; anti-Myc (clone 9E10, Abcam); DM1A (Sigma); anti-GFP (Roche, 11814460001); DSas-4 (from J. Raff); Bld10 (from T. Megraw); rabbit anti-CP110 (G. Mao, unpublished reagent). Secondary antibodies were conjugated with Rhodamine Redex, FITC (Sigma), Alexa Fluor 488 and 594 (Invitrogen), CY5 and peroxidase (Jackson Immunochemicals). Protein A affinity purification was as described[18]. *In vitro* binding assays were carried out with ^{35}S-methionine-labelled proteins made from plasmid or PCR-generated DNA templates using the TNT T7 Quick Coupled Transcription/Translation System (Promega). These were applied equally to MBP/GST-tagged bait and MBP/GST alone as negative controls. U2OS cells were transfected using Oligofectamine (Invitrogen) for siRNA and either FuGENE HD (Roche) or Lipofectamine 2000 (Invitrogen) for plasmid DNA.

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

Received 23 December 2009; accepted 27 August 2010.
Published online 19 September 2010.

1. Kleylein-Sohn, J. *et al.* Plk4-induced centriole biogenesis in human cells. *Dev. Cell* **13,** 190–202 (2007).
2. Cunha-Ferreira, I. *et al.* The SCF/Slimb ubiquitin ligase limits centrosome amplification through degradation of SAK/PLK4. *Curr. Biol.* **19,** 43–49 (2009).

3. Bettencourt-Dias, M. *et al.* SAK/PLK4 is required for centriole duplication and flagella development. *Curr. Biol.* **15,** 2199–2207 (2005).
4. Rodrigues-Martins, A., Riparbelli, M., Callaini, G., Glover, D. M. & Bettencourt-Dias, M. Revisiting the role of the mother centriole in centriole biogenesis. *Science* **316,** 1046–1050 (2007).
5. Rogers, G. C., Rusan, N. M., Roberts, D. M., Peifer, M. & Rogers, S. L. The SCF Slimb ubiquitin ligase regulates Plk4/Sak levels to block centriole reduplication. *J. Cell Biol.* **184,** 225–239 (2009).
6. Habedanck, R., Stierhof, Y. D., Wilkinson, C. J. & Nigg, E. A. The Polo kinase Plk4 functions in centriole duplication. *Nature Cell Biol.* **7,** 1140–1146 (2005).
7. Peel, N., Stevens, N. R., Basto, R. & Raff, J. W. Overexpressing centriole-replication proteins *in vivo* induces centriole overduplication and *de novo* formation. *Curr. Biol.* **17,** 834–843 (2007).
8. Varmark, H. *et al.* Asterless is a centriolar protein required for centrosome function and embryo development in *Drosophila. Curr. Biol.* **17,** 1735–1745 (2007).
9. Blachon, S. *et al. Drosophila* asterless and vertebrate Cep152 are orthologs essential for centriole duplication. *Genetics* **180,** 2081–2094 (2008).
10. Dobbelaere, J. *et al.* A genome-wide RNAi screen to dissect centriole duplication and centrosome maturation in *Drosophila. PLoS Biol.* **6,** e224 (2008).
11. Rodrigues-Martins, A. *et al.* DSAS-6 organizes a tube-like centriole precursor, and its absence suggests modularity in centriole assembly. *Curr. Biol.* **17,** 1465–1472 (2007).
12. Leung, G. C. *et al.* The Sak polo-box comprises a structural domain sufficient for mitotic subcellular localization. *Nature Struct. Biol.* **9,** 719–724 (2002).
13. Elia, A. E., Cantley, L. C. & Yaffe, M. B. Proteomic screen finds pSer/pThr-binding domain localizing Plk1 to mitotic substrates. *Science* **299,** 1228–1231 (2003).
14. Elia, A. E. *et al.* The molecular basis for phosphodependent substrate targeting and regulation of Plks by the Polo-box domain. *Cell* **115,** 83–95 (2003).
15. Lowery, D. M. *et al.* Proteomic screen defines the Polo-box domain interactome and identifies Rock2 as a Plk1 substrate. *EMBO J.* **26,** 2262–2273 (2007).
16. Lowery, D. M., Lim, D. & Yaffe, M. B. Structure and function of Polo-like kinases. *Oncogene* **24,** 248–259 (2005).
17. Chang, J., Cizmecioglu, O., Hoffmann, I. & Rhee, K. PLK2 phosphorylation is critical for CPAP function in procentriole formation during the centrosome cycle. *EMBO J.* **29,** 2395–2406 (2010).
18. D'Avino, P. P. *et al.* Isolation of protein complexes involved in mitosis and cytokinesis from *Drosophila* cultured cells. *Methods Mol. Biol.* **545,** 99–112 (2009).

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

Acknowledgements We thank E. Wegel for technical help. We thank P. Coelho, G. Mao, P. Gönczy, T. Megraw and J. Raff for antibodies. We thank Cancer Research UK for a Programme Grant to D.M.G. and a studentship to Q.D.Y. K.W. was a visiting scholar of the Winston Churchill Foundation of the United States. Fundação Calouste Gulbenkian, Fundação para a Ciência e Tecnologia (FCT) provided support to M.B.-D. and scholarships to I.C.-F. and A.R.-M. The Royal Society provided an International Joint Project Grant for collaboration between M.B.-D. and D.M.G.

Author Contributions N.S.D. undertook interaction assays, mutagenesis and *Drosophila* cell culture work; Q.D.Y. worked on *Drosophila* embryos/eggs and *de novo* centriole formation; K.W. and I.C.-F. performed PrA–Plk4/Asl purifications; K.W. studied Asl depletion/overexpression in *Drosophila* cell culture. G.T. performed the human cell culture experiments. A.R.-M. and M.B.-D. overexpressed Sas-4 in embryos/eggs. M.R. and G.C. performed EM. N.S.D. and D.M.G. planned experiments and wrote the paper that was discussed by all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.S.D. (nsd23@cam.ac.uk) or D.M.G. (dmg25@cam.ac.uk).

METHODS

Flies and husbandry. pUASp-Asl and pUASp-Asl-M1 transgenic flies were generated by BestGene Inc. Expression in the female germ line was driven by V32-gal4 or poly-ubiquitin promoters. OreR stocks were used as wild-type flies. pUASp-Asl and pUASp-AslM1 embryos were fertilized by OreR males. All flies were reared according to standard procedures and maintained at 25 °C.

Constructs. *Drosophila* expression vectors were made using the Gateway technology (Invitrogen). Full-length Asl (clone GH02902) and Plk4 (clone RE70136), and fragments of Asl and Plk4, were cloned by recombination of PCR products into the pDONR221 vector (Invitrogen). The same sequences were then recombined into the relevant destination vectors for expression in fusion with protein A, GFP, or Myc, at the N or C terminus, and under the control of the Actin 5C, UASp, or MT (Metallothionein) promoter. Bacterial expression constructs for GST and MBP fusions of Asl and Plk4 were made using the following destination vectors: pDEST24 (Invitrogen) and pKM596[19] (Addgene plasmid 8837).

Cell culture, transfections, cell lines and RNAi. D.Mel-2 cells (Invitrogen) were cultured in Express Five SFM medium (Invitrogen) supplemented with 1 mM glutamine, penicillin and streptomycin. Transfections and selection of stable cell lines were carried out as described[20]. For RNAi, 1×10^6 cells were transfected with 20 μg of dsRNA using Transfast (Promega).

Production of dsRNA and transfections were performed as described[3]. Templates used in PCR: GH02902 (Asl cDNA), RE70136 (Plk4 cDNA) and pDEST24 (GST-full-length-CDS-containing vector, Invitrogen).

Primers used for dsRNA production were. Sak-F, 5′-TAATACGACTCACTATAGGGAGAATACGGGAGGAATTTAAGCAAGTC-3′, Sak-R, 5′-TAATACGACTCACTATAGGGAGATTATAACGCGTCGGAAGCAGTCT-3′; GST-F, 5′-TAATACGACTCACTATAGGGAGAAAGGTGATAAATGGCGAAACAAAA-3′, GST-R, 5′-TAATACGACTCACTATAGGGAGACAACATCAAGAGCGTCATACAACA-3′; Asl-F, 5′-TAATACGACTCACTATAGGGAGAATGAACACGCCAGGTATAAG-3′, Asl-R, 5′-TAATACGACTCACTATAGGGAGATATTGGAGCACGTCTCTTT-3′.

For the RNAi-based silencing of endogenous Asl, the following 5′-UTR-3′-UTR hybrid 181-bp-long sequence was generated by Overlap Extension PCR: 5′-TTGCCTACGAAAATAGCGCCTGCGCCTCGGAATTTATTTGAATTTGATGCTTGTGGCGTCCCCTTAGCCGAATGACTTAGGAAAATATATATATGTATATGTAGGTCTCTTTAGTGGGATATCGTTACTGCATGCACTGCCAAATTGTCTAAACAAATAAAAGTCGCTGTACATTCCTAAA-3′.

Site-directed mutagenesis. For the construction of Asl-M1–M5 mutants, the following mutations in Asl cDNA (GH02902) or in the Asl-Entry clone were introduced: Asl-M1 (deleting residues 24–55), Asl-M2 (deleting residues 101–115), Asl-M3 (deleting residues 150–159), Asl-M4 (deleting residues 236–241), Asl-M5 (deleting residues 273–296). We used the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene). Primers used for Site-Directed Mutagenesis were: Asl-M1-F, 5′-AACATAAATTCTACACTGGACCGGAACACAATCGACTCGACGACGAAT-3′, Asl-M1-R, 5′-ATTCGTCGTCGAGTCGATTGTGTTCCGGTCCAGTGTAGAATTTATGTT -3′; Asl-M2-F, 5′-GCCAACGAGATCGAGATGCATCGCAACCAGGTGGCCAACGCTGCGCAC-3′, Asl-M2-R, 5′-GTGCGCAGCGTTGGCCACCTGGTTGCGATGCATCTCGATCTCGTTGGC-3′; Asl-M3-F, 5′-CGCGCCATTCGCGAGAAGCAGAACACGTGCTCCAATAAGGACAGCGAT-3′, Asl-M3-R, 5′-ATCGCTGTCCTTATTGGAGCACGTGTTCTGCTTCTCGCGAATGGCGCG-3′; Asl-M4-F, 5′-GCGCACCGAGCCCAGTCCGATCTTCAGATGAAGGATCAGCTTGATCGC-3′, Asl-M4-R, 5′-GCGATCAAGCTGATCCTTCATCTGAAGATCGGACTGGGCTCGGTGCGC-3′; Asl-M5-F, 5′-CAGTCTGGTCACGAAACAATGCTCAGTGCAAGACCCGATTTGCAGGCT-3′, Asl-M5-R, 5′-AGCCTGCAAATCGGGTCTTGCACTGAGCATTGTTTCGTGACCAGACTG-3′.

S-phase arrest experiment. For the S-phase arrest of D.Mel-2 cells we used the same conditions as described previously[5]. In brief, control (non-transfected) and cells stably expressing pMT-Asl–GFP were treated with 50 μM $CuSO_4$ to induce expression of the transgene. Centrosome numbers in both cell lines were scored every 24 h (day 0–5) in cycling cells and in S-phase arrest (induced by 10 μM aphidicolin and 1.5 mM HU, added 8 h after the induction with $CuSO_4$).

Antibodies. For generation of rabbit anti-Asl serum, three regions of Asl, that cover the entire protein length, were expressed in *Escherichia coli* as GST fusions (Asl(1–300)–GST, Asl(301–647)–GST and Asl(648–995)–GST) and affinity purified. Immunizations were performed by Scottish National Blood Transfusion Service with a mixture of the three Asl peptides. The anti-Asl antibody was used for western blotting (1:2,000).

We used the following antibodies for immunofluorescence or western blotting: rat anti-α-tubulin-YL1/2 (Oxford Bioscences, immunofluorescence 1:50); mouse anti-γ-tubulin-GTU88 (Sigma, immunofluorescence 1:25); rabbit anti-γ-tubulin, AK-15 (Sigma, immunofluorescence 1:250); mouse anti-centrin 3 (Santa Cruz, immunofluorescence 1:300); rabbit anti-centrin 1 (Sigma, C7736, immunofluorescence 1:400); rabbit anti-CEP152 (Bethyl Laboratories, A302-480A, immunofluorescence 1:2,500); rat anti-mouse Plk4 antibody (provided by P. Coelho, unpublished reagent, immunofluorescence 1:4,000); rabbit anti-CPAP antibody (provided by P. Gönczy, immunofluorescence 1:500); chicken anti-D-PLP[4] (immunofluorescence 1:1,000), mouse anti-Myc (clone 9E10 from Abcam, immunofluorescence 1:1000, western blotting 1:5000), mouse anti-α-tubulin (clone DM1A Sigma, western blotting 1:5,000); mouse anti-GFP (Roche, 11814460001, immunofluorescence 1:600, western blotting 1:500); rabbit anti-DSas-4 (immunofluorescence 1:750)[21], rabbit anti-Bld10 (immunofluorescence 1:1,000)[22] and rabbit anti-CP110 (G. Mao, unpublished reagent, immunofluorescence 1:750). Secondary antibodies used were conjugated with Rhodamine Redex, FITC (Sigma), Alexa Fluor 488 and 594 (Invitrogen), CY5 (Jackson Imunochemicals) and Peroxidase (Jackson Immunochemicals). All secondary antibodies had minimal cross reactivity to other species.

Immunostaining and microscopy. Embryos and eggs from 4–5-day-old non-virgin/virgin females were collected at 25 °C on agar plates. Agar plates were changed every 15 min for four times to synchronize embryos and eggs before collections. Embryos and eggs were collected according to different time points as indicated. Embryos and eggs were dechorionated, washed and had their vitelline membrane removed as described[4]. Embryos and eggs were fixed and stained as described[23]. Mounting was in Vectashield (Vector laboratories) and TOTO-3 iodide (Invitrogen, 1:1,000) was used for DNA counterstaining.

Embryos and eggs were all analysed with a Zeiss LSM 510 Meta Lazer Scanning Confocal Microscope with 63×/1.4 oil objective and 20×/0.5 Plan Apochromat objectives. Images were acquired with the manufacture software LSM 510 Version 4.2. Images shown are the projections of optical sections.

D.Mel-2 cells were grown on coverslips (either pre-treated with concanavalin A or without pre-treatment) for 4–6 h and fixed in cold methanol for 5 min. Blocking (10% fetal calf serum), antibody incubations and washes were done in PBS + 0.1% Triton X-100. Cells were mounted in Vectashield with DAPI (Vector laboratories). Microscopy was performed on a Carl Zeiss Axiovert 200M microscope with 63×/1.4 and 100×/1.4 Plan Apochromat objectives. Images were acquired with a Photometrics Cool SNAP HQ camera and the image analysis software MGA Biosystem Software Series 7.6.

Transmission electron microscopy. 2–3-h wild type and embryos/eggs overexpressing Asl or Asl-M1 were collected on small agar plates at 25 °C. Embryos and eggs were dechorionated in 50% bleach and washed with distilled water. Dechorionated embryos and eggs were incubated in a solution containing 1:1 ratio of 25% glutaraldehyde in PBS and heptane for 3 min. Embryos and eggs were then incubated in 2.5% glutaraldehyde in PBS for 30 min before the vitelline envelope was removed with tungsten needles. The devitellinized embryos and eggs were incubated in 2.5% glutaraldehyde overnight at 4 °C. The samples were washed in PBS and post-fixed in 1% osmium tetroxide for 2 h at 4 °C. The samples were dehydrated in graded alcohol series, embedded in an Epon-Araldite mixture and polymerized at 60 °C for 48 h. Random and serial sections were cut with an LKB Nova ultramicrotome. Sections were mounted on copper grids and stained with uranyl acetate and lead citrate. The preparations were observed and photographed with a Philips CM10 electron microscope.

Cells from the line stably expressing Act5-Asl–GFP were fixed in 2.5% glutaraldehyde in 0.1 M phosphate buffer (pH 7.2) overnight at 4 °C. Cells were washed 3× for 30 min in PBS after fixation. The cells were post-fixed with 1% OsO_4 for 1 h at 4 °C, washed first in PBS, then in distilled water. Cells were then dehydrated in a graded series of ethanol and flat-embedded in a mixture of Epon and Araldite. The samples were incubated at 60 °C for 2 days for polymerization. The samples were then briefly immersed in liquid nitrogen and removed from the resin. Sections were obtained with a LKB ultratome, stained with uranyl acetate and lead citrate, and observed and photographed with a Philips CM10 electron microscope at 80 kV.

Western blotting. Standard procedures were used for western blotting. Total protein extracts from tissue cultured cells were prepared by homogenizing cell pellets in SDS–PAGE sample buffer, boiling for 5 min and spinning at high speed to clear the lysate.

Protein A affinity purification and mass spectrometry. Protein A affinity purification was performed as described[18]. For identification of PrA–Plk4-interacting proteins, PrA–Plk4 expression was induced in the pMT-PrA–Plk4 cell line by adding $CuSO_4$ to 100 μM for 16 h. At the same time cells were treated with 10 μM MG-132 to inhibit proteasomal degradation. Approximately 1.5–2.0×10^9 cells were used to purify complexes. Cells were homogenized on ice in 10 ml cold Extraction Buffer (EB; 75 mM NaHEPES, 150 mM KCl, 2 mM $MgCl_2$, 0.1% NP-40 substitute, 5 mM DTT, 2 mM EGTA, 5% glycerol, complete protease inhibitors (Roche)). Lysates were clarified by centrifuging at 20,000*g* for 30 min in a table-top centrifuge at 4 °C. Supernatants were transferred to new tubes and incubated with 250 μl magnetic Dynabeads M-270 Epoxy (Invitrogen, catalogue number 143.01) conjugated to rabbit IgG (MP Biomedicals, catalogue number

0855272). Samples were incubated on a rotating wheel for 4 h at 4 °C to allow the complexes associated with the PrA-tagged bait to bind to the rabbit IgG. Beads were washed a total of five times with EB and then proteins were eluted in 0.5 M NH_4OH. Protein samples were prepared for mass spectrometry by desiccation using a Concentrator 5301 (Eppendorf). Dry protein pellets were submitted to the Mass Spectrometry Laboratory at the Institute of Biochemistry and Biophysics (Warsaw, Poland) for gel-free analysis.

For the PrA–Plk4–Asl–GFP co-immunoprecipitation experiment, cells (pMT-PrA–Plk4 and pMT-PrA-alone stable cell lines) growing in 25 cm^2 flasks were transfected with constructs for the expression of GFP, CID–GFP, GFP–Asl and Asl–GFP 3 days before harvesting. PrA and PrA–Plk4 expression was induced with 100 μM $CuSO_4$ 16 h before collection. MG-132 was not used in this experiment. The CID–GFP construct and PMT–GFP construct that were used as controls were provided by M. Przewloka and T. Takeda, respectively. Between 1.5–3.0×10^7 cells were used for each purification. A volume of 30 μl bead slurry of IgG-coupled Dynabeads was used to bind PrA-associated complexes. A volume of 1 ml EB was used for washes. Complexes were dissociated from the beads using Laemmli sample buffer. Samples were then analysed by western blot.

Expression and purification of GST- and MBP-tagged proteins from *E. coli*. Standard procedures were followed. MBP and GST fusion proteins were expressed in BL21 Star (DE3)pLysS *E. coli* cells (Invitrogen) by IPTG induction at 37 °C for 4 h. Bacterial cells were lysed by sonication and fusion proteins were purified onto Amylose resin (New England Biolabs) or Glutathione sepharose 4b resin (GE Healthcare) according to the manufacturers' instructions. Wherever necessary, proteins were eluted by D-maltose or reduced gluthatione according to manufacturers' protocols. Note that purified Plk4–GST (as in Supplementary Fig. 2c) is not full length. We were never able to purify any N- or C-terminal fusions of wild-type Plk4 from bacteria that were full length, owing to extensive *in vivo* degradation. Because the GST tag in this case in at the C terminus, we presume that in our Plk–GST purifications Plk4 is truncated from its N-terminal kinase domain.

Recombinant full-length Asl-WT–GST and Asl-M1–GST are insoluble in standard purification conditions, therefore we purified these under special 'solubilizing' conditions using sarkosyl as described[24].

Myc–Plk4–Asl–GFP co-immunoprecipitations. D.Mel-2 cell lines stably co-expressing Act5C-Myc–Plk4 and Act5C-Asl-WT–GFP or Act5C-Myc–Plk4 and Act5C-Asl-M1–GFP were lysed in extraction buffer (EB: 50 mM HEPES pH 7.5, 1 mM EGTA, 1 mM Mg_2Cl, 1 mM DTT, 150 mM NaCl, 0.25% NP-40, complete EDTA-free protease inhibitors (Roche)) by 3×30 s pulses with a Power Gen 125 homogenizer (Fisher Scientific) at power level 6. Approximately 1×10^7 cells were used for each co-immunoprecipitation. Lysates were cleared by centrifugation (20 min, 20,000*g* at 4° C) in a refrigerated bench-top centrifuge to obtain soluble extracts. These were then applied to Protein-G-coupled Dynabeads (Invitrogen) pre-bound with 6 μg of mouse anti-C-Myc antibody (clone 9E10 from Abcam). After incubation for 1 h at 4 °C, beads were washed 5×5 min in EB and then boiled in SDS–PAGE sample buffer and processed for western blotting.

***In vitro* binding assays.** For mapping of the interaction regions in Asl and Plk4, ^{35}S-methionine-labelled Asl and Plk4 fragments were produced by *in vitro* transcription-translation (IVT) using the TNT T7 Quick Coupled Transcription/Translation System (Promega) according to the manufacturer's instructions. In brief, either full-length cDNA-containing plasmids or PCR-generated linear fragments of Asl and Plk4 (in both cases with an upstream T7 promoter) were used in 25 μl labelling reactions. The upstream primers for the PCR fragments included T7 promoter, Kozak sequence and START codon, followed by an in-frame gene-specific sequence: 5′-GAATTAATACGACTCACTATAGGG-AGAGCCGCCA CC-ATG-in-frame gene specific sequence-3′. The downstream primers only included an in-frame STOP-codon after the gene-specific sequence.

The 25-μl reactions were divided equally and used for two pull-downs each: GST-only on beads (negative control) and either Plk4–GST or Asl(1–400)–GST beads. Binding and washes were performed in the following buffer: 50 mM HEPES pH 7.5, 1 mM EGTA, 1 mM Mg_2Cl, 1 mM DTT, 150 mM NaCl, 0.1% Triton X-100, complete EDTA-free protease inhibitors (Roche). Additionally, in the binding reactions 0.5 mg ml^{-1} BSA were added as blocking agent. After binding (0.5–1 h at ambient temperature) and washes (4×5 min at ambient temperature), the beads were boiled in SDS–PAGE sample buffer, separated on an acrylamide gel and transferred onto nitrocellulose membranes. After developing the autoradiograms, signals were compared between experimental (Plk4–GST or Asl(1–400)–GST) pull downs and GST-only control.

For direct-interaction assays with purified recombinant proteins, MBP and MBP–Plk4(376–525) and also Asl-WT–GST and Asl-M1–GST were purified and eluted from the resin as described above. Each of these proteins was then mixed with beads bound to GST and Asl(1–400)–GST (and in the case of Asl-WT/M1–GST to PrA–Plk4 on IgG-Dynabeads). Binding and washes were performed exactly as for the IVT-produced fragments. However, the analysis here was done directly on the SDS–PAGE gel by Coomassie staining (BioSafe Coomassie, Bio-Rad).

Functional studies of CEP152 in human U2OS cells. U2OS cells were grown in standard DMEM-GLUTAMAX media containing 10% FCS (Fetal calf serum), 1% penicillin-streptomycin and 1% NEAA (non-essential amino acids). Cells were transfected according to the manufacturers' recommendations using Oligofectamine (Invitrogen) for siRNA transfections and either FuGENE HD (Roche) or Lipofectamine 2000 (Invitrogen) for plasmid DNA transfections. The negative control, RSC (random sequence control)-siRNA-oligonucleotide had the following sequence: GCUAUGUGACGUAGAGCGA. Both PLK4 and CEP152 were depleted by using a pool of three siRNA oligonucleotides. For CEP152: A, ACCACAUCAUCCUGAGCAA; B, GAGCAGAAGUGCACAA UCC; C, GAAGUCCAGAACAACUGAA. For Plk4: A, AAUGCCUGGA CUGAUACAA; B, UGAAGGACUUGGUCUUACA; C, GGACCUUAUU CACCAGUUA. U2OS cells transfected with *Plk4* siRNA were analysed 48 h after transfection, whereas for *CEP152* siRNA U2OS cells were analysed 72–96 h after transfection. For immunostaining, U2OS cells were grown on coverslips and fixed with cold methanol (−20 °C). Immunostaining and microscopy were carried out as described above for *Drosophila* D.Mel-2 cells. All molecular cloning for the human genes was done in the Gateway System (Invitrogen) as described above using publicly available full-length cDNA clones: *PLK4* (NM_014264, IMAGE: 5273226), *CEP152* (BC117182, IMAGE: 40125733), and *CPAP* (BC113664, IMAGE: 8322716). For the overexpression of GFP–PLK4 and CEP152–GFP, full-length PLK4 and CEP152 CDSs were cloned into two novel Gateway destination expression vectors that we generated based on the pcDNA4/TO vector from Invitrogen. All of these have a constitutive CMV promoter, TetO2 regulatory sequence, puromycin and ampicillin resistance as well as a Gateway cassette containing either N-terminal EGFP (p4NT2-GFP) or C-terminal EGFP (p4CT3-GFP). For the centrosome reduplication experiment U2OS cells were transfected with siRNAs targeting *PLK4*, *CEP152* or control (RSC). S-phase arrest was induced by addition of 5 μM aphidicolin 24 h after transfection (for PLK4) or 48 h after transfection (in the case of CEP152 and RSC). Cells were analysed 72 h after aphidicolin addition.

19. Fox, J. D., Routzahn, K. M., Bucher, M. H. & Waugh, D. S. Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers. *FEBS Lett.* **537,** 53–57 (2003).
20. D'Avino, P. P., Savoian, M. S., Capalbo, L. & Glover, D. M. RacGAP50C is sufficient to signal cleavage furrow formation during cytokinesis. *J. Cell Sci.* **119,** 4402–4408 (2006).
21. Basto, R. *et al.* Flies without centrioles. *Cell* **125,** 1375–1386 (2006).
22. Mottier-Pavie, V. & Megraw, T. L. *Drosophila* Bld10 is a centriolar protein that regulates centriole, basal body, and motile cilium assembly. *Mol. Biol. Cell* **20,** 2605–2614 (2009).
23. Warn, R. M. & Warn, A. Microtubule arrays present during the syncytial and cellular blastoderm stages of the early *Drosophila* embryo. *Exp. Cell Res.* **163,** 201–210 (1986).
24. Frangioni, J. V. & Neel, B. G. Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins. *Anal. Biochem.* **210,** 179–187 (1993).

doi:10.1038/nature09390

Phosphorylation of the CPC by Cdk1 promotes chromosome bi-orientation

Tatsuya Tsukahara[1,2], Yuji Tanno[1] & Yoshinori Watanabe[1,2]

Successful partition of replicated genomes at cell division requires chromosome attachment to opposite poles of mitotic spindle (bi-orientation). Any defects in this regulation bring about chromosomal instability, which may accelerate tumour progression in humans. To achieve chromosome bi-orientation at prometaphase, the chromosomal passenger complex (CPC), composed of catalytic kinase Aurora B and regulatory components (INCENP, Survivin and Borealin), must be localized to centromeres to phosphorylate kinetochore substrates[1–7]. Although the CPC dynamically changes the subcellular localization, the regulation of centromere targeting is largely unknown[1]. Here we isolated a fission yeast cyclin B mutant defective specifically in chromosome bi-orientation. Accordingly, we identified Cdk1 (also known as Cdc2)–cyclin-B-dependent phosphorylation of Survivin. Preventing Survivin phosphorylation impairs centromere CPC targeting as well as chromosome bi-orientation, whereas phosphomimetic Survivin suppresses the bi-orientation defect in the cyclin B mutant. Survivin phosphorylation promotes direct binding with shugoshin[8,9], which we now define as a conserved centromeric adaptor of the CPC. In human cells, the phosphorylation of Borealin has a comparable role. Thus, our study resolves the conserved mechanisms of CPC targeting to centromeres, highlighting a key role of Cdk1–cyclin B in chromosome bi-orientation.

The dynamic regulation of kinetochore-microtubule attachment is governed by the CPC, consisting of Aurora B, INCENP, Survivin and Borealin/Dasra (Ark1, Pic1, Bir1 and Nbl1 in fission yeast)[1,10], which concentrates on centromeres by passing through the chromosome arms from prophase to metaphase, and suddenly relocating to the spindle midzone at the metaphase-anaphase transition[1]. A central question is how the CPC is targeted to centromeres in coordination with the cell cycle progression. The fission yeast cyclin B Cdc13 might have unique functions in mid-M phase in addition to its cell cycle role at the G2/M transition[11]. By performing PCR-based mutagenesis of the *cdc13*$^+$ gene, we isolated a unique allele *cdc13-M7* (amino acid change E255G), which showed a specific defect in chromosome segregation (marked sensitivity to TBZ (thiabendazole), a microtubule depolymerizing drug) but not in G2/M transition (no cell elongation) (Fig. 1a, b). Nuclear staining of anaphase cells showed that the *cdc13-M7* mutant, but not the conventional *cdc13-117* mutant, often exhibited lagging chromosomes at anaphase (Fig. 1c). Time-lapse observations of centromere 2 (*cen2*–GFP) in *cdc13-M7* cells confirmed that single chromatids are indeed lagging on the anaphase spindle, indicative of merotelic attachment (Fig. 1d). Thus, in addition to the canonical role in cell cycle progression at G2/M, the Cdk1–cyclin B complex may have an important role in chromosome bi-orientation.

Because the *cdc13-M7* mutant has a phenotype very similar to that of the Aurora B mutant (*ark1-T7*) (Supplementary Fig. 1), we envisaged that the CPC might be the relevant target of Cdc2 (fission yeast Cdk1). By performing *in vitro* phosphorylation of all CPC components purified from bacteria, we identified Bir1/Survivin as the best Cdc2 substrate (Fig. 1e). Immunoblots using cell extracts prepared from asynchronous (mostly G2 phase) or mitotically arrested cells identified slower-migrating forms of Bir1 only in mitotic extracts (Fig. 1f). Phosphatase treatment abolished the slower migration, indicating that the Bir1 protein is phosphorylated in mitosis (Fig. 1f). Accordingly, the Bir1 protein has eight consensus residues for CDK phosphorylation (see Fig. 4a), which are phosphorylatable *in vitro* by Cdc2 (Fig. 1g). We replaced endogenous *bir1*$^+$ with the non-phosphorylatable mutant *bir1-8A*, in which all CDK sites were replaced with alanines, and analysed the cell extracts by immunoblot, showing that the mitotic band-shift was abolished with the mutant Bir1-8A protein (Fig. 1h). The *in vivo* phosphorylation of Bir1 at prometaphase but not interphase was further confirmed by a phospho-specific antibody against one of the CDK sites, Bir1-pS244 (Supplementary Fig. 2). Moreover, the phosphorylation of Bir1 was reduced in the *cdc13-M7* mutant in prometaphase (Fig. 1i). These results indicate that Bir1 is phosphorylated at the CDK consensus sites *in vivo*, most likely by Cdc2–Cdc13. Accordingly, *bir1-8A* cells showed marked sensitivity to TBZ, whereas the phosphomimetic *bir1-8D* mutation, in which all CDK sites were replaced with aspartic acid, did not increase the sensitivity (Fig. 1j). Moreover, *bir1-8A* cells, but not *bir1-8D* cells, showed a high incidence of lagging chromosomes at anaphase (Fig. 1d, k). If Bir1 is a crucial substrate of the Cdc2–Cdc13 complex for chromosome bi-orientation, *bir1-8D* might suppress the *cdc13-M7* mutant. Indeed, this was the case (Fig. 1j, k), whereas *cdc13-117* and another cell division cycle mutant, *cdc25-22*, were not suppressed by *bir1-8D* (Supplementary Fig. 3). Taken together, we conclude that the phosphorylation of Bir1 by Cdc2–Cdc13 is crucial for chromosome bi-orientation.

Given that Bir1 has an important role in centromere targeting of the CPC[12,13], we examined the behaviour of the CPC in the *bir1-8A* mutant. Bir1 usually distributes in the nucleolus during interphase, whereas it concentrates at centromeres on the metaphase spindle, with further relocation to the spindle midzone at anaphase. In contrast, Bir1-8A persists in mitosis with very little relocation to centromeres even at metaphase, whereas the anaphase localization remains intact (Supplementary Fig. 4). Accordingly, in *bir1-8A* cells centromeric signals of Ark1/Aurora B were reduced selectively in metaphase (Fig. 2a, b). A chromatin immunoprecipitation (ChIP) assay confirmed the reduced centromeric localization of Ark1 and Bir1 (Supplementary Fig. 5). In contrast, the phosphomimetic *bir1-8D* mutant retained the correct localization of Ark1 in both metaphase and anaphase (Fig. 2a). Consistent with these results, *cdc13-M7* cells, in which Bir1 phosphorylation is impaired, show reduced centromeric localization of CPC, but the localization is restored if the *bir1-8D* mutation is also present (Supplementary Fig. 6). These results indicate that Bir1 phosphorylation by Cdc2 is important for targeting the CPC to centromeres at metaphase. To examine whether Bir1-8A is solely deficient in its ability to localize at centromeres, we fused the Bir1-8A protein to cyan fluorescent protein (CFP) and a chromodomain (CD), which binds to Lys-9-methylated histone H3 located principally at centromeric heterochromatin[12]. The expression of Bir1-8A–CD suppressed TBZ sensitivity and the incidence of lagging chromosomes at comparable level to

[1]Laboratory of Chromosome Dynamics, Institute of Molecular and Cellular Biosciences, University of Tokyo, Yayoi, Tokyo 113-0032, Japan. [2]Graduate Program in Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Yayoi, Tokyo 113-0032, Japan.

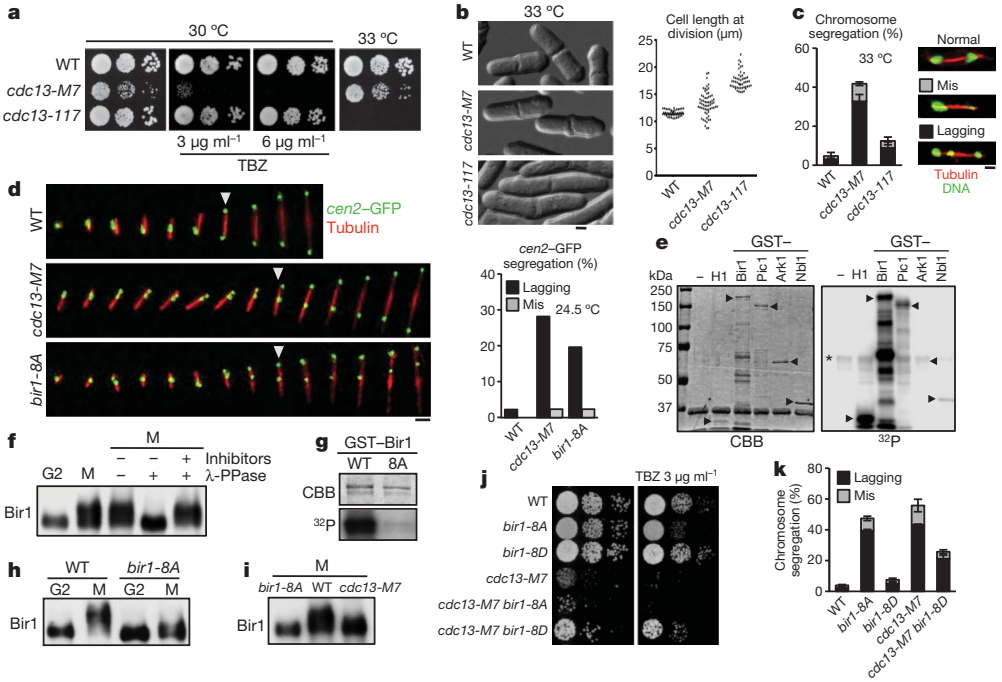


Figure 1 | Cdk1–cyclin B has a role in chromosome bi-orientation by phosphorylating Bir1. **a**, Serial dilutions of the indicated strains were spotted on yeast extract (YE) plates containing 0, 3 or 6 μg ml^{-1} TBZ and incubated at 30 °C. YE plate incubation at 33 °C is also shown. WT, wild type. **b**, Cultures of the indicated strains were shifted from 25 °C to 33 °C for 6 h and then fixed. Lengths of the septated cells were measured with the Image J software ($n > 50$ cells). Representative images of septated cells in each strain are shown. **c**, Cells in **b** were stained for DNA. The spindles were visualized by expressing mCherry–Atb2 (α2-tubulin). Frequencies of lagging chromosomes and mis-segregation at anaphase cells ($n > 100$ cells) were examined. Error bars represent s.d. ($n = 3$ experiments). **d**, Time-lapse observation of *cen2*–GFP and tubulin (mCherry–Atb2) in the indicated cells at 24.5 °C. Arrowheads indicate the onset of anaphase. Frequencies of lagging chromosomes or mis-segregation (nondisjunction) of *cen2*–GFP are shown ($n > 80$ cells). **e**, GST–Bir1, –Pic1, –Ark1-KR and –Nbl1 were incubated with Suc1-bound Cdc2, and analysed for phosphate incorporation (^{32}P) and protein levels (Coomassie brilliant blue, CBB). Arrowheads indicate the full-length recombinant proteins and histone H1; the asterisk indicates the phosphorylation of Cdc13 in Suc1-beads bound complex. **f**, Lambda protein phosphatase (λ-PPase) treatment of fission yeast extracts was analysed by immunoblot using the anti-Bir1 antibody. G2, asynchronous culture; M, mitotic arrest (produced by overexpressing *mad2*$^{+}$). **g**, GST–Bir1-WT and -8A were incubated with Suc1-bound Cdc2, and analysed for phosphate incorporation (^{32}P) and protein levels (CBB). **h**, Immunoblot analysis of Bir1 protein in wild-type and *bir1-8A* cells at interphase and prometaphase (arrested by *nda3-KM311* inactivation). **i**, Immunoblot analysis of Bir1 protein in wild-type, *bir1-8A* and *cdc13-M7* cells at prometaphase. **j**, Serial dilutions of the indicated strains were spotted onto YE plates containing 0 or 3 μg ml^{-1} TBZ and incubated at 28 °C. **k**, The indicated strains expressing mCherry–Atb2 were grown at 18 °C for 5.5 h and stained for DNA. Frequencies of lagging chromosomes and mis-segregation at anaphase ($n > 100$ cells) were examined. Error bars represent s.d. ($n = 3$ experiments).

Bir1–CD, whereas neither Bir1-8A nor chromodomain alone suppressed them (Fig. 2c, d). These results indicate that once they are tethered at centromeres, the functionality is indistinguishable between Bir1 and Bir1-8A. Supporting this conclusion, complex formation of the CPC was intact in *bir1-8A* cells (Supplementary Fig. 7).

Because CPC localization depends largely on shugoshin Sgo2 in fission yeast[12,13], we examined whether the phosphorylation of Bir1 influences this pathway. Two-hybrid assays revealed that Sgo2 interacts with the amino-terminal region of Bir1 (Bir1-N), which contains a CDK phosphorylation cluster (Supplementary Fig. 8a), but not with other CPC components (Fig. 2e and Supplementary Fig. 8b). The interaction is mediated through the coiled-coil region of Sgo2 (Fig. 2f), a region required for CPC targeting to centromeres (Supplementary Fig. 9). This finding is comparable to the fact that another shugoshin Sgo1 interacts with its effector Par1 (PP2A-B56 subunit) through the coiled-coil domain[12,14]. The fact that Par1 does not interact with Sgo2 as Bir1-N does not interact with Sgo1 (Fig. 2g), indicates the distinct specificity of the interaction between these shugoshins and their effectors. To explore the contribution of the CDK phosphorylation cluster in Bir1-N to the interaction with Sgo2, we made non-phosphorylatable Bir1-N-5A and phosphomimetic Bir1-N-5D, and examined their interaction in a two-hybrid assay. Bir1-N-5A abolished the interaction with Sgo2, whereas Bir1-N-5D retained the interaction (Fig. 2h). The *bir1-5A* and *bir1-8A* mutants have comparable defects in centromere targeting of the CPC (Supplementary Fig. 10), implying that the phosphorylation cluster within Bir1-N has a predominant role in this function. Finally, immunoprecipitation assays showed that Bir1-8A had lost the ability to form a complex with Sgo2 (Fig. 2i). Collectively, we conclude that the Cdc2-dependent phosphorylation of Bir1 regulates CPC targeting to centromeres, at least in part, by promoting the interaction with Sgo2.

We next examined the conservation of the mechanism of CPC targeting to centromeres in HeLa (human; h) cells. Although the depletion of either hSgo1 or hSgo2 did not fully impair hAurora B localization[12], co-depletion of hSgo1 and hSgo2, or single hBub1 depletion that displaces both hSgo1 and hSgo2[15–17], strikingly abolished the centromeric enrichment of hAurora B (Fig. 3a and Supplementary Fig. 11). Two-hybrid assays indicated that the coiled-coil domains of both hSgo1 and hSgo2 interact specifically with hBorealin/Dasra[18,19] rather than hSurvivin (Fig. 3b and Supplementary Fig. 12). It came to our notice that hSurvivin is much shorter than fission yeast Survivin/Bir1 and lacks the phosphorylation cluster (see Fig. 4a). Instead, hBorealin but not fission yeast Borealin/Nbl1 possesses accessory sequences[10], in which we found a cluster of putative CDK phosphorylation sites that are indeed phosphorylatable *in vitro* (Fig. 3c and Supplementary Fig. 13). To examine the *in vivo* phosphorylation of hBorealin, we expressed GFP-tagged hBorealin in 293T cells and analysed mitotic extracts by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and subsequent immunoblot using GFP antibodies. Slower migration appeared with wild-type hBorealin, but never with hBorealin-7A in which all putative Cdk1 consensus sites are replaced with alanines

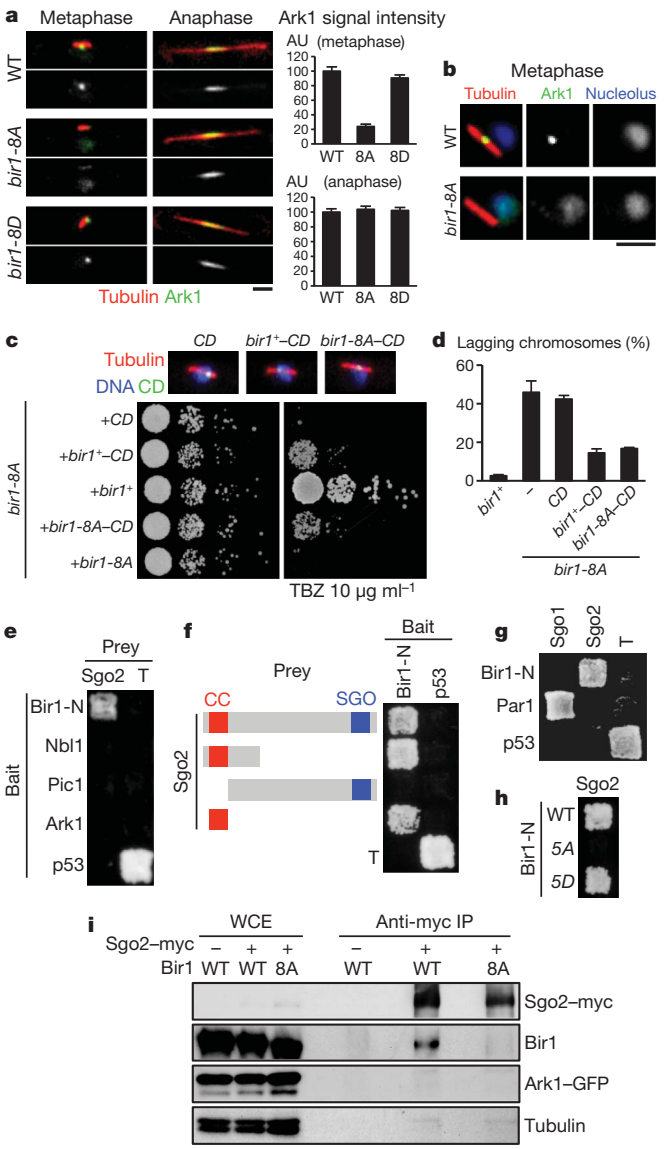


Figure 2 | Phosphorylation of Bir1 promotes the association with Sgo2 and targeting the CPC to centromeres. **a**, Ark1–GFP was detected at metaphase and anaphase in wild-type, *bir1-8A* and *bir1-8D* cells expressing mCherry–Atb2. Ark1–GFP signals were measured at metaphase and anaphase. Error bars indicate s.e.m. ($n > 25$ cells). AU, arbitrary units. **b**, Ark1–GFP was detected at metaphase in wild-type and *bir1-8A* cells expressing mCherry–Atb2 and Gar1–CFP (Nucleolus). **c**, Serial dilutions of the indicated strains were spotted on YE plates containing 0 or 10 μg ml^{-1} TBZ and incubated at 28 °C. Photos show the representative localization of chromodomain, Bir1–CD and Bir1-8A–CD at metaphase. **d**, Frequencies of lagging chromosomes at anaphase ($n > 100$ cells) were examined at 18 °C. Error bars represent s.d. ($n = 3$ experiments). **e**, Yeast two-hybrid assay indicates that Sgo2 interacts with the N terminus of Bir1 (Bir1-N). The pair of p53-T-antigen serves as positive controls (also see Supplementary Fig. 8). **f**, Yeast two-hybrid assay with Sgo2 deletions. **g**, Yeast two-hybrid assay indicates that Sgo1 interacts with Par1 but not Bir1, whereas Sgo2 interacts with Bir1 but not Par1. **h**, Yeast two-hybrid assay with Bir1-N-WT, -5A and -5D. **i**, Co-immunoprecipitation of Bir1 and Sgo2. Whole cell extracts (WCE) were prepared from mitotic cells expressing Sgo2–myc and Ark1–GFP. Sgo2–myc was immunoprecipitated (IP) with anti-myc antibodies.

(Fig. 3d). Further, the slower migration was diminished by treating the extracts with (lambda-)phosphatase (Fig. 3d), indicating that the Cdk1 consensus sites in hBorealin are indeed phosphorylated *in vivo*. Two-hybrid assays indicated that the phosphorylation of hBorealin is important for the interaction with hSgo1 and hSgo2 (Fig. 3e). Moreover, immunoprecipitation assays indicated that hBorealin

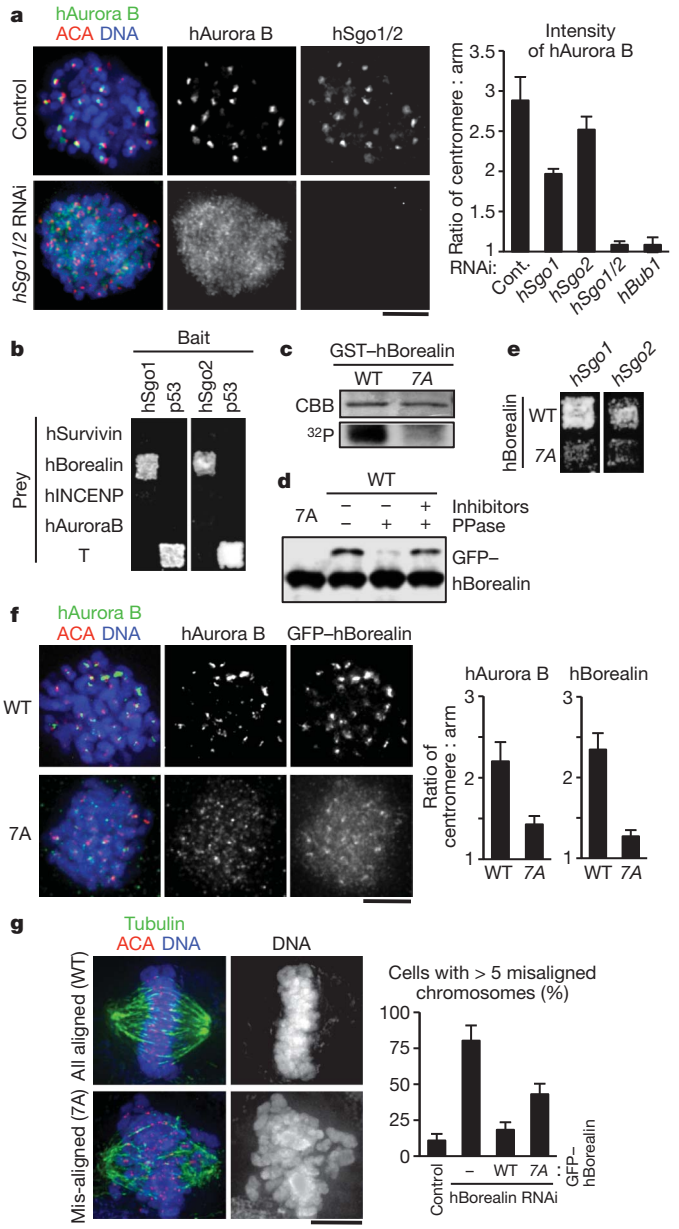


Figure 3 | Phosphorylation of Borealin by Cdk1 promotes the association with shugoshins and centromere targeting of the CPC in human cells. **a**, hAurora B signals at prometaphase were examined by immunostaining control and hSgo1/hSgo2 RNAi cells. The ratio of centromere and arm hAurora B signals was quantified (also see Supplementary Fig. 11a). Error bars represent s.e.m. ($n = 5$ cells). **b**, Yeast two-hybrid assay indicates that both hSgo1 and hSgo2 interacts with hBorealin (also see Supplementary Fig. 12b). **c**, GST–hBorealin-WT and -7A were incubated with Suc1-bound Cdc2, and analysed for phosphate incorporation (^{32}P) and protein levels (CBB). **d**, Mitotic extracts were prepared from prometaphase-arrested 293T cells expressing GFP–hBorealin-WT and -7A, treated with or without lambda-phosphatase (PPase) and analysed by immunoblot using the anti-GFP antibody. **e**, Yeast two-hybrid assay with hBorealin-WT and -7A. **f**, Signals of GFP–hBorealin and hAurora B were examined at prometaphase in hBorealin RNAi cells expressing RNAi-resistant GFP–hBorealin-WT and -7A. The ratios of centromere and arm GFP–hBorealin signals or hAurora B signals were quantified in the indicated cells. Error bars represent s.e.m. ($n = 8$ cells). **g**, HeLa cells expressing RNAi-resistant GFP–hBorealin-WT and -7A were treated with hBorealin siRNA. The indicated cells arrested at metaphase were examined for chromosome alignment. Representative images of GFP–hBorealin-WT and -7A cells are shown. Frequencies of cells exhibiting more than five misaligned chromosomes or multipolar spindles were measured ($n > 30$ cells). Error bars represent s.d. ($n = 3$ experiments). Scale bars, 5 μm.

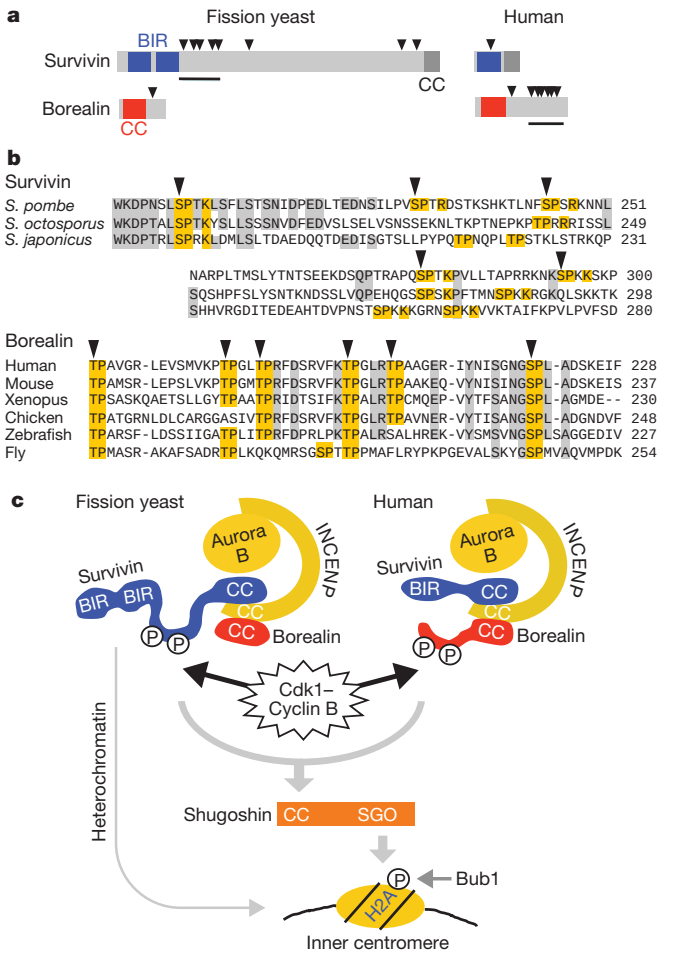


Figure 4 | Fission yeast Survivin and human Borealin have comparable roles in binding with shugoshin and CPC targeting. **a**, Schematic diagrams of Survivin and Borealin in fission yeast and humans. Conserved BIR domains (blue), coiled-coil regions (CC; grey and red) and CDK phosphorylation sites (arrowheads) are highlighted. **b**, Alignments of the cluster of CDK consensus sequences (yellow shade) of fission yeast Survivin and metazoan Borealin (underlined in **a**) are shown. **c**, Schematic model for the conserved centromere targeting of the CPC. Phosphorylation of fission yeast Survivin or human Borealin by Cdk1 promotes the CPC binding to the coiled-coil region of shugoshin, whereas shugoshin is targeted to centromeres by interacting with centromeric nucleosomes only when H2A is phosphorylated by Bub1 (ref. 9). The intertwined configuration of Survivin, Borealin and INCENP is based on the crystallographic study of the complex[26]. Note that a redundant centromere CPC targeting pathway that depends on heterochromatin assembly also acts in fission yeast[12] (also see Supplementary Fig. 18).

forms a complex with hSgo1 and hSgo2 in prometaphase, and this complex can be disrupted by inhibiting Cdk1 (Supplementary Fig. 14).

To delineate the significance of hBorealin phosphorylation *in vivo*, we examined the localization of hBorealin-7A in HeLa cells. Whereas wild-type hBorealin localized exclusively at centromeres in prometaphase, hBorealin-7A showed diffuse localization along the chromosomes (Fig. 3f and Supplementary Fig. 15). These results are consistent with a previous report identifying the carboxy-terminal region of hBorealin as being required for centromere targeting[18]. We further examined the ability of hBorealin-7A to complement the chromosome alignment defects in hBorealin RNAi cells. In control MG-132-treated cells, most chromosomes are aligned on the metaphase plate, whereas hBorealin RNAi cells show multiple misaligned chromosomes or multipolar spindles under the same conditions. The expression of wild-type hBorealin largely restored this alignment defect, whereas the expression of hBorealin-7A did so to a much lesser extent (Fig. 3g). Collectively, we conclude that phosphorylation of hBorealin promotes the interaction with shugoshins and CPC targeting to centromeres and that the mechanisms of CPC targeting to centromeres are overall conserved between fission yeast and human cells.

Our identification of a specific mutant allele of cyclin B in fission yeast showed that Cdk1 has a crucial role in chromosome bi-orientation. Recently, data with similar implications were obtained by observing mitotic HeLa cells in which cyclin B1 was inactivated by shRNA-based depletion[20]. Our results show that Cdk1 phosphorylates Survivin/Bir1 in fission yeast, but Borealin does so in human cells, both components of the CPC that have an essential role in localizing to centromeres by associating with shugoshin. This implies that a functionally comparable domain might have moved between different subunits of the CPC during evolution (Fig. 4a, see also Supplementary Fig. 16). Notably, the primary sequences around the phosphorylation cluster of hBorealin are conserved among metazoan Borealin family members, whereas the similarity of the sequences to those of fission yeast Survivin/Bir1 is not obvious (Fig. 4b). A recent study, however, showed that CDK phosphorylation sites are not always conserved at the primary sequence level, but are conserved in their formation as a cluster in a disordered region[21], supporting our conclusion.

Another chief finding of this study is that the shugoshin proteins turn out to be the conserved centromeric adaptor of the CPC (Fig. 4c). The localization of shugoshin at centromeres generally depends on the interaction between the conserved SGO motif in shugoshin and centromeric nucleosomes, which are phosphorylated by Bub1 (ref. 9, Fig. 4c). However, fission yeast Sgo2 localization is diminished by mutations in *bir1*$^{+}$ including *bir1-8A* (ref. 12, Supplementary Fig. 17), and the CPC affects the centromeric localization of shugoshin also in *Drosophila*, *Xenopus* and humans[12,22,23]. Therefore, shugoshin might not be a mere centromeric anchor to recruit the CPC; instead the complex formation of the CPC and shugoshin is important for co-targeting to centromeres, allowing us to define shugoshin as a centromeric CPC adaptor linked to the centromeric nucleosomes marked by Bub1 (Fig. 4c). In fission yeast, Sgo2 acts solely to facilitate CPC targeting to centromeres, whereas meiosis-specific Sgo1 acts solely to recruit PP2A (ref. 12). In contrast, human Sgo1 and Sgo2 share both roles[24,25] (Fig. 3). We therefore suggest that the ancestral shugoshin molecule might have had dual roles in targeting CPC as well as other centromeric proteins such as PP2A, which improves the fidelity of bi-orientation by strengthening cohesion at centromeres. These roles might be divided respectively into Sgo2 and Sgo1 in fission yeast, but shared between human Sgo1 and Sgo2. Finally, it should be noted that the shugoshin-Cdk1 is the major but not sole pathway to target the CPC to centromeres in fission yeast, as the heterochromatin assembly supports the residual localization of the CPC in *sgo2Δ* or *bir1-8A* cells[12] (Fig. 4c and Supplementary Fig. 18). Further characterization of this redundant pathway will help achieve a full understanding of the mechanisms setting up chromosome bi-orientation in eukaryotes.

METHODS SUMMARY

All *Schizosaccharomyces pombe* strains used in this study are listed in Supplementary Table 1. Methods used to generate the strains are described in Methods. To quantify the centromeric fluorescent signals, in-focus images of Ark1–GFP and Sgo2–GFP were taken with Axio Vision software (Carl Zeiss). We measured the average intensity of the centromeric dots on the spindles and subtracted the average background intensity. To measure the cell length, in-focus differential interference contrast images were taken and we measured the length of septated cells with the Image J software. ChIP assay and immunoprecipitation were conducted essentially as described previously[12,27]. For time-lapse imaging of mitotic progression, we used logarithmically growing cells and live recordings were performed on a Delta Vision P system (Applied precision) in an air-conditioned room maintained at 24.5 °C. Immunofluorescence staining of HeLa cells was performed as described[28]. To quantify the centromeric enrichment of hAurora B and GFP–hBorealin, in-focus images were taken with Delta Vision RT system (Applied precision). We measured the centromeric and chromosome arm signal intensities and subtracted by cytoplasmic background intensities, then the ratio between centromeres and arms were calculated. Immunoprecipitations of mitotic HeLa cells were performed as described in Methods. For *in vitro* phosphorylation,

Suc1-bound Cdc2 complex was purified from mitotic fission yeast cells. Cdc2 complex was incubated with γ-^{32}P-labelled ATP and recombinant proteins for 30 min at 30 °C.

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

Received 26 May; accepted 5 August 2010.
Published online 25 August 2010.

1. Ruchaud, S., Carmena, M. & Earnshaw, W. C. Chromosomal passengers: conducting cell division. *Nature Rev. Mol. Cell Biol.* **8,** 798–812 (2007).
2. Liu, D., Vader, G., Vromans, M. J., Lampson, M. A. & Lens, S. M. Sensing chromosome bi-orientation by spatial separation of aurora B kinase from kinetochore substrates. *Science* **323,** 1350–1353 (2009).
3. Tanaka, T. U. Bi-orienting chromosomes: acrobatics on the mitotic spindle. *Chromosoma* **117,** 521–533 (2008).
4. Cheeseman, I. M. & Desai, A. Molecular architecture of the kinetochore-microtubule interface. *Nature Rev. Mol. Cell Biol.* **9,** 33–46 (2008).
5. Lampson, M. A., Renduchitala, K., Khodjakov, A. & Kapoor, T. M. Correcting improper chromosome-spindle attachments during cell division. *Nature Cell Biol.* **6,** 232–237 (2004).
6. Hauf, S. *et al.* The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *J. Cell Biol.* **161,** 281–294 (2003).
7. Meraldi, P., Honda, R. & Nigg, E. A. Aurora kinases link chromosome segregation and cell division to cancer susceptibility. *Curr. Opin. Genet. Dev.* **14,** 29–36 (2004).
8. Watanabe, Y. Shugoshin: guardian spirit at the centromere. *Curr. Opin. Cell Biol.* **17,** 590–595 (2005).
9. Kawashima, S. A., Yamagishi, Y., Honda, T., Ishiguro, K. & Watanabe, Y. Phosphorylation of H2A by Bub1 prevents chromosomal instability through localizing shugoshin. *Science* **327,** 172–177 (2010).
10. Bohnert, K. A., Chen, J. S., Clifford, D. M., Vander Kooi, C. W. & Gould, K. L. A link between aurora kinase and Clp1/Cdc14 regulation uncovered by the identification of a fission yeast borealin-like protein. *Mol. Biol. Cell* **20,** 3646–3659 (2009).
11. Nasmyth, K. & Nurse, P. Cell division cycle mutants altered in DNA replication and mitosis in the fission yeast *Schizosaccharomyces pombe*. *Mol. Gen. Genet.* **182,** 119–124 (1981).
12. Kawashima, S. A. *et al.* Shugoshin enables tension-generating attachment of kinetochores by loading Aurora to centromeres. *Genes Dev.* **21,** 420–435 (2007).
13. Vanoosthuyse, V., Prykhozhij, S. & Hardwick, K. G. Shugoshin 2 regulates localization of the chromosomal passenger proteins in fission yeast mitosis. *Mol. Biol. Cell* **18,** 1657–1669 (2007).
14. Yamagishi, Y., Sakuno, T., Shimura, M. & Watanabe, Y. Heterochromatin links to centromeric protection by recruiting shugoshin. *Nature* **455,** 251–255 (2008).
15. Kitajima, T. S., Hauf, S., Ohsugi, M., Yamamoto, T. & Watanabe, Y. Human Bub1 defines the persistent cohesion site along the mitotic chromosome by affecting shugoshin localization. *Curr. Biol.* **15,** 353–359 (2005).
16. Tang, Z., Sun, Y., Harley, S. E., Zou, H. & Yu, H. Human Bub1 protects centromeric sister-chromatid cohesion through Shugoshin during mitosis. *Proc. Natl Acad. Sci. USA* **101,** 18012–18017 (2004).
17. Huang, H. *et al.* Tripin/hSgo2 recruits MCAK to the inner centromere to correct defective kinetochore attachments. *J. Cell Biol.* **177,** 413–424 (2007).
18. Gassmann, R. *et al.* Borealin: a novel chromosomal passenger required for stability of the bipolar mitotic spindle. *J. Cell Biol.* **166,** 179–191 (2004).
19. Sampath, S. C. *et al.* The chromosomal passenger complex is required for chromatin-induced microtubule stabilization and spindle assembly. *Cell* **118,** 187–202 (2004).
20. Chen, Q., Zhang, X., Jiang, Q., Clarke, P. R. & Zhang, C. Cyclin B1 is localized to unattached kinetochores and contributes to efficient microtubule attachment and proper chromosome alignment during mitosis. *Cell Res.* **18,** 268–280 (2008).
21. Holt, L. J. *et al.* Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* **325,** 1682–1686 (2009).
22. Resnick, T. D. *et al.* INCENP and Aurora B promote meiotic sister chromatid cohesion through localization of the Shugoshin MEI-S332 in *Drosophila*. *Dev. Cell* **11,** 57–68 (2006).
23. Boyarchuk, Y., Salic, A., Dasso, M. & Arnaoutov, A. Bub1 is essential for assembly of the functional inner centromere. *J. Cell Biol.* **176,** 919–928 (2007).
24. Kitajima, T. S. *et al.* Shugoshin collaborates with protein phosphatase 2A to protect cohesin. *Nature* **441,** 46–52 (2006).
25. Xu, Z. *et al.* Structure and function of the PP2A-shugoshin interaction. *Mol. Cell* **35,** 426–441 (2009).
26. Jeyaprakash, A. A. *et al.* Structure of a Survivin–Borealin–INCENP core complex reveals how chromosomal passengers travel together. *Cell* **131,** 271–285 (2007).
27. Yokobayashi, S., Yamamoto, M. & Watanabe, Y. Cohesins determine the attachment manner of kinetochores to spindle microtubules at meiosis I in fission yeast. *Mol. Cell. Biol.* **23,** 3965–3973 (2003).
28. Liu, S. T. *et al.* Human CENP-I specifies localization of CENP-F, MAD1 and MAD2 to kinetochores and is essential for mitosis. *Nature Cell Biol.* **5,** 341–345 (2003).

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

Acknowledgements We thank S. Hauf for critically reading the manuscript and the Yeast Genetic Resource Center (YGRC) for yeast strains, and all the members of our laboratory, particularly S. A. Kawashima, for their valuable support and discussion. This work was supported in part by Global COE Program (Integrative Life Science Based on the Study of Biosignaling Mechanisms), JSPS Research Fellowship (to T.T. and Y.T.) and a Grant-in-Aid for Specially Promoted Research, MEXT, Japan (to Y.W.).

Author Contributions Most experiments were performed by T.T. except those in Fig. 3a and Supplementary Fig. 11, which were performed by Y.T. Experimental design and interpretation of data were conducted by all authors. Y.W. supervised the project, and T.T. and Y.W. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.W. (ywatanab@iam.u-tokyo.ac.jp).

METHODS

***Schizosaccharomyces pombe* strains.** Deletion and tagging of endogenous *sgo2*$^+$, *ark1*$^+$ and *bir1*$^+$ by GFP or 13 copies of Myc (abbreviated as myc) were performed using the PCR-based gene targeting method for *S. pombe*[29]. For screening TBZ-sensitive *cdc13* mutants, we used a low-fidelity PCR method as reported previously[30]. We screened for alleles that showed marked sensitivity to TBZ (10 μg ml^{-1}) at 30 °C. To express Bir1–CFP–2CD, a sequence encoding CFP and two copies of the chromodomain (CD) of Swi6 (residues 69–215) were fused to the C terminus of Bir1 and cloned under the promoter P*adh81* (a weak version of the *adh1*$^+$ promoter)[12]. The resulting plasmid was linearized and integrated at the *lys1*$^+$ locus of chromosome I by using the *hyg*^{r} marker. To generate the *bir1-8A* and *8D* strains, S202, S229, S244, S278, S294, T393, T831 and T908 were changed to alanines or aspartic acids, respectively, using Kunkel methods. The genomic *bir1* fragments carrying the mutations were then transformed into *bir1::ura4*$^+$ cells (additional copy of *bir1*$^+$ is integrated at the *lys1*$^+$ locus) and the integration at *bir1::ura4*$^+$ was selected by 5-fluoroorotic acid (5-FOA) resistance and confirmed by PCR. To generate the *sgo2-Δcoiled coil* strain, 54 amino acids (residues 10–63) were deleted using a PrimeSTAR mutagenesis kit (Takara). The genomic *sgo2-Δcoiled coil* fragment was then transformed into *sgo2::ura4*$^+$ cells and selected for 5-FOA resistance. To generate the *bir1-5A* or *-5D* strains, the genomic fragments of *bir1* ORF (1.8 kilobases from the 5′ end) carrying the mutations at S202, S229, S244, S278 and S294 were transformed into the temperature-sensitive *bir1-T1* mutant. Colonies grown at 34 °C were selected and the genotype was confirmed by sequencing. The *GFP–bir1*$^+$ and *-8A* strains were generated as described previously[31]. To express mCherry–Tubulin, a sequence encoding mCherry was fused to the N terminus of *atb2*$^+$, cloned under the promoter P*adh15* (a weak version of the *adh1*$^+$ promoter), and integrated into the locus adjacent to the *zfs1*$^+$ gene of chromosome II (denoted by the *z* locus) using the *nat*^{r} marker[14]. All strains used are listed in Supplementary Table 1.

Two-hybrid assay. The constructs of *bir1*, *pic1*, *nbl1*, *par1*, *hSgo1*, *hSgo2*, *hINCENP*, *hAurora B*, *hSurvivin* and *hBorealin* derivatives were amplified by PCR and cloned into pGBKT7 vectors and used as bait. The constructs of *bir1*, *nbl1*, *sgo1*, *sgo2*, *hINCENP*, *hAurora B*, *hSurvivin* and *hBorealin* derivatives were amplified by PCR and cloned into pGADT7 vectors and the construct of *pic1* was cloned into pAct2 vector then used as prey. hAurora B plasmids were mutated at K106 to R by PrimeSTAR mutagenesis (Takara). pGBKT9 and pGADT9 containing *ark1-KR* vectors were generous gifts from T. Hunter. These plasmids were transformed into *Saccharomyces cerevisiae* AH109 strain. Plates lacking histidine or both histidine and adenine were used as selective media, with the addition of appropriate amounts of 3-amino-1,2,4-triazole.

Growth. For prometaphase arrest, we used the *nda3-KM311* mutation[32] and grew cells at 17 °C for 11 to 13 h (for ChIP experiments for the *cdc13-M7* mutant, we grew *nda3-KM311* cells for 8 h to adjust the mitotic population). For metaphase arrest, we used the pREP1(mad2$^+$) plasmids expressed by the *nmt1*$^+$ promoter (gift from T. Matsumoto) and grew cells in minimal medium at 30 °C for 22 h (for wild-type cells) or 25 h (for *bir1-8A* mutants). For analysis of the *cdc13-117* mutant, cells were grown at 33 °C for 6 h.

Quantification of fluorescent signals. To quantify the fluorescent signals at centromeres and spindle midzones, in-focus images of Ark1–GFP and Sgo2–GFP cells were taken with Axio Vision imaging software (Carl Zeiss). We measured the average intensity of the centromeric dots on the spindles (centromeres) or line signals on the spindles (spindle midzones) and subtracted the average background intensity.

Measurement of cell length at division. To measure cell length at division, in-focus differential interference contrast images were taken with the Axio Vision imaging software (Carl Zeiss). We measured the length of septated cells with the Image J software.

Antibodies. To generate anti-Bir1 antibodies, the His-tagged full-length Bir1 protein was expressed, purified from bacterial cells and used to immunize rabbits (Keari). Antibodies were purified from anti-serum with His-Bir1-conjugated cyanogen bromide-activated Sepharose (GE) and dialysed against PBS. To generate phospho-specific antibodies against Bir1-pS244, phosphorylated peptides (C+LNF(pS)PSRKNN) were used to immunize rabbits (OPERON). Antibodies were purified from anti-serum with phosphorylated-peptide-conjugated cyanogen bromide-activated Sepharose and dialysed against PBS. Antibodies were further purified by passing them through non-phospho-peptide-conjugated cyanogen bromide-activated Sepharose.

Chromatin immunoprecipitation (ChIP) assay. The procedure was carried out essentially as described previously[27]. Anti-Bir1 polyclonal antibodies, anti-GFP polyclonal antibodies (Living Colours Full-length A.v. Polyclonal Antibody, BD), anti-H3 polyclonal antibodies (Abcam) and anti-H3S10 phospho-specific antibodies (UPSTATE) were used for immunoprecipitation. DNA prepared from whole cell extracts or immunoprecipitated fractions was analysed by quantitative PCR with the ABI PRISM7000 system (Applied Biosystems) using SYBR Premix Ex Taq (Perfect Real Time) (Takara). The primers used for PCR were all described previously[12,27,33]. We included control IgG immunoprecipitation in each experiment to account for nonspecific binding in the ChIP fractions.

Co-immunoprecipitation from fission yeast extracts. Cells (PM255, PL236 and PL237 for Sgo2 IP (Fig. 2i)) were cross-linked after incubation in minimal medium at 30 °C for 22 h (PM255 and PL236) or 25 h (PL237) (to induce the expression of Mad2) by treatment with 0.9% formaldehyde (for Ark1–GFP immunoprecipitation in Supplementary Figure 7, JY902, PL335 and PL336 were arrested at prometaphase by *nda3-KM311* inactivation before cross-linking). After washing with buffer 1 (50 mM HEPES-KOH (pH 7.5), 140 mM NaCl, 1 mM EDTA, 0.1% TritonX-100), the cells were resuspended in buffer 1 containing 1 mM PMSF, complete protease inhibitors (Roche), 3 mM $CaCl_2$, and 4 units per 5×10^9 cell nuclease micrococcal (SIGMA), and then lysed with Multi beads shocker (Yasui Kikai). Crude cell extracts were sonicated five times (15 s each time) and the supernatant was collected after centrifugation. Cell extracts were incubated with anti-myc antibody (9E10, Santa Cruz) or anti-GFP antibody (BD) for 1.5 h at 4 °C. Protein A beads (GE) were added and incubation was continued for 1.5 h at 4 °C. After washing with buffer 1, we analysed the immunoprecipitates by SDS–PAGE and western blotting with anti-Bir1 (1:1,000), anti-GFP (1:800, Roche), anti-myc (1:1,000, 9E10), anti-Cdc2 (1:1,000, PSTAIR) and TAT1 (1:5,000) antibodies.

Time-lapse imaging. To measure the onset of anaphase, wild-type, *bir1-8A* or *cdc13-M7* cells were cultured in YE medium at 23 °C, and then live cell recordings were obtained on a DeltaVision P system (Applied Precision) in an air-conditioned chamber maintained at 24.5 °C. A glass bottom dish (Matsunami) coated with 0.2% Concanavalin A (Sigma) was used to mount cells. During observation, the cells transferred to the glass-bottom dish were supplied with minimal medium. Images were acquired by *Z*-sectioning and stacked using 'quick projection' in the softWoRx software.

Immunoprecipitation to detect *in vivo* phosphorylation of Bir1. Cultured cells were collected, washed with HB buffer (25 mM MOPS (pH 7.2), 15 mM $MgCl_2$, 15 mM EGTA, 60 mM β-glycerophosphate, 0.1 mM Na-orthovanadate, 0.1 mM NaF, 15 mM *p*-nitrophenylphosphate, 1% Triton-X100, 1 mM dithiothreitol, 1 mM PMSF, complete protease inhibitor (Roche)), disrupted using the Multi-bead shocker (Yasui Kikai) and centrifuged, and the crude lysates were boiled with 1% SDS and 10 mM dithiothreitol. The supernatants were collected after centrifugation and diluted 10 times in the same buffer. Cell extracts were incubated with protein A beads pre-bound with anti-Bir1 antibody for 2 h at 4 °C. After washing with HB buffer, we analysed the immunoprecipitates by SDS–PAGE and western blotting with anti-Bir1 and anti-Bir1 S244 phospho-specific antibodies.

***In vitro* kinase assay.** Kinases and substrates were incubated with HB buffer at 30 °C for 30 min in the presence of [γ-^{32}P]ATP. Incorporation of the radioactive phosphate group was visualized by autoradiography and protein level was analysed by staining with Coomassie brilliant blue. For expressing His-Bir1 derivatives, *bir1*$^+$ ORF was amplified by PCR and cloned into pET19b vectors. For expressing GST–Bir1 or GST–hBorealin derivatives, *bir1*$^+$, *hBorealin*$^+$ ORF were amplified by PCR and cloned into pGEX4T-2 or pGEX-4T-3 vectors. For the purification of mitotic Cdc2, we pulled down the Cdc2 complex from prometaphase-arrested fission yeast cell extracts using Suc1 beads (Millipore).

Lambda-phosphatase assay. Mitotic extracts prepared as for the co-immunoprecipitation assay were treated with lambda-phosphatase (NEB) for 30 min at 30 °C. We used 10 mM Na_3VO_4, 50 mM NaF, 50 mM EDTA and 50 mM β-glycerophosphate as phosphatase inhibitors.

Cell culture and synchronization. HeLa cells and 293T cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS, 0.03% L-glutamine, 100 U ml^{-1} penicillin and 100 μg ml^{-1} streptomycin. Nocodazole, MG-132 and thymidine were used at 330 nM, 20 μM and 2 mM, respectively, for synchronization. For measurement of Aurora B localization at centromeres, cells were synchronized by single thymidine block and release and then arrested by adding nocodazole. For chromosome alignment assay, cells were synchronized by single thymidine block and release, then arrested at metaphase by adding MG-132 for 5 h.

Expression of GFP–hBorealin. Full-length wild type and hBorealin-7A (T106A, T172A, T185A, T189A, T199A, T204A and S219A) tagged with GFP at their N termini were inserted into pcPURO vectors for transient expression. Vectors were transfected into HeLa cells or 293T cells using Lipofectamine LTX (Invitrogen) and FuGENE 6 (Roche) respectively.

RNAi and immunostaining of HeLa cells. Synthetic sense and antisense oligonucleotides for RNAi of hSgo1 (ref. 15), hSgo2 (ref. 24), hBub1 (ref. 15) and hBorealin (ref. 34) were obtained from JbioS. siRNA transfection was performed as described[15]. Sequences of siRNA are hSgo1: 5′-GUCUACUGAUAAUGUC UUATT-3′, hSgo2: 5′-GCACUACCACUUUGAAUAATT-3′, hBub1: 5′-CCA

GUGAGUUCCUAUCCAATT-3′, hBorealin: 5′-AGGUAGAGCUGUCUGUU CATT-3′.

Immunofluorescent staining was performed as described[28] with anti-hSgo1, anti-hSgo2 (ref. 15, 1:1,000), anti-hAurora B (1:1,000, BD Biosciences), anti-hBub1 (1:1,000, Chemicon), anti-GFP (1:1,000, Invitrogen), anti-GFP (1:1,000, Roche), anti-Tubulin (1:5,000, DM1A; Sigma), anti-Hec1 (1:500, Abcam), anti-CENP-A pS7 (1:500, Cell Signaling), anti-CENP-A (1:1,000, MBL) and ACA (anticentromere antibody) (1:100, MBL). Secondary antibodies were Alexa Fluor 588 anti-rabbit or mouse antibodies (1:1,000, Molecular Probes), Cy3-conjugated anti-rabbit or mouse antibodies (1:1,000, Chemicon), and Alexa Fluor 647 anti-human antibodies (1:1,000, Molecular Probes). DNA was stained with 3 μg ml^{-1} Hoechst 33342. Images were captured by DeltaVision softWoRx software (Applied Precision) and processed by deconvolution and *z*-stack projection.

Quantification of the fluorescent signals in HeLa cells. Immunostained images were obtained with the use of DeltaVision softWoRx software (Applied Precision). The fluorescent intensities at centromeres and in the chromosome arm regions were corrected by subtracting the intensity of the cytoplasm, and the ratios of the centromeres and arms were calculated.

Immunoprecipitation from human cell extracts. For immunoprecipitation using control or hSgo1/2 RNAi cells, nocodazole-arrested cells were collected and washed once with PBS, then resuspended in CSK buffer (10 mM PIPES pH 6.8, 100 mM NaCl, 10% glycerol, 3 mM $MgCl_2$, 1 mM EGTA, 1 mM dithiothreitol, 0.25 mM PMSF, 0.1 mM ATP, complete protease inhibitor mixture, PhosSTOP phosphatase inhibitor mixture) containing 1 U ml^{-1} DNase I. After three freeze-thaw cycles, the suspension was centrifuged at 16,100*g* for 15 min and supernatants were used for immunoprecipitation. The preparation of chromatin extracts and immunoprecipitation were performed as described previously[12] using anti-GFP antibodies (BD). To detect the transient interaction between CPC and Shugoshin, we cross-linked cells with dithiobis(succinimidylpropionate) (DSP) (Thermo scientific). HeLa cells were synchronized by single thymidine block and release and then arrested at prometaphase by adding nocodazole. Prometaphase cells were further treated with nocodazole and MG-132 in the presence or absence of 100 μM roscovitine (Calbiochem) for 40 min. Cells were washed with PBS then treated with cross-link buffer (1 mM $MgCl_2$, 100 mM sucrose, 0.01% Triton-X100 and PhosSTOP (Roche) in PBS) containing 2 mM DSP for 40 min at 4 °C. The cross-link reaction was quenched by adding 1 M Tris-HCl (pH 7.5, final 50 mM) after which the cells were washed three times with PBS. The cells were then suspended in extraction buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM $MgCl_2$, 0.2% NP-40, 10% glycerol, 1 mM NaF, 1 mM Na-orthovanadate, 20 mM β-glycero-phosphate, 10 mM β-mercaptoethanol, 0.5 mM PMSF, complete protease inhibitors). After three rounds of freezing and thawing, the insoluble materials were suspended in extraction buffer containing 1% SDS and sonicated eight times (15 s each) on ice. After centrifugation, the cleared chromatin extracts were diluted tenfold in extraction buffer, and then rotated with anti-hBorealin (MBL) or mouse control IgG (SIGMA) for 1.5 h at 4 °C, followed by 1.5 h rotation with Protein A Sepharose at 4 °C. The beads were washed six times with extraction buffer. We analysed the immunoprecipitates by western blotting with anti-GFP (1:1,000), anti-hAurora B (1:1,000), anti-hINCENP (Cell Signaling, 1:100), anti-hSurvivin, anti-hBorealin (1:1,000), anti-hSgo1 (1:1,000), anti-hSgo2 (1:1,000), anti-Actin (1:1,000, Abcam) and anti-GAPDH (1:1,000, Millipore) antibodies. Total (cytoplasmic and chromatin) protein amounts of cyclin B and Securin were also analysed by western blotting using anti-cyclin B (1:1,000, Santa Cruz) and anti-Securin (1:1,000, Abcam) antibodies.

Roscovitine treatment of HeLa cells. HeLa cells were synchronized by single thymidine block and release and then arrested at prometaphase by adding nocodazole. Prometaphase cells were further treated with nocodazole and MG-132 in the presence or absence of 100 μM Roscovitine for 0, 40 or 80 min, and then fixed with paraformaldehyde and stained with anti-cyclin B (1:1,000, CHEMICON), anti-hAurora B and anti-ACA antibodies. The hAurora B signals in cells with condensed DNA and cyclin B signals were quantified.

Protein analysis of hBorealin. 293T cells were transfected with GFP-tagged hBorealin-WT or -7A, and extracted with RIPA buffer (50 mM Tris-HCl (pH7.5), 150 mM NaCl, 1 mM EDTA, 1.2% NP-40) following treatment with 330 nM nocodazole for 12 h. Wild-type extracts were treated with lambda-phosphatase as described above (we used 50 mM NaF as phosphatase inhibitor). Proteins were subjected to SDS–PAGE containing phos-tag acrylamide (40 μM final, NARD Institute)[35] and analysed by immunoblot using the anti-GFP antibody (1:1,000, Roche). To confirm the RNAi efficiency, crude lysates were analysed by immunoblot using anti-hBorealin (1:1,000), anti-hSgo1 (1:1,000) and anti-Tubulin (1:5,000, DM1A; Sigma) antibodies.

29. Bähler, J. *et al.* Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14,** 943–951 (1998).
30. Nonaka, N. *et al.* Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast. *Nature Cell Biol.* **4,** 89–93 (2002).
31. Rajagopalan, S. & Balasubramanian, M. K. *Schizosaccharomyces pombe* Bir1p, a nuclear protein that localizes to kinetochores and the spindle midzone, is essential for chromosome condensation and spindle elongation during mitosis. *Genetics* **160,** 445–456 (2002).
32. Hiraoka, Y., Toda, T. & Yanagida, M. The *NDA3* gene of fission yeast encodes β-tubulin: a cold-sensitive *nda3* mutation reversibly blocks spindle formation and chromosome movement in mitosis. *Cell* **39,** 349–358 (1984).
33. Yokobayashi, S. & Watanabe, Y. The kinetochore protein Moa1 enables cohesion-mediated monopolar attachment at meiosis I. *Cell* **123,** 803–817 (2005).
34. Klein, U. R., Nigg, E. A. & Gruneberg, U. Centromere targeting of the chromosomal passenger complex requires a ternary subcomplex of Borealin, Survivin, and the N-terminal domain of INCENP. *Mol. Biol. Cell* **17,** 2547–2558 (2006).
35. Kinoshita, E., Kinoshita-Kikuta, E., Takiyama, K. & Koike, T. Phosphate-binding tag, a new tool to visualize phosphorylated proteins. *Mol. Cell. Proteomics* **5,** 749–757 (2006).

doi:10.1038/nature09372

Direct visualization of secondary structures of F-actin by electron cryomicroscopy

Takashi Fujii[1], Atsuko H. Iwane[1], Toshio Yanagida[1] & Keiichi Namba[1]

F-actin is a helical assembly of actin, which is a component of muscle fibres essential for contraction and has a crucial role in numerous cellular processes, such as the formation of lamellipodia and filopodia[1,2], as the most abundant component and regulator of cytoskeletons by dynamic assembly and disassembly (from G-actin to F-actin and vice versa). Actin is a ubiquitous protein and is involved in important biological functions, but the definitive high-resolution structure of F-actin remains unknown. Although a recent atomic model well reproduced X-ray fibre diffraction intensity data from a highly oriented liquid-crystalline sol specimen[3], its refinement without experimental phase information has certain limitations. Direct visualization of the structure by electron cryomicroscopy, however, has been difficult because it is relatively thin and flexible. Here we report the F-actin structure at 6.6 Å resolution, made obtainable by recent advances in electron cryomicroscopy. The density map clearly resolves all the secondary structures of G-actin, such as α-helices, β-structures and loops, and makes unambiguous modelling and refinement possible. Complex domain motions that open the nucleotide-binding pocket on F-actin formation, specific D-loop and terminal conformations, and relatively tight axial but markedly loose interprotofilament interactions hydrophilic in nature are revealed in the F-actin model, and all seem to be important for dynamic functions of actin.

Many of the atomic models of F-actin proposed over the years[3–5] were obtained by maximizing the agreement between experimental intensities of X-ray fibre diffraction and those calculated from a model. However, because the diffraction patterns are cylindrically averaged and layer lines are broadened in an arc owing to the finite disorientation, it is extremely difficult to obtain individual Fourier–Bessel components of the structure factors for three-dimensional (3D) density reconstruction. That is why these analyses have been done by building many models, calculating layer-line intensities and maximizing the agreement between observed and calculated intensities to find a best possible model, except in one case with tobacco mosaic virus[6]. Therefore, there is always some concern as to whether or not such models are unique.

Technical advances in the electron cryomicroscopy (cryoEM) and image analysis of frozen hydrated specimens in recent years have allowed the structural analysis of helical assemblies of biological macromolecules, such as the bacterial flagellar filament, at near-atomic resolution by aligning and averaging tens of thousands of molecular images using helical symmetry[7–9]. CryoEM image analyses of F-actin have also been carried out but only up to 13 Å resolution[3,5,10], so the atomic model is of limited accuracy in studying the mechanisms of actin polymerization and depolymerization. Because F-actin is a flexible, ribbon-like filament with a diameter of 100 Å and a mass-per-length of 1.5 kDa $Å^{-1}$, the image contrast of frozen hydrated specimens is markedly lower than that of thicker tubular structures of the flagellar filament (230 Å, 11.0 kDa $Å^{-1}$) and tobacco mosaic virus (180 Å, 12.6 kDa $Å^{-1}$), making accurate image alignment and high-resolution analysis extremely difficult. However, recent technical advances[11] have allowed us to obtain a 3D density map of skeletal muscle F-actin at 6.6 Å resolution.

Key to this is our use of an in-column Ω-type energy filter and charge-couple-device (CCD) camera as well as a field-emission gun and a liquid-helium-cooled specimen stage. A remarkable ~5-fold gain in contrast has been achieved by energy filtering, controlling ice thickness and using a specimen temperature of 50 K instead of 4 K (ref. 11). We can now see the two-stranded helical features of F-actin even at small defocus levels (Fig. 1a). The CCD camera made the collections of high-quality images far more efficient, but to avoid undesirable removal of high-resolution contrast by its poor modulation transfer function[12] we used a magnification of 172,000 × (0.87 Å per pixel). We collected 490 such high-quality images in two days. We used a single-particle image analysis method using helical symmetry[11,13]. Because we fully automated the whole procedure including corrections for the contrast transfer function, the image analysis was completed within two days. The helical symmetry and the axial repeat distance were refined in the iterative process of image analysis and converged to a subunit rotation of −166.6°, corresponding to a helical symmetry of between 28/13 and 13/6 (subunits/turn), and an axial repeat of 27.6 Å. The resolution was 6.6 Å at a Fourier shell correlation of 0.143 (ref. 14; Supplementary Figs 1 and 2). Layer lines are visible out to 6.8 Å in the power spectrum of the 3D reconstruction (Fig. 1b; see also Supplementary Fig. 3). The variance map shows the reliability of the density map as well as the rigid helical backbone of the filament (Supplementary Fig. 4).

Because the resolution was better than that needed to identify individual secondary structures of actin, including loops and an extended amino-terminal chain previously unresolved in the crystal structure (Fig. 2 and Supplementary Movie 1), we were able to build a highly reliable atomic model of F-actin. The situation is analogous to protein crystallography at 3 Å resolution, where side chains can be identified and used to build a reliable atomic model. Although the flexibility of F-actin, especially in its helical order, has been debated over the years[10], the high-resolution map obtained by using over 97% of the collected images indicates that F-actin is not so flexible. Multireference alignment with reference 3D volumes of F-actin with different helical symmetries produced a narrow angular distribution (Fig. 1c), indicating that the twisting variability by thermal motion is less than ~1°. This also demonstrates the importance of careful cryospecimen preparation with an automated vitrification device (Methods).

To build a reliable atomic model of F-actin, we used FlexEM[15], which refines the atomic model by fitting it into the electron microscopy map by simulated annealing molecular dynamics with stereochemical and non-bonded interaction terms restrained. We used the crystal structure of uncomplexed actin[16] (Protein Data Bank ID, 1J6Z) as the model representing G-actin and divided it into four domains (Fig. 3), which have been conventionally named subdomains[17]. We call them domains because domains 1, 3 and 4 have well-defined hydrophobic cores and behave as rigid, independent units on conformational change from G-actin[16] to F-actin (Fig. 3a), as indicated by the small root-mean-squared displacements of Cα atoms (Supplementary Table 1). This in turn ensures the correctness and reliability of the present F-actin model. The model and the refinement process are shown in Fig. 2 and Supplementary Movie 2, respectively.

[1]Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamadaoka, Suita, Osaka 565-0871, Japan.

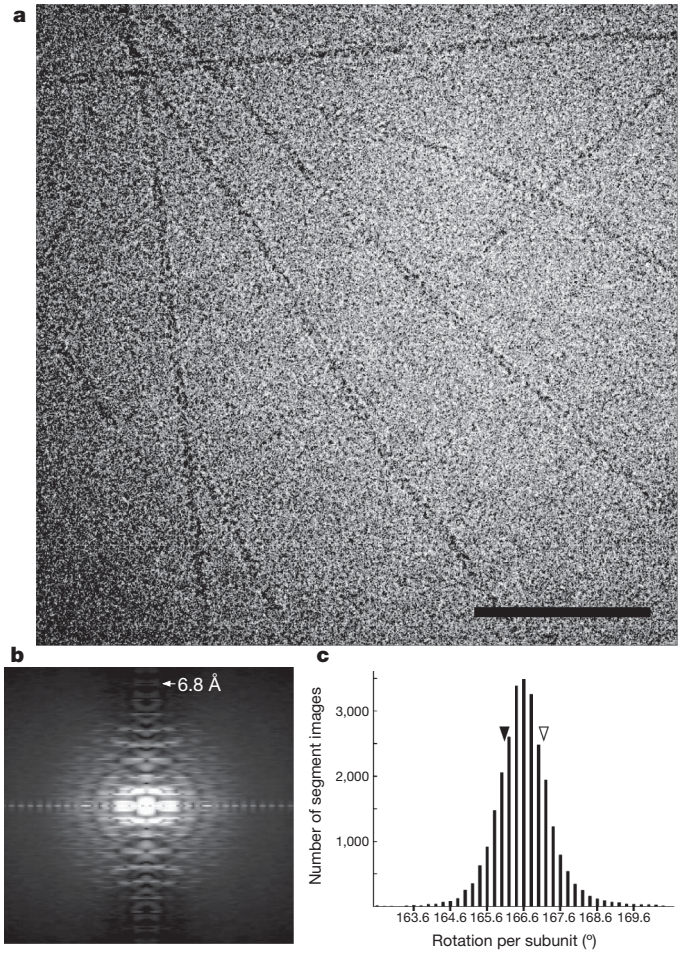


Figure 1 | High-contrast cryoEM image of F-actin, its power spectrum and angular distribution of rotation per subunit. **a**, A CCD frame of frozen hydrated F-actin filaments embedded in vitreous ice film recorded by zero-loss energy-filtered microscopy. The defocus value is 1,500 nm. Scale bar, 100 nm. **b**, The power spectrum of the 3D reconstruction without imposing helical symmetry. The layer line is visible out to 6.8 Å. **c**, Histogram showing the distribution of the rotation angle per subunit at every 0.2° along the 1-start helix of F-actin by multireference alignment. The peak is at 166.6°. The 13/6 helix of F-actin corresponds to 166.15° (black arrowhead), and the 28/13 helix corresponds to 167.14° (white arrowhead). Eighty-seven per cent of all image segments were classified within ±1°. The angular distribution clearly shows that the twisting variability of F-actin by thermal motion is less than ~1°.

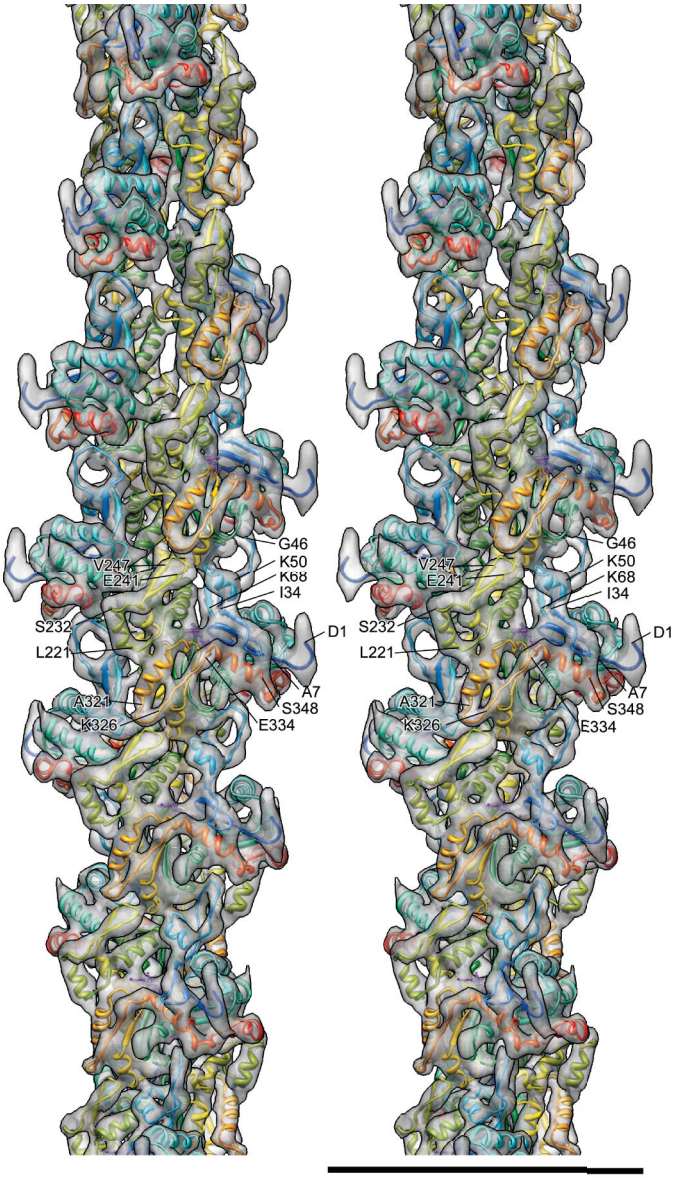


Figure 2 | Three-dimensional density map of F-actin with a fitted atomic model in stereo. The atomic model is presented as a Cα ribbon diagram coloured with a rainbow gradient from the amino terminus in blue to the carboxy terminus in red. Approximately fifteen subunits of actin are shown, to display all the possible azimuthal orientations. D1, Asp 1. Scale bar, 100 Å.

Domain 2 is the only exception. The two-turn α-helix (residues 40–48) at the tip of the D-loop (the DNase-I-binding loop[17]) of G-actin[16] becomes an extended loop (38–53) (Fig. 3a, upper right). Although such a change has been predicted from its variable conformation in crystal structures and its possible involvement in the axial intersubunit interactions, this D-loop conformation is unique, indicating that it is totally dependent on the molecule to which it binds (Supplementary Fig. 5).

Although the overall conformation of F-actin is similar to the previous ones[3,5], our model shows distinct features in local conformation, such as loops and terminal chains (Supplementary Fig. 6). Even the positions of main-chain atoms are distinctly different, as indicated by the root-mean-squared deviations in the positions of corresponding Cα atoms (Supplementary Fig. 7 and Supplementary Table 2). Because the nature of the conformational change from G-actin to F-actin is of immense importance for the biological implications of actin function, we carefully compared the F-actin structure with that of G-actin[16]. Whereas the two major domains (domains 1–2 and 3–4) are twisted in G-actin, they become flat in F-actin. However, the relative domain motions are more complex than those described previously[3]. When domain 1 is superposed, domain 2 is tilted towards domain 4 by 10° and towards the outside of F-actin (out of the page in Fig. 3b) by 20° around the red 'spear' (Fig. 3b, c, red arrow). Together with the conformational change in the D-loop, the slightly bent domain 1–2 in G-actin becomes significantly flatter in F-actin (Fig. 3c). When domain 3 is superposed, domain 4 is tilted towards the inside of F-actin (into the page in Fig. 3d) by 7° around the blue spear (Fig. 3d, e, blue arrows), making the slightly bent domain 3–4 in G-actin flatter in F-actin. At the same time, domain 1 is rotated relative to domain 3 by 13° (Fig. 3f, green arrows) around the green spear passing diagonally through domain 4 to the outside of domain 1 (Fig. 3d, f).

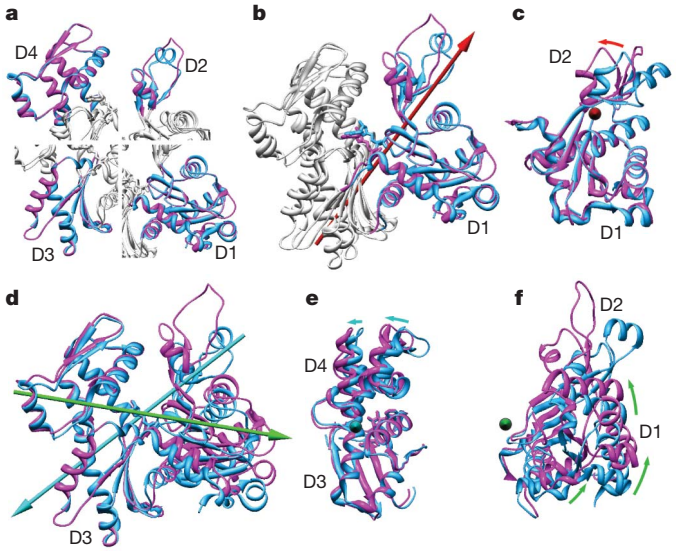


Figure 3 | Comparison of actin models in F-actin with the uncomplexed crystal structure[16]**.** The Cα ribbon diagram in magenta is the atomic model in F-actin, and that in blue is the crystal structure. **a**, Superposition of the two structures for each domain. The Cα atoms of residues 34–37 and 54–68 were used for superposition of domain 2. **b**, **c**, Domain 1 (D1) is superposed to compare domain 2 (D2) in two different views. Only domains 1 and 2 are coloured in **b** and shown in **c**. The solid red spear in **b** and **c** is the axis of rotation of domain 2. **c** is viewed against the direction of the spear. **d**–**f**, Domain 3 (D3) is superposed to compare domains 4 (D4) and 1 (D1) in two different views. The spears in **d**–**f** are the axes of rotation: blue for domain 4; green for domain 1. **e** and **f** are viewed against the directions of the blue and green spears, respectively. Short arrows in red in **c**, blue in **e** and green in **f** indicate the motions of domain 2 relative to 1, domain 4 relative to 3 and domain 1 relative to 3, respectively. The rotation axes were obtained using DYNDOM[30].

Although we cannot discuss side-chain conformations and their interactions in detail owing to the limited resolution, reasonably accurate main-chain positions allow us to look into possible side-chain interactions within and between actin molecules. The domain-1 rotation described above also involves a 5° anticlockwise rotation around the nucleotide within the plain of actin to open the nucleotide-binding pocket as well as the gap between domains 1 and 3 (Fig. 3d). This rotation causes Gln 137 to move closer to the position of the γ-phosphate. Because actin ATPase is activated on F-actin formation[18], and because Gln 137 is crucial in ATP hydrolysis[19], our model offers a possible explanation for the ATPase activation mechanism.

Axial interactions along the protofilament are tight (Figs 2 and 4a, b). Subunits in Fig. 4 are numbered −2, −1, 0, 1 and 2 along the 1-start helix from the barbed-end side. Domains and residues mentioned hereafter will be those of subunit 0 unless otherwise specified. In each pair of interacting residues described below, the first residue is always of subunit 0. Axial interactions are extensive but not to the extent of the previous model[3]. The difference is partly due to the distinct D-loop conformation; residues Gly 46–Leu 50 of the D-loop are not associated with subunit 2 (Supplementary Fig. 6). However, the D-loop still extends into the bottom pocket formed by domains 1 and 3 of subunit 2: Val 45–Tyr 143 in the interaction with domain 1 and His 40–Tyr 169 with domain 3. The remaining part of domain 2 also forms a few bonds with domain 3 of subunit 2: Ile 64–Tyr 166, Lys 61–Glu 167 and Arg 62–Asp 288. Domain 4 is also in close contact with domain 3 of subunit 2, forming a few bonds: Glu 205–Asp 286, Asp 244–Arg 290, Asp 241–Thr 324 and Gly 245–Pro 322. Therefore, the axial interactions are mostly electrostatic and hydrophilic, although there are some hydrophobic interactions.

In contrast, interprotofilament interactions are surprisingly modest, with only two points of contact (Figs 2 and 4c–f). One is between the plug (Gln 263–Gly 273) in domain 3 and the beginning of the D-loop of subunit −1: Glu 270–Arg 39, Ser 265–His 40 and Gly 268–His 40. The main-chain oxygen of Ile 267 at the top of this plug may also form a bond with His 173 in domain 3 of subunit 1. This plug has been called 'hydrophobic plug' because it was modified to an extended hairpin for hydrophobic interactions with the opposite strand[4]. However, the plug conformation is preserved in F-actin (Fig. 2, two actin subunits on the top right), and because the plug forms a salt bridge and hydrophilic interactions with subunits −1 and 1 to stabilize the two-stranded F-actin structure (Fig. 4d), it should be called 'hydrophilic plug'.

The other interprotofilament contact is just above the plug, where domain 4 is in contact with domains 1 and 3 of subunit 1 (Fig. 4c–f). Lys 191 and Thr 194 in domain 4 are in close proximity to Leu 110 in domain 1 and Arg 177 in domain 3 of subunit 1 (Fig. 4f). Glu 195 in domain 4 is located where it might form a salt bridge with Lys 113 in domain 1 of subunit 1 (Fig. 4f). The main-chain oxygen atoms of Thr 194 and Glu 195 in domain 4 seem to form hydrogen bonds with the main-chain oxygen of Leu 110 of subunit 1, either directly or indirectly through water molecules, because a strong density is observed here.

Taken together, the interactions between actin subunits are mostly electrostatic or hydrophilic. This explains F-actin depolymerization in concentrated salt solutions[20]. However, F-actin is not fragile, because a number of electrostatic and hydrophilic bonds can make a protein stable in a similar manner to hydrophobic ones.

In the previous model, the concatenation of salt bridges between the major domains was thought to be responsible for stabilizing the flat actin conformation[3], but our model does not show such extensive salt bridges. However, the disruption of hydrophobic interactions between domain 1 (Pro 109 and Leu 110) and domain 3 (Val 163 and Ile 175) by the rotation of domain 1 relative to domain 3 by 20° (Fig. 3d) is consistently observed, although this rotation is nearly within the plane of actin (Fig. 3f) rather than untwisting the plane (Fig. 3d, f). The disruption of the hydrophobic interactions puts actin into a higher-energy conformation, but it is possibly balanced by the interactions between actin subunits to make F-actin a dynamic molecular assembly with polymerization/depolymerization for cellular activities. The complex domain motions are likely to be key to the asymmetric association/dissociation rate of ATP–actin for steady-state treadmill-like dynamics[21].

The N-terminal chain conformation is distinct from any previous models of actin (Fig. 2). Because the negatively charged N-terminal residues (Asp-Glu-Asp-Glu) are important for interactions with many actin-binding proteins, including myosin for muscle contraction[22], this conformation must be important in the dynamic functions. The helical symmetry of F-actin changes when it binding other proteins, such as scruin[23] and cofilin[24], suggesting that the axial twist can be modified easily by binding a protein over two neighbouring actin subunits, even though the twist is largely unchanged by thermal motion.

Although cryoEM image analysis has advanced enough to allow the visualization of side chains for icosahedral viruses[25–27] and chaperonins[28,29], their large particle sizes that give rise to high-contrast images are still essential for accurate image alignment to achieve high-resolution analysis. It is therefore thought that high-resolution analysis of small or thin objects, such as F-actin, is extremely difficult owing to their intrinsically low image contrast. We have demonstrated that our cryoEM technologies now allow us to visualize the secondary structures of such thin objects as F-actin in a few days of work, including image data collection and processing and 3D image reconstruction. It would be easy to determine the structure of F-actin in complex with actin-binding proteins, such as myosin, troponin and tropomyosin, in similar detail, and such structures will bring us deep insights into the mechanisms of muscle contraction and regulation, cellular motility and morphogenesis. There is also room for further improvement, to reach atomic resolution. The present work offers a new opportunity to look into cellular mechanisms essential for the activities of life.

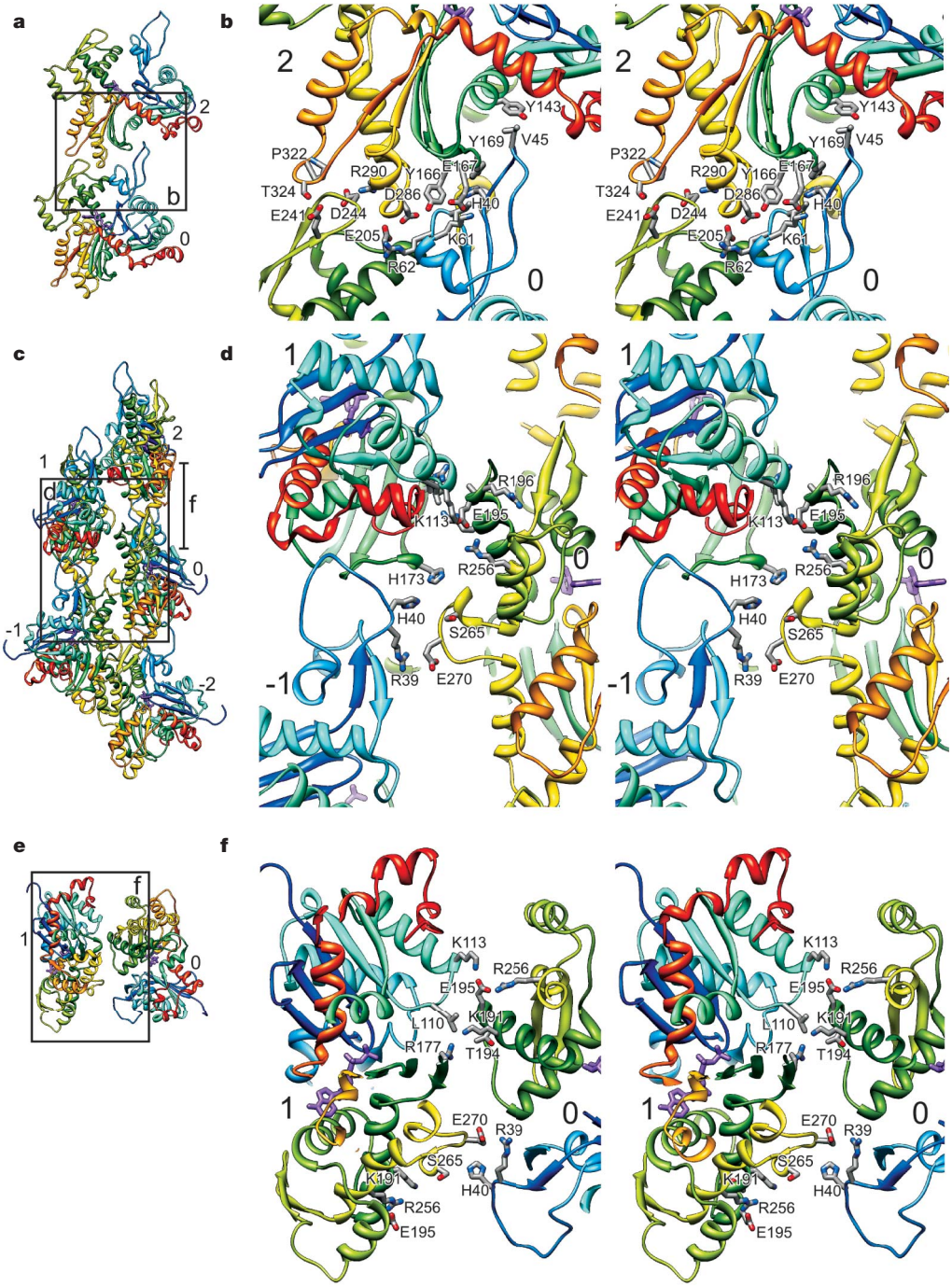


Figure 4 | Axial and lateral interactions of actin subunits in F-actin. **a**, **b**, Axial interactions in side view. **c**, **d**, Interprotofilament interactions in side view. **e**, **f**, Interprotofilament interactions in end-on view from the bottom (barbed-end side). **a**, **c** and **e** are guides to show the portions displayed in stereo in **b**, **d** and **f**, respectively. Only subunits 0 and 1 are shown in **e** and **f**. The models are shown as Cα ribbon diagram coloured in a rainbow gradient as in Fig. 2 with relevant side chains in stick representation and coloured according to the element: carbon, grey; nitrogen, blue; oxygen, red. The side-chain conformation of Lys 113, His 173 and Glu 195 was manually modified to indicate their potential involvement in the interprotofilament interactions.

METHODS SUMMARY

We polymerized 7.5 μM G-actin from rabbit skeletal muscle in a 30-μl solution of 25 mM Hepes buffer (pH 7.5), 50 mM KCl, 1 mM $MgCl_2$ and 1 mM ATP for ~2 h at 25 °C. The F-actin filaments were spun down by centrifugation at 100,000*g* for 60 min to remove monomeric actin. The pellet was gently resuspended in a 30-μl polymerization buffer, and a 2.1-μl aliquot was applied onto an electron microscopy grid, blotted and rapidly frozen by Vitrobot (FEI), and observed at temperatures of 50–60 K using a JEOL JEM3200FSC electron cryomicroscope equipped with a liquid-helium-cooled specimen stage, an Ω-type energy filter and a field-emission electron gun operated at 200 kV. We recorded zero-energy-loss images, with a slit setting to remove electrons with an energy loss larger than 10 eV, on a 4,000 × 4,000 15-μm-per-pixel slow-scan CCD camera (TemCam-F415MP, TVIPS) at a magnification of ×172,414, a defocus range of 1.0–2.5 μm and an electron dose of ~20 electrons per square ångström. Image processing and analysis was done using the iterative helical real-space refinement method[11,13].

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

Received 12 January; accepted 21 July 2010.
Published online 15 September 2010.

1. Pollard, T. D. & Borisy, G. G. Cellular motility driven by assembly and disassembly of actin filaments. *Cell* **112,** 453–465 (2003).
2. Carlier, M. F. & Pantaloni, D. Control of actin assembly dynamics in cell motility. *J. Biol. Chem.* **282,** 23005–23009 (2007).
3. Oda, T., Iwasa, M., Aihara, T., Maeda, Y. & Narita, A. The nature of the globular- to fibrous-actin transition. *Nature* **457,** 441–445 (2009).
4. Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. Atomic model of the actin filament. *Nature* **347,** 44–49 (1990).
5. Holmes, K. C. *et al.* Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide. *Nature* **425,** 423–427 (2003).
6. Namba, K. & Stubbs, G. Structure of tobacco mosaic virus at 3.6 A resolution: implications for assembly. *Science* **231,** 1401–1406 (1986).
7. Yonekura, K., Maki-Yonekura, S. & Namba, K. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **424,** 643–650 (2003).
8. Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423,** 949–955 (2003).
9. Sachse, C. *et al.* High-resolution electron microscopy of helical specimens: a fresh look at tobacco mosaic virus. *J. Mol. Biol.* **371,** 812–835 (2007).
10. Galkin, V. E., Orlova, A., Cherepanova, O., Lebart, M. C. & Egelman, E. H. High-resolution cryo-EM structure of the F-actin-fimbrin/plastin ABD2 complex. *Proc. Natl Acad. Sci. USA* **105,** 1494–1498 (2008).
11. Fujii, T., Kato, T. & Namba, K. Specific arrangement of α-helical coiled coils in the core domain of the bacterial flagellar hook for the universal joint function. *Structure* **17,** 1485–1493 (2009).
12. Sander, B., Golas, M. M. & Stark, H. Advantages of CCD detectors for de novo three-dimensional structure determination in single-particle electron microscopy. *J. Struct. Biol.* **151,** 92–105 (2005).
13. Egelman, E. H. A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy* **85,** 225–234 (2000).
14. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333,** 721–745 (2003).
15. Topf, M. *et al.* Protein structure fitting and refinement guided by cryo-EM density. *Structure* **16,** 295–307 (2008).
16. Otterbein, L. R., Graceffa, P. & Dominguez, R. The crystal structure of uncomplexed actin in the ADP state. *Science* **293,** 708–711 (2001).
17. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. Atomic model of the actin:DNase I complex. *Nature* **347,** 37–44 (1990).
18. Pollard, T. D. Regulation of actin filament assembly by Arp2/3 complex and formins. *Annu. Rev. Biophys. Biomol. Struct.* **36,** 451–477 (2007).
19. Iwasa, M. *et al.* Dual roles of Q137 of actin revealed by recombinant human cardiac muscle alpha-actin mutants. *J. Biol. Chem.* **283,** 21045–21053 (2008).
20. Béla, N. & Jencks, W. P. Depolymerization of F-actin by concentrated solutions of salts and denaturing agents. *J. Am. Chem. Soc.* **87,** 2480–2488 (1965).
21. Fujiwara, I., Vavylonis, D. & Pollard, T. D. Polymerization kinetics of ADP- and ADP-Pi-actin determined by fluorescence microscopy. *Proc. Natl Acad. Sci. USA* **104,** 8827–8832 (2007).
22. Sutoh, K. Identification of myosin-binding sites on the actin sequence. *Biochemistry* **21,** 3654–3661 (1982).
23. Schmid, M. F., Sherman, M. B., Matsudaira, P. & Chiu, W. Structure of the acrosomal bundle. *Nature* **431,** 104–107 (2004).
24. McGough, A., Pope, B., Chiu, W. & Weeds, A. Cofilin changes the twist of F-actin: implications for actin filament dynamics and cellular function. *J. Cell Biol.* **138,** 771–781 (1997).
25. Yu, X., Jin, L. & Zhou, Z. H. 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature* **453,** 415–419 (2008).
26. Zhang, X. *et al.* Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc. Natl Acad. Sci. USA* **105,** 1867–1872 (2008).
27. Zhang, X., Jin, L., Fang, Q., Hui, W. H. & Zhou, Z. H. 3.3 A cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell* **141,** 472–482 (2010).
28. Cong, Y. *et al.* 4.0-Å resolution cryo-EM structure of the mammalian chaperonin TRiC/CCT reveals its unique subunit arrangement. *Proc. Natl Acad. Sci. USA* **107,** 4967–4972 (2010).
29. Zhang, J. *et al.* Mechanism of folding chamber closure in a group II chaperonin. *Nature* **463,** 379–383 (2010).
30. Hayward, S. & Berendsen, H. J. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins* **30,** 144–154 (1998).

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

Acknowledgements We thank T. Kato for setting up and managing our cryoEM and computing facilities for high-throughput, high-resolution image analysis, and I. Tokita, M. Urabe and JEOL for maintaining the electron cryomicroscope in the best conditions. We also thank E. Egelman for his help in the use of the iterative helical real-space refinement method in the early stages of this study, and F. Oosawa, S. Asakura, H. Hotani and D. L. D. Caspar for their discussions, continuous support and encouragement. T.F. was a research fellow of the Japan Society for the Promotion of Science. This work was supported in part by Grants-in-Aid for Scientific Research to K.N. (16087207 and 21227006) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Author Contributions T.F. made various improvements to the cryoEM method and performed all the cryoEM experiments, the image analysis and the model-building of F-actin. A.H.I. prepared G-actin. T.Y. and K.N. planned the project, and K.N. supervised the project. T.F. and K.N. wrote the paper on the basis of discussions with A.H.I. and T.Y.

Author Information The reconstructed density map has been deposited in the Electron Microscopy Data Bank under accession code EMD-5168, and the atomic coordinates have been deposited in the Protein Data Bank under accession code 3MFP. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.N. (keiichi@fbs.osaka-u.ac.jp).

METHODS

Sample preparation and electron microscopy. G-actin (7.5 μM) from rabbit skeletal muscle was polymerized in a 30-μl solution of 25 mM Hepes buffer (pH 7.5), 50 mM KCl, 1 mM $MgCl_2$ and 1 mM ATP for ~2 h at 25 °C. The F-actin filaments were spun down by centrifugation at 100,000*g* for 60 min to remove monomeric actin. The pellet was gently resuspended in a 30-μl polymerization buffer.

A 2.1-μl sample solution was applied onto a Quantifoil holey carbon molybdenum grid (R0.6/1.0, Quantifoil) and was plunge-frozen into liquid ethane using a fully automated vitrification device (Vitrobot, FEI). The control of temperature, 100% humidity and the timing between blotting and plunging is important to make ice-embedded F-actin as straight as possible and ice thickness as low as possible for high-contrast, high-quality imaging. The specimen was observed at temperatures of 50–60 K using a JEOL JEM3200FSC electron microscope, which is equipped with a liquid-helium-cooled specimen stage, an Ω-type energy filter and a field-emission electron gun operated at 200 kV. Zero-energy-loss images, with a slit setting to remove electrons of an energy loss larger than 10 eV, were recorded on a 4,000 × 4,000 15-μm-per-pixel slow-scan CCD camera (TemCam-F415MP, TVIPS) at a magnification of around ×172,000, a defocus range of 1.0–2.5 μm and an electron dose of ~20 electrons per square ångström. The magnification was calibrated by measuring the layer-line spacing of 23.0 Å in the Fourier transform of images of tobacco mosaic virus mixed in the sample solution. The image pixels of 0.87 Å per pixel at this magnification were twofold binned to have 1.74 Å per pixel for image analysis, to increase the signal-to-noise ratio. In total, 490 CCD images were collected manually in two days and used for image analysis.

Image analysis. Helical image analysis was carried out by using the iterative helical real-space reconstruction method[13] with EMAN 1.9[31] and SPIDER 15.06[32], as previously described[11]. Defocus and astigmatism of each image were determined using a slightly modified version of CTFFIND3[33]. Images of F-actin in 490 CCD frames were boxed into 27,240 segments of 700 × 700 pixels with a step shift of 70 pixels along the filament axis using EMAN's boxer program[31]. The number of actin molecules used for this reconstruction corresponds to ~120,000. The in-plane orientation of F-actin in each segment was recorded in a list to avoid interpolation when rotating the image. Images were then corrected for a phase and amplitude contrast transfer function (CTF) by multiplying the CTF calculated from the defocus and astigmatism. We used an amplitude CTF/phase CTF ratio of 7% (ref. 34). This procedure for CTF correction results in the multiplication of the original structure factor by the square of the CTF (CTF^2) and suppresses the noise around the nodes of the CTF, allowing more accurate image alignment. The amplitude modification by CTF^2 was corrected in the last stage of image analysis as described later. The images were then high-pass-filtered (285 Å) to remove a density undulation of low spatial frequency, normalized, decimated by a factor of two and cropped to 320 × 320 pixels. Image processing was mainly carried out with the SPIDER package[32] on a PC cluster computer with 40 CPUs (RC server Calm2000, Real Computing).

A series of reference projection images were generated for each reference volume by rotating the volume azimuthally about the filament axis between 0° and 360° and projecting the volume at every 1° to produce all the views. The variation of the out-of-plane tilt angle was limited to ±10° and was also sampled at every 1°. The raw images of the boxed F-actin segments were translationally and rotationally aligned and cross-correlated with the set of reference projections to produce the following information: an in-plane rotation angle, an *x* shift, a *y* shift, an azimuthal angle and a cross-correlation coefficient for each segment. Particles with a small cross-correlation coefficient were discarded. The polarity of the particles was tracked with respect to their respective filaments. Even with our high-contrast imaging technique, the orientation of each individual particle was sometime ambiguous owing to the relatively low contrast and high noise level of the segment images. Therefore, the orientation was defined as that of the majority of the segments for each filament during each alignment cycle, and all the segments identified to have the opposite orientation were discarded. On average, 95% of the segments from an F-actin filament showed the same polarity, indicating the reliability of polarity determination. A 3D image was then reconstructed by back-projection.

We used a solid cylinder with a diameter of 70 Å as the initial reference volume to avoid any model bias in image alignment and reconstruction. The initial helical symmetry parameters were imposed on the first reconstruction to produce the new reference volume for the second round of image alignment. After this cycle, every time a 3D image was reconstructed, the symmetry of this new volume was determined by a least-squares fitting algorithm, and this symmetry was imposed on the reconstruction[13]. The new symmetry-enforced volume was used as a reference for the next round of alignment. This process was repeated iteratively until the symmetry values converged to a stable solution. In the refinement phase, to reduce the noise in the reference and increase the accuracy of aliment, we used automask2 from the EMAN package to impose the edge smooth mask to the reconstruction.

The resulting reconstruction was then modified by multiplying the transform of the reconstruction by $1/[\sum CTF^2 + 1/SNR]$ (SNR, signal-to-noise ratio) to compensate for the amplitude distortion by the CTF. The map was scaled with a B factor of $-180\,Å^2$. More than 97% of the image data were used for the final reconstruction presented.

Model refinement. We used the crystal structure of uncomplexed actin[16] (PDB ID, 1J6Z) as an initial model and divided it into four domains. In the first stage of the fitting process, we treated the four domains as rigid bodies and allowed the joints of these domains to be flexible. But residues 1–8, 39–56, 221–234 and 337–375 were clearly outside the density map. In the second stage, we allowed these residues to move flexibly to fit into the map under stereochemical restraints. We then applied the helical symmetry to this subunit model to build an F-actin model and minimized the conformational energy further using FlexEM to remove intermolecular clashes of atoms. All the figures were made using UCSF Chimera[35].

Multireference alignment. On the basis of the refined atomic model of F-actin, we built 41 F-actin models with different rotation angles per subunit along the 1-start helix, from 162.6° to 170.6° at every 0.2°. We converted these models to 3D volumes at 8.0 Å resolution using pdb2mrc[31] and used them as the reference volumes to carry out multireference alignment for the 27,240 overlapping segment images that we used for 3D image reconstruction. It should be noted that the relatively high level of random noise in cryoEM images due to low electron dose tends to cause the misalignment and, thereby, broadening of the angular distribution. Therefore, the actual distribution should be narrower than the one shown in the histogram of Fig. 1c.

31. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstruction. *J. Struct. Biol.* **128,** 82–97 (1999).
32. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).
33. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142,** 334–347 (2003).
34. Yonekura, K., Braunfeld, M. B., Maki-Yonekura, S. & Agard, D. A. Electron energy filtering significantly improves amplitude contrast of frozen-hydrated protein at 300kV. *J. Struct. Biol.* **156,** 524–536 (2006).
35. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

LETTER

doi:10.1038/nature09391

Crystal structure of the human symplekin–Ssu72–CTD phosphopeptide complex

Kehui Xiang[1], Takashi Nagaike[1]*, Song Xiang[1]*†, Turgay Kilic[1], Maia M. Beh[1], James L. Manley[1] & Liang Tong[1]

Symplekin (Pta1 in yeast) is a scaffold in the large protein complex that is required for 3′-end cleavage and polyadenylation of eukaryotic messenger RNA precursors (pre-mRNAs)[1–4]; it also participates in transcription initiation and termination by RNA polymerase II (Pol II)[5,6]. Symplekin mediates interactions between many different proteins in this machinery[1,2,7–9], although the molecular basis for its function is not known. Here we report the crystal structure at 2.4 Å resolution of the amino-terminal domain (residues 30–340) of human symplekin in a ternary complex with the Pol II carboxy-terminal domain (CTD) Ser 5 phosphatase Ssu72 (refs 7, 10–17) and a CTD Ser 5 phosphopeptide. The N-terminal domain of symplekin has the ARM or HEAT fold, with seven pairs of anti-parallel α-helices arranged in the shape of an arc. The structure of Ssu72 has some similarity to that of low-molecular-mass phosphotyrosine protein phosphatase[18,19], although Ssu72 has a unique active-site landscape as well as extra structural features at the C terminus that are important for interaction with symplekin. Ssu72 is bound to the concave face of symplekin, and engineered mutations in this interface can abolish interactions between the two proteins. The CTD peptide is bound in the active site of Ssu72, with the pSer 5-Pro 6 peptide bond in the *cis* configuration, which contrasts with all other known CTD peptide conformations[20,21]. Although the active site of Ssu72 is about 25 Å from the interface with symplekin, we found that the symplekin N-terminal domain stimulates Ssu72 CTD phosphatase activity *in vitro*. Furthermore, the N-terminal domain of symplekin inhibits polyadenylation *in vitro*, but only when coupled to transcription. Because catalytically active Ssu72 overcomes this inhibition, our results show a role for mammalian Ssu72 in transcription-coupled pre-mRNA 3′-end processing.

Human symplekin contains 1,274 amino-acid residues (Fig. 1a) and its sequence is well conserved in higher eukaryotes (Supplementary Fig. 1). In comparison, symplekin shares only weak sequence similarity with yeast Pta1 (ref. 1) (Supplementary Fig. 2), and Pta1 lacks the C-terminal 500 residues of symplekin (Fig. 1a). Symplekin and Pta1 do not have any recognizable homology with other proteins. Predictions of secondary structure suggest the presence of an all-helical segment in the N-terminal region of symplekin and Pta1 (Fig. 1a and Supplementary Figs 1 and 2). Recent studies in yeast suggested that the N-terminal segment of Pta1 is important for interaction with Ssu72 (ref. 9). Ssu72 is required for pre-mRNA 3′-end cleavage in yeast[7], although its phosphatase activity is not necessary for this function[13]. The catalytic activity of Ssu72 may instead be important for Pol II transcription and termination and for gene looping[17]. Ssu72 is highly conserved in the eukaryotes (Supplementary Fig. 3), but so far no evidence exists implicating mammalian Ssu72 in 3′-end processing.

To determine the structure of a symplekin–Ssu72–CTD phosphopeptide ternary complex, residues 30–360 of human symplekin and full-length human Ssu72 were overexpressed and purified separately. The two proteins were mixed, with Ssu72 in slight molar excess, and the symplekin–Ssu72 complex was purified by gel-filtration chromatography. This procedure also demonstrated strong interactions between the two human proteins, consistent with observations on their yeast counterparts[9]. The decamer CTD phosphopeptide used in this study, Ser-Tyr 1-Ser 2-Pro 3-Thr 4-pSer 5-Pro 6-Ser 7-Tyr-Ser, where Ser 5 is phosphorylated, contained an entire CTD heptad repeat as well as a serine residue from the previous repeat and Tyr-Ser from the following repeat. To prevent hydrolysis, the active-site nucleophile Cys 12 of Ssu72 was mutated to Ser in the ternary complex. We have also determined the crystal structure of the symplekin–Ssu72(wild-type) binary complex and the structures of the symplekin N-terminal domain alone (for residues 30–395 or 1–395). All the structures are in excellent agreement with the crystallographic data and the expected geometric parameters (Supplementary Table 1).

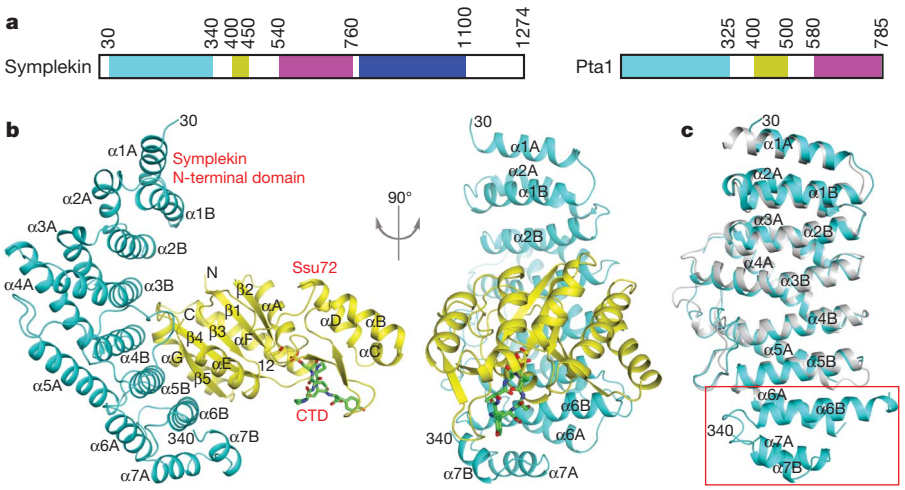


Figure 1 | Structure of the human symplekin–Ssu72–CTD phosphopeptide ternary complex. **a**, Domain organization of human symplekin and *Saccharomyces cerevisiae* Pta1. The domains are suggested by secondary-structure predictions, and the middle region of Pta1 is suggested by functional studies[9]. **b**, Schematic drawing of the structure of human symplekin–Ssu72–CTD phosphopeptide ternary complex, in two views. The N-terminal domain of symplekin is shown in cyan, and Ssu72 in yellow. The CTD phosphopeptide is shown as a stick model (in green for carbon atoms). **c**, Overlay of the structures of the N-terminal domain of human (in cyan) and *Drosophila* (in grey) symplekin[22]. *Drosophila* symplekin lacks the last two pairs of helical repeats (boxed). All structure figures were produced with PyMOL (http://www.pymol.org).

[1]Department of Biological Sciences, Columbia University, New York, New York 10027, USA. †Present address: Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China.
*These authors contributed equally to this work.

The structures show that residues 30–340 of symplekin (Symp-N) form seven pairs of antiparallel α-helices, whereas residues 1–29 and 341–395 are disordered (Supplementary Fig. 1). The pairs of helices are arranged in an arc, with the first helix in each pair, the αA helix, being located on the convex face of the arc, and the αB helix on the concave face (Fig. 1b). Most of the loops connecting the helices are short, except for that linking helices α4B and α5A, with 31 residues (Fig. 1b). The overall fold of Symp-N is found in many other proteins, including those with the ARM or HEAT repeats. These structures are often involved in protein–protein interactions, which is consistent with the proposed scaffold function of symplekin.

The structure of the N-terminal domain of *Drosophila* symplekin (residues 22–270) was reported recently[22]. Its overall conformation is similar to that of human Symp-N, with a root mean squared distance of 1.0 Å among their equivalent Cα atoms, although the *Drosophila* symplekin structure lacks the two pairs of helices (α6 and α7) at the C terminus (Fig. 1c). Noting the good sequence conservation of residues in this region (Supplementary Fig. 1), it is likely that these helices are also present in the N-terminal domain of *Drosophila* symplekin. Our studies show that helix α6B is important for interactions with Ssu72 (Fig. 1b; see below).

The structure of Ssu72 contains a central five-stranded β-sheet (β1–β5) that is surrounded by helices on both sides (Fig. 2a and Supplementary Fig. 4). The closest structural homologue of Ssu72 is the low-molecular-mass phosphotyrosine protein phosphatase (Fig. 2b)[18,19], as suggested previously[11,12,14], even though the two proteins share only 16% sequence identity. However, our studies show that Ssu72 possesses three unique structural features compared with this other phosphatase (Fig. 2b), which are formed by highly conserved residues (Supplementary Fig. 3) and have important functions. A small, two-stranded antiparallel β-sheet (β2A and β2B) is located near the active site (Fig. 2a). The αD helix is in a different conformation in Ssu72 and also contributes to phosphopeptide binding. Finally, Ssu72 contains an extra helix (αG) and a β-strand (β5) at the C terminus, which are essential for interactions with symplekin (Fig. 1b).

Our structure of the ternary complex showed that the CTD phosphopeptide, with good electron density for residues Thr 4 to Ser 7 (Fig. 2c), is bound with the peptide bond between pSer 5 and Pro 6 in the *cis* configuration (Fig. 2d). This is in sharp contrast to the conformations of the CTD phosphopeptides observed in other structures, which all have the Pro residue(s) in the *trans* configuration (Supplementary Fig. 5)[20,21]. With the *cis* configuration, the backbone of the phosphopeptide makes a 180° turn at the pSer-Pro residues, whereas the peptide in the *trans* configuration (Supplementary Fig. 5) would clash with Ssu72. Therefore, Ssu72 can only bind and dephosphorylate CTD substrates

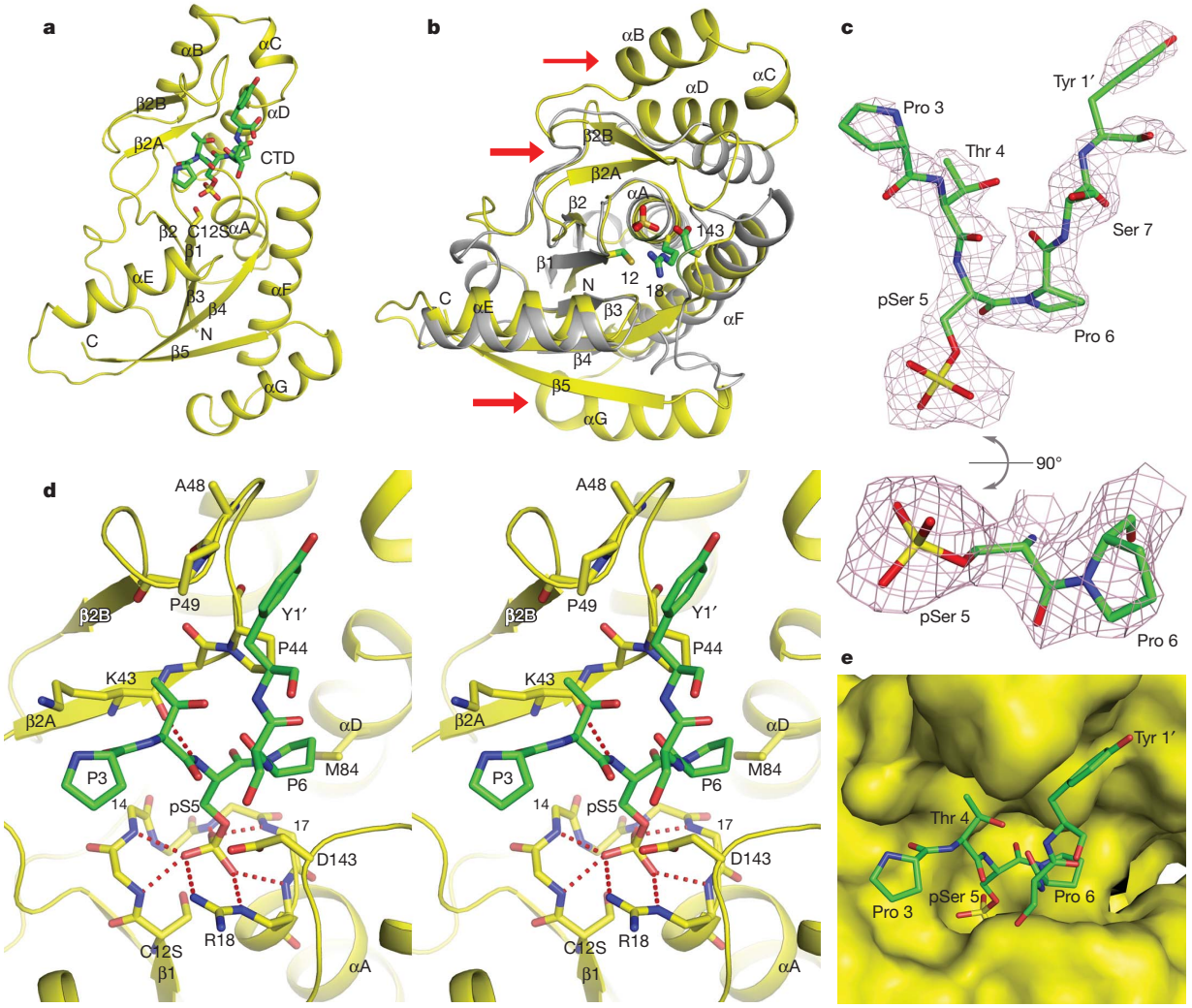


Figure 2 | Recognition of the CTD phosphopeptide by human Ssu72. **a**, Schematic drawing of the structure of human Ssu72–CTD phosphopeptide complex. **b**, Overlay of the structures of Ssu72 (yellow) and low-molecular-mass phosphotyrosine protein phosphatase (grey)[18,19]. Arrows indicate unique structural features in Ssu72. Stereo versions of **a** and **b** are given in Supplementary Fig. 4. **c**, Two views of the omit $F_o - F_c$ electron density at 2.4 Å resolution for the CTD phosphopeptide, contoured at 3σ. **d**, Stereo pair showing detailed interactions between the CTD phosphopeptide and the active site of Ssu72. Ion-pair and hydrogen-bonding interactions are indicated by red dashed lines. **e**, Molecular surface of the active-site region of Ssu72. The CTD phosphopeptide is shown as a stick model.

with the pSer-Pro peptide bond in the *cis* configuration, in contrast to all other known CTD phosphatases (Supplementary Information).

Our observation of a *cis* configuration for the CTD also provides a different interpretation for the role of the peptidyl-prolyl isomerase Pin1 (Ess1 in yeast) in regulating Pol II transcription[23–25]. It has been proposed that Pin1/Ess1 promotes the *trans* configuration of the CTD for dephosphorylation by Ssu72 (refs 24, 25), whereas our structure indicates that the opposite must be true. Our *in vitro* phosphatase assays demonstrate that Pin1 strongly stimulates the phosphatase activity of Ssu72 (Supplementary Information and Supplementary Fig. 6), consistent with its specificity for the *cis* configuration.

The active site of Ssu72 is located at the bottom of a narrow groove (Fig. 2e), one wall of which is formed by the small β-sheet (β2A and β2B) and the loop linking the two strands (Fig. 2d). This severely limits the possible conformation of the CTD, ensuring that only the *cis* configuration of the pSer-Pro peptide bond can be accommodated in the active site. In fact, the Thr 4-pSer 5 peptide bond is π-stacked with the Pro 6-Ser 7 peptide bond (Fig. 2d), suggesting a highly restrained conformation for the CTD phosphopeptide in this region. Residues Thr 4, pSer 5 and Pro 6 of the same repeat as well as Tyr 1 of the following repeat have interactions with the enzyme (Fig. 2d and Supplementary Information), explaining the preference for pSer 5 by Ssu72 and consistent with results from earlier biochemical studies on yeast Ssu72 (ref. 14).

The phosphate group of the peptide is bound deepest in the structure, having extensive ion-pair and hydrogen-bonding interactions with the enzyme (Fig. 2d). In addition, the main-chain amide group of pSer 5 is hydrogen-bonded to the main-chain carbonyl of Lys 43 (in β2A). The catalytic nucleophile of Ssu72, Cys 12, is located directly below the phosphate group and can be in the correct position for the inline nucleophilic attack on the phosphorus atom to initiate the reaction (Supplementary Fig. 7). The side chain of Asp 143 is located 3.5 Å from the Oγ atom of Ser 5, consistent with its role as the general acid to protonate the leaving group. There are some conformational changes in the active-site region of Ssu72, especially for the β2A–β2B loop, on binding of the CTD phosphopeptide (Supplementary Fig. 7), although this loop seems to be flexible and can assume different conformations in the various structures.

In the structures of the binary and ternary complexes, Ssu72 is bound to the concave face of Symp-N (Fig. 1b). About 950 $Å^2$ of the surface area of each protein is buried in the interface of this complex, which involves helices α3B–α6B of Symp-N (Fig. 3a and Supplementary Fig. 1), and helix αE, the following αE–β4 loop, helix αG and strand β5 of Ssu72 (Supplementary Fig. 3). In addition, residue Arg 206, at the tip of the long loop connecting helices α4B and α5A of Symp-N, is also located in the interface (Fig. 3a). Ion-pair, hydrogen-bonding and hydrophobic interactions make contributions to the formation of this complex (Supplementary Information). In particular, the side chains of Val 191 and Phe 193 of Ssu72 (in strand β5) establish hydrophobic interactions with those of Lys 185 (α4B) and Ile 251 (α5B) of symplekin in the centre of this interface (Fig. 3a). In addition, the side-chain hydroxyl group of Thr 190 (β5) of Ssu72 is hydrogen-bonded to the side chain of Asn 300 (α6B) of symplekin. The relative positions of Symp-N and Ssu72 seem to be somewhat variable among

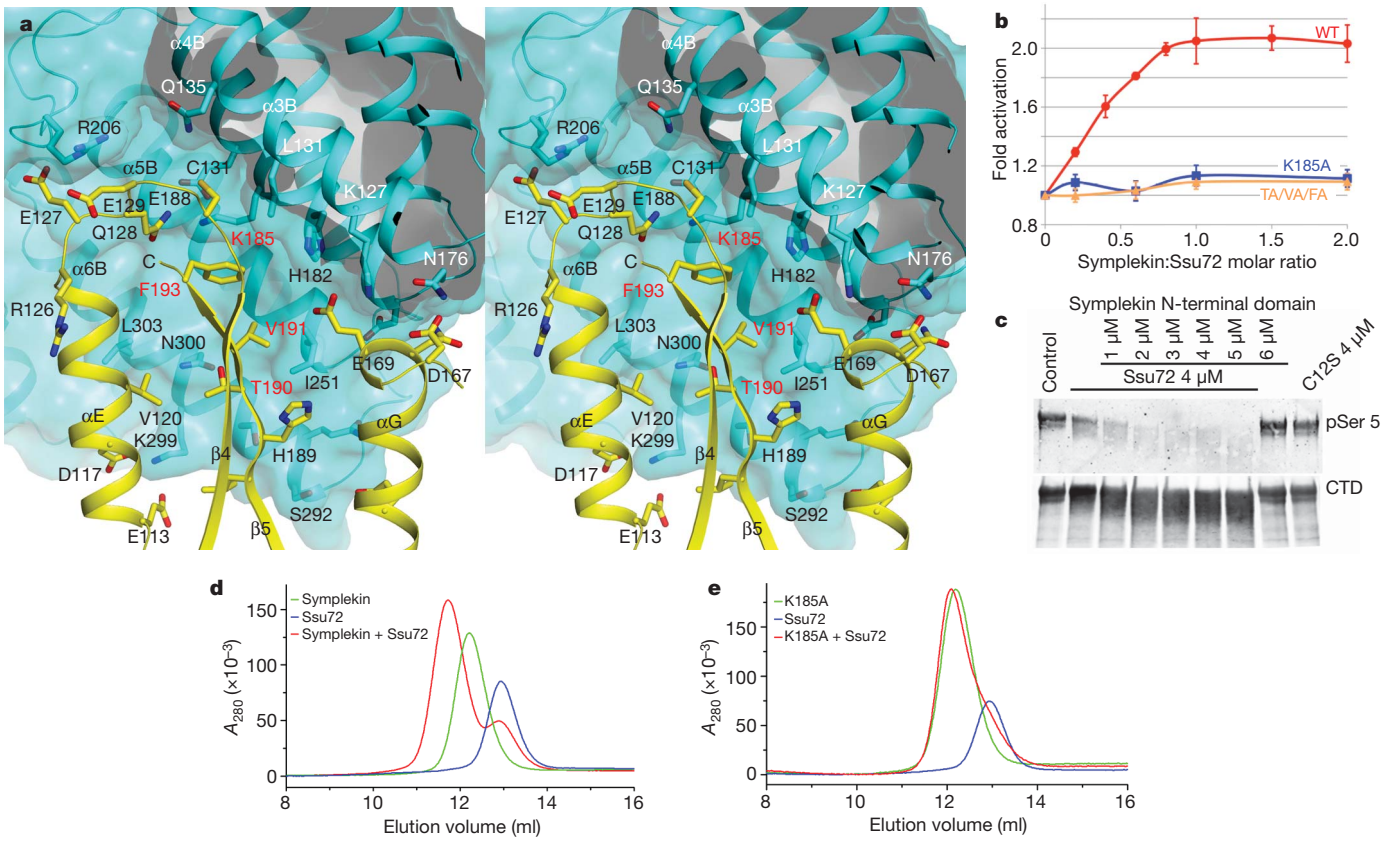


Figure 3 | Structural and biochemical characterizations of the symplekin–Ssu72 interface. **a**, Stereo pair showing detailed interactions between symplekin (in cyan) and Ssu72 (in yellow) in the interface. The molecular surface of symplekin is also shown (in cyan). Side chains making important contributions to the interface are shown as stick models. Residues labelled in red were selected for mutagenesis. **b**, The activity of Ssu72, measured by the hydrolysis of pNPP, as a function of the molar ratio of symplekin. TA/VA/FA, T190A/V191A/F193A mutant of Ssu72 (stimulation by wild-type (WT) symplekin). Results are shown as means ± s.d. for three independent experiments. **c**, Stimulation of the CTD Ser 5 phosphatase activity of Ssu72 by symplekin. The levels of pSer 5 and total CTD were determined using the H14 and 8WG16 antibodies, respectively. **d**, Gel-filtration profiles for wild-type symplekin N-terminal domain alone, wild-type human Ssu72 (full-length) alone, and a mixture of the two (with Ssu72 present in roughly twofold molar excess). **e**, Gel-filtration profiles for wild-type Ssu72 alone, K185A mutant of symplekin alone, and a mixture of the two.

the binary and ternary complexes (Supplementary Information and Supplementary Fig. 8).

The symplekin–Ssu72 interface is located about 25 Å from the active site of Ssu72 (Fig. 1b). However, phosphatase assays measuring the hydrolysis of a *p*-nitrophenyl phosphate (pNPP) model substrate[11,12] showed that Symp-N stimulated Ssu72 activity (Fig. 3b), and maximal activation was achieved when the two proteins were at a 1:1 molar ratio. To assess whether this stimulation also occurs with a natural substrate, we first used the decamer CTD phosphopeptide in the assay, monitoring the release of inorganic phosphate, and observed a similar stimulation (Supplementary Fig. 6). We next prepared a glutathione *S*-transferase (GST)–CTD fusion protein that had been phosphorylated on Ser 2 and Ser 5 with HeLa nuclear extract[26]. As demonstrated by western blotting with a pSer 5-specific antibody, Ssu72 dephosphorylated this protein on Ser 5, in a manner that was also stimulated by Symp-N (Fig. 3c). Ssu72 was specific for dephosphorylating pSer 5, because Ser 2 phosphorylation, as monitored by a pSer 2-specific antibody, was not affected (data not shown).

Our data indicate that the symplekin–Ssu72 interaction activated Ssu72 phosphatase activity, probably through stabilization of the Ssu72 structure and/or an allosteric mechanism. This is consistent with previous studies on the R129A mutant (*ssu72-2*) of yeast Ssu72, equivalent to Arg 126 in human Ssu72 (Supplementary Fig. 3). This mutant shows a twofold decrease in catalytic activity compared with wild-type Ssu72 and produces a severe growth defect at the non-permissive temperature[16]. Arg 126 is far from the active site and is in fact near the interface with symplekin (Fig. 3a). However, it does not contribute directly to interactions with symplekin, and the R126A mutation did not disrupt interaction with Symp-N (data not shown).

To assess the importance of individual residues for the stability of the symplekin–Ssu72 complex, we introduced mutations in the interface and characterized their effects on the complex by using gel-filtration chromatography and phosphatase assays. The presence of wild-type Ssu72 gave rise to a clear shift in the peak for Symp-N from a gel-filtration column (Fig. 3d), corresponding to the formation of the symplekin–Ssu72 complex. Ssu72 was present in twofold molar excess in this experiment, and only half of this protein was incorporated into the complex (Fig. 3d), demonstrating a 1:1 stoichiometry for the complex. Mutation of a symplekin residue in the interface, K185A (Fig. 3a), essentially abolished the interaction with wild-type Ssu72 (Fig. 3e), and mutation of three Ssu72 residues in the interface, T190A/V191A/F193A, abolished the interaction with wild-type symplekin. The chromatographic behaviour of the mutants alone was similar to that of the wild-type protein (Fig. 3e), suggesting that the mutations did not disrupt the structure of the proteins. This was also confirmed by the crystal structure of the K185A mutant (data not shown). Consistent with the gel-filtration data, the symplekin K185A mutant failed to stimulate Ssu72 phosphatase activity, and the T190A/V191A/F193A mutant of Ssu72 could not be stimulated by wild-type Symp-N (Fig. 3b).

We next wished to assess the functional importance of the symplekin–Ssu72 interaction with respect to 3′-end formation. Given the roles of their yeast counterparts in both transcription and polyadenylation, we used a transcription-coupled 3′-end processing assay[27]. HeLa nuclear extract was preincubated with increasing concentrations of Symp-N, which led to a pronounced inhibition of polyadenylation (Fig. 4a), similar to an effect observed earlier with the yeast Pta1 N-terminal domain in a transcription-independent assay[9]. Transcription, as measured by the accumulation of unprocessed pre-mRNA, was not affected (Fig. 4b). RNase protection assays showed that 3′-end cleavage was also not affected (Supplementary Fig. 9), indicating that Symp-N affects only the polyadenylation step of 3′-end formation. Inclusion of purified Ssu72 during the preincubation with Symp-N blocked the inhibition, whereas Ssu72 alone had no effect (Fig. 4c). The K185A mutation in Symp-N abolished this inhibitory effect, whereas the T190A/V191A/F193A mutant of Ssu72 failed to overcome the inhibition by wild-type Symp-N (Fig. 4c). These results provide strong evidence that the

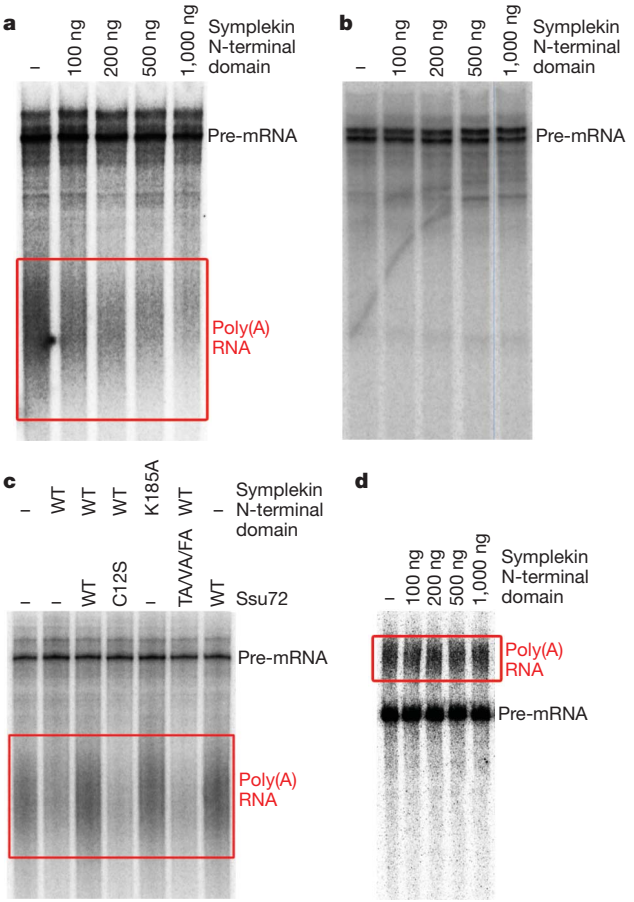


Figure 4 | Functional characterization of the symplekin–Ssu72 interaction. **a**, Transcription-coupled polyadenylation is inhibited in a dose-dependent fashion by symplekin N-terminal domain. Transcription-processing was performed in HeLa nuclear extract; RNAs were purified and separated into poly(A)$^-$ and poly(A)$^+$ fractions, and resolved by denaturing PAGE. Positions of unprocessed 'run-off' RNA (pre-mRNA) and cleaved and polyadenylated RNA (poly(A) RNA) are indicated. **b**, Symplekin N-terminal domain does not inhibit transcription. Poly(A)$^-$ RNAs (2%) from the transcription-coupled polyadenylation assays in the presence of increasing concentrations of symplekin N-terminal domain are shown. **c**, Ssu72 overcomes the inhibition of transcription-coupled polyadenylation by symplekin. The effects of the K185A mutant of symplekin and the C12S and T190A/V191A/F193A (TA/VA/FA) mutants of Ssu72 on polyadenylation are also shown. **d**, Polyadenylation of SV40 late pre-mRNA, uncoupled to transcription, is not inhibited by symplekin N-terminal domain.

inhibitory effect of Symp-N reflects its interaction with Ssu72, and thus implicates Ssu72 in mammalian 3′-end processing. In contrast with studies in yeast[7,9], the catalytically inactive C12S mutant of Ssu72 failed to overcome this inhibition (Fig. 4c), and Symp-N had no detectable effect on transcription-independent polyadenylation (Fig. 4d, and data not shown). Together, these results indicate that Ssu72 phosphatase activity is required for polyadenylation of pre-mRNAs, but only when processing is coupled to transcription.

Our finding that a CTD phosphopeptide is bound to Ssu72 with the pSer-Pro peptide bond in the *cis* configuration indicates the existence of a novel CTD conformation. Although Ssu72 has been well studied in yeast and has functions in transcription and 3′-end processing, essentially nothing was known about its mammalian counterpart. In fact, whereas the yeast enzyme is a stable component of the polyadenylation machinery and is required for processing, mammalian Ssu72 has not been found associated with polyadenylation factors and was not detected in a recent proteomic analysis of the assembled polyadenylation complex[28]. Consistent with this, our results provide evidence that in mammals Ssu72 is only necessary for polyadenylation when processing is coupled to

transcription. A parsimonious model is that symplekin recruits Ssu72 to the transcription complex and activates its phosphatase activity, which promotes polyadenylation. Conceivably, this occurs by facilitating the recruitment of poly(A) polymerase, known for many years to be only weakly associated with other 3′-end processing factors[28,29], to the complex. Given that the CTD is necessary for efficient 3′-end formation in mammalian cells[26,30], and that CTD pSer 5 is the only known target of Ssu72, CTD pSer 5 dephosphorylation may well be important in facilitating polyadenylation during transcription.

METHODS SUMMARY

Crystallography. The N-terminal domain of human symplekin and full-length human Ssu72 were overexpressed separately in *Escherichia coli* and purified. The symplekin–Ssu72 complex was purified by gel filtration of a mixture of the two proteins. Crystals were obtained by the sitting-drop vapour-diffusion method, and the structures were determined by the selenomethionyl single-wavelength anomalous diffraction method and the molecular replacement method.

***In vitro* transcription-coupled polyadenylation assay.** Transcription-coupled polyadenylation was performed with a DNA construct containing GAL4-binding sites upstream of the adenovirus E4 core promoter and SV40 late poly(A) site downstream. Recombinant symplekin and Ssu72 proteins were preincubated with HeLa nuclear extract before transcription was started by adding the DNA template and purified GAL4–VP16. RNA products were purified, separated into nonpolyadenylated and polyadenylated fractions and analysed on 5% denaturing gel. Radioactivity was detected with a PhosphorImager. Assays were repeated multiple times with consistent results.

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

Received 16 March; accepted 3 August 2010.
Published online 22 September 2010.

1. Takagaki, Y. & Manley, J. L. Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol. Cell. Biol.* **20,** 1515–1525 (2000).
2. Zhao, J., Kessler, M. M., Helmling, S., O'Connor, J. P. & Moore, C. L. Pta1, a component of yeast CFII, is required for both cleavage and poly(A) addition of mRNA precursor. *Mol. Cell. Biol.* **19,** 7733–7740 (1999).
3. Zhao, J., Hyman, L. & Moore, C. L. Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* **63,** 405–445 (1999).
4. Mandel, C. R., Bai, Y. & Tong, L. Protein factors in pre-mRNA 3′-end processing. *Cell. Mol. Life Sci.* **65,** 1099–1122 (2008).
5. Calvo, O. & Manley, J. L. Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev.* **17,** 1321–1327 (2003).
6. Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136,** 688–700 (2009).
7. He, X. *et al.* Functional interactions between the transcription and mRNA 3′ end processing machineries mediated by Ssu72 and Sub1. *Genes Dev.* **17,** 1030–1042 (2003).
8. Nedea, E. *et al.* Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3′-ends. *J. Biol. Chem.* **278,** 33000–33010 (2003).
9. Ghazy, M. A., He, X., Singh, B. N., Hampsey, M. & Moore, C. The essential N terminus of the Pta1 scaffold protein is required for snoRNA transcription termination and Ssu72 function but is dispensible for pre-mRNA 3′-end processing. *Mol. Cell. Biol.* **29,** 2296–2307 (2009).
10. Dichtl, B. *et al.* A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. *Mol. Cell* **10,** 1139–1150 (2002).
11. Meinhart, A., Silberzahn, T. & Cramer, P. The mRNA transcription/processing factor Ssu72 is a potential tyrosine phosphatase. *J. Biol. Chem.* **278,** 15917–15921 (2003).
12. Ganem, C. *et al.* Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. *EMBO J.* **22,** 1588–1598 (2003).
13. Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C. L. & Hampsey, M. Ssu72 is an RNA polymerase II CTD phosphatase. *Mol. Cell* **14,** 387–394 (2004).
14. Hausmann, S., Koiwa, H., Krishnamurthy, S., Hampsey, M. & Shuman, S. Different strategies for carboxyl-terminal domain (CTD) recognition by serine 5-specific CTD phosphatases. *J. Biol. Chem.* **280,** 37681–37688 (2005).
15. St-Pierre, B. *et al.* Conserved and specific functions of mammalian ssu72. *Nucleic Acids Res.* **33,** 464–477 (2005).
16. Reyes-Reyes, M. & Hampsey, M. Role for the Ssu72 C-terminal domain phosphatase in RNA polymerase II transcription elongation. *Mol. Cell. Biol.* **27,** 926–936 (2007).
17. Ansari, A. & Hampsey, M. A role for the CPF 3′-end processing machinery in RNAP II-dependent gene looping. *Genes Dev.* **19,** 2969–2978 (2005).
18. Su, X.-D., Taddei, N., Stefani, M., Ramponi, G. & Nordlund, P. The crystal structure of a low-molecular-weight phosphotyrosine protein phosphatase. *Nature* **370,** 575–578 (1994).
19. Zhang, M., Van Etten, R. L. & Stauffacher, C. V. Crystal structure of bovine heart phosphotyrosyl phosphatase at 2.2-Å resolution. *Biochemistry* **33,** 11097–11105 (1994).
20. Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S. & Cramer, P. A structural perspective of CTD function. *Genes Dev.* **19,** 1401–1415 (2005).
21. Zhang, Y. *et al.* Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1. *Mol. Cell* **24,** 759–770 (2006).
22. Kennedy, S. A. *et al.* Crystal structure of the HEAT domain from the pre-mRNA processing factor symplekin. *J. Mol. Biol.* **392,** 115–128 (2009).
23. Xu, Y.-X., Hirose, Y., Zhou, X. Z., Lu, K. P. & Manley, J. L. Pin1 modulates the structure and function of human RNA polymerase II. *Genes Dev.* **17,** 2765–2776 (2003).
24. Krishnamurthy, S., Ghazy, M. A., Moore, C. & Hampsey, M. Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries. *Mol. Cell. Biol.* **29,** 2925–2934 (2009).
25. Singh, N. *et al.* The Ess1 prolyl isomerase is required for transcription termination of small noncoding RNAs via the Nrd1 pathway. *Mol. Cell* **36,** 255–266 (2009).
26. Hirose, Y. & Manley, J. L. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395,** 93–96 (1998).
27. Rozenblatt-Rosen, O. *et al.* The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3′ mRNA processing factors. *Proc. Natl Acad. Sci. USA* **106,** 755–760 (2009).
28. Shi, Y. *et al.* Molecular architecture of the human pre-mRNA 3′ processing complex. *Mol. Cell* **33,** 365–376 (2009).
29. Takagaki, Y., Ryner, L. C. & Manley, J. L. Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. *Cell* **52,** 731–742 (1988).
30. McCracken, S. *et al.* The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385,** 357–361 (1997).

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

Acknowledgements We thank F. Forouhar and J. Seetharaman for help with data collection; T. Auperin for initial experiments with symplekin; R. Abramowitz and J. Schwanof for setting up the X4C beamline; S. Myers for setting up the X29A beamline at the National Synchrotron Light Source; M. Heidemann and D. Eick for providing the 3E10 antibody; N. Rao and C. Logan for HeLa nuclear extract and help with *in vitro* assays; and J. Decatur for characterizing the phosphopeptide by NMR. This research is supported in part by grants from the National Institutes of Health to L.T. (GM077175) and J.L.M. (GM028983).

Author Contributions K.X., S.X., T.K. and M.M.B. performed protein expression, purification and crystallization experiments. K.X., S.X. and L.T. conducted crystallographic data collection, structure determination and refinement. T.N. and K.X. performed polyadenylation experiments. K.X. performed Ssu72 phosphatase assays. All authors commented on the manuscript. L.T. and J.L.M. designed the experiments, analysed the data and wrote the paper.

Author Information Atomic coordinates have been deposited at the Protein Data Bank (accession codes 3O2Q, 3O2S, 3O2T, 3ODR and 3ODS). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.T. (ltong@columbia.edu).

METHODS

Protein expression and purification. Residues 30–395 of human symplekin were subcloned into the pET28a vector (Novagen). The recombinant protein carries a hexahistidine tag at the N terminus. The plasmids were transformed into *E. coli* BL21(DE3) Star cells. After induction with 0.5 mM isopropyl β-D-thiogalactoside, the cells were allowed to grow at 20 °C for 14–16 h, collected by centrifugation, and lysed by sonication. Soluble symplekin was purified by Ni^{2+}-nitrilotriacetate (Qiagen) and gel-filtration (Sephacryl S-300; GE Healthcare) chromatography. Purified protein was concentrated to 15 mg ml^{-1} in a buffer containing 20 mM Tris-HCl pH 8.5, 200 mM NaCl, 10 mM dithiothreitol (DTT) and 5% (v/v) glycerol, flash-frozen in liquid nitrogen and stored at −80 °C. The N-terminal His-tag was not removed for crystallization.

The selenomethionyl protein sample was produced in *E. coli* B834(DE3) cells, and the bacteria were grown in defined LeMaster medium supplemented with selenomethionine[31]. The purification procedure was the same as for the native protein.

To prepare the symplekin–Ssu72 complex, residues 30–360 of human symplekin and full-length human Ssu72 were cloned separately into the pET28a vector. Both proteins carried a hexahistidine tag at the N terminus and were purified separately by Ni^{2+}-nitrilotriacetate (Qiagen) and gel-filtration (Sephacryl S-300; GE Healthcare) chromatography. The purified proteins were then mixed, with Ssu72 present in slight molar excess (1:1.2 molar ratio), and the symplekin–Ssu72 complex was purified by gel-filtration chromatography. Purified protein was concentrated to 10 mg ml^{-1} in a buffer containing 20 mM Tris-HCl pH 8.5, 200 mM NaCl, 10 mM DTT and 5% (v/v) glycerol, flash-frozen in liquid nitrogen and stored at −80 °C.

Protein crystallization. Crystals of symplekin (residues 30–395) were obtained with the sitting-drop vapour-diffusion method at 20 °C. The reservoir solution contained 50 mM Bis-Tris pH 6.5, 40 mM ammonium sulphate and 40% (v/v) pentaerythritol ethoxylate (15/4 EO/OH). The crystals belong to space group $P2_1$, with unit cell parameters of $a = 41.6$ Å, $b = 63.0$ Å, $c = 62.3$ Å and $\beta = 90.6°$. There is one symplekin molecule in the asymmetric unit.

Crystals of the symplekin–Ssu72 complex were obtained with the sitting-drop vapour-diffusion method at 20 °C. The reservoir solution contained 0.32 M ammonium sulphate and 26% (w/v) PEG 3350. The crystals belong to space group $P2_12_12$, with unit cell parameters of $a = 99.4$ Å, $b = 113.6$ Å and $c = 59.1$ Å. There is one symplekin–Ssu72 complex in the asymmetric unit.

Crystals of the symplekin–Ssu72–CTD phosphopeptide ternary complex were obtained with the sitting-drop vapour-diffusion method at 20 °C. The reservoir solution contained 1.6 M ammonium chloride and 27% (w/v) PEG 3350. Sodium potassium tartrate (10 mM) was included as an additive in the drop. These crystals of the symplekin–Ssu72 binary complex were soaked overnight in a solution containing 30 mM CTD phosphopeptide, 25% (w/v) PEG 3000 and 100 mM Tris-HCl pH 8.5 at 20 °C. The crystal belongs to space group $P2_1$, with unit cell parameters of $a = 67.0$ Å, $b = 97.6$ Å, $c = 105.0$ Å and $\beta = 98.7°$. There are two copies of the symplekin–Ssu72 complex in the asymmetric unit. The CTD phosphopeptide was observed in only one of the Ssu72 molecules, and a phosphate is bound in the active site of the other Ssu72 molecule. We found that commercial PEG 3350 contains some phosphate (about 1 mM) as a contaminant.

The crystals were cryoprotected by the reservoir solution, supplemented with 25% (v/v) ethylene glycol if necessary, and flash-frozen in liquid nitrogen for data collection at 100 K.

Data collection and structure determination. The structure of symplekin alone was determined by the selenomethionyl anomalous diffraction method[32]. A single-wavelength anomalous diffraction (SAD) data set to 2.0 Å resolution was collected on a SeMet-substituted crystal at beamline 9-2 of the Stanford Synchrotron Radiation Laboratory. A native reflection data set to 1.4 Å resolution was collected at the same beamline. The diffraction data were processed and scaled with the HKL package[33]. The data processing statistics are summarized in Supplementary Table 1.

The Se atoms were located with the program BnP[34], and the reflections were phased with the program Solve[35]. Most of the residues were built automatically by the program Resolve, and the model was completed by manual building with the programs O[36] and Coot[37]. The structure refinement was performed with the programs CNS[38] and Refmac[39].

The structure of the symplekin–Ssu72 complex was determined by the molecular replacement method, with the program COMO[40]. A native reflection data set to 2.5 Å resolution was collected at the X4C beamline of the National Synchrotron Light Source. The structures of symplekin and *Drosophila* Ssu72 (PDB ID 3FDF) were used as the search models.

A native reflection data set to 2.4 Å resolution was collected at the X29 beamline on the symplekin–Ssu72–CTD phosphopeptide ternary complex.

Symplekin–Ssu72 interactions. Symplekin and Ssu72 mutants were made with the QuikChange kit (Stratagene) and verified by sequencing. The mutant proteins were expressed and purified by following the same protocol as that for the wild-type protein.

Analytical gel-filtration experiments were carried out on a Superose-12 10/30 column (GE Healthcare), with a buffer containing 20 mM Tris-HCl pH 7.5 and 200 mM NaCl. Symplekin (210 μg) and Ssu72 (380 μg) were mixed and diluted to a final volume of 500 μl with the gel-filtration buffer. The mixture was incubated on ice for 1 h before being loaded on the column. The proteins were also run separately on the column to determine their migration behaviour alone.

Ssu72 CTD peptide phosphatase assays. Reaction mixtures (50 μl) containing 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM DTT, the indicated amount of CTD phosphopeptide, Ssu72 and symplekin were incubated at 37 °C. Time-point samples were taken and quenched by the addition of 0.5 ml of malachite green reagent (BIOMOL Research Laboratories). Phosphate release was determined by measuring A_{620} and comparing it with a phosphate standard curve.

To study the effect of Pin1 on Ssu72, reaction mixtures (50 μl) containing 50 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM DTT, 500 μM CTD phosphopeptide and the indicated amount of Ssu72 and GST–Pin1 were incubated at 10 °C. Phosphate release was determined with the malachite green reagent.

Ssu72 GST–CTD phosphatase assays. GST–CTD fusion protein was expressed, purified and phosphorylated *in vitro* as described previously[41]. GST–CTD phosphatase assays were performed in a total volume of 10 μl containing 50 mM Tris-HCl pH 7.9, 100 mM KCl, 12.5 mM $MgCl_2$, 0.5 mM EDTA, 5% (v/v) glycerol and the indicated amount of phosphorylated GST–CTD as a substrate, Ssu72, symplekin or their mutants. Reactions were incubated for various durations at 37 °C, stopped by the addition of SDS loading buffer and resolved by 8% SDS–PAGE. pSer 2 and pSer 5 levels were detected by western blot with 3E10 (ref. 42) and H14 (Covance) antibodies, respectively.

***In vitro* transcription-coupled polyadenylation assay.** The DNA construct used for the transcription-coupled polyadenylation assay contained GAL4-binding sites upstream of an E4 core promoter and an SV40 late poly(A) site downstream. Transcription-coupled polyadenylation was performed at 30 °C for 1 h in 20-μl reaction mixtures containing 10 μl of nuclear extract, 100 ng of Gal4–VP16, recombinant proteins (symplekin and Ssu72), 12 mM HEPES pH 7.9, 500 ng of the DNA templates, 0.5 mM each of ATP, GTP and CTP, 15 μM unlabelled UTP, 10 μCi of [α-^{32}P]UTP, 4 mM $MgCl_2$, 20 mM creatine phosphate (di-tris), 2.4% PEG 8000, 12% glycerol, 60 mM KCl, 0.12 mM EDTA, 0.12 mM DTT and 0.3 mM phenylmethylsulphonyl fluoride. Recombinant symplekin and Ssu72 proteins were preincubated with nuclear extract (30 min at 23 °C) before transcription was started by adding the DNA templates. The reaction was stopped by adding proteinase K. RNA products were separated into non-polyadenylated and polyadenylated fractions by oligo(dT) selection; thereafter 2% of non-polyadenylated and 100% of polyadenylated fractions were analysed on 5% denaturing gel. Radioactivity was detected with a PhosphorImager.

RNase protection assay. To make the probe, pG3SVL-A was linearized with *Sal*I and transcribed with T7 RNA polymerase (Promega) for 2 h at 37 °C, uniformly labelled with [α-^{32}P]UTP. The RNA was gel purified and resuspended in hybridization buffer containing 40 mM PIPES pH 6.4, 400 mM NaCl, 1 mM EDTA and 80% (v/v) formamide.

Transcription-coupled polyadenylation was performed as mentioned above except that 0.5 mM each of ATP, GTP, UTP and CTP, but no [α-^{32}P]UTP, were used in the reaction mixture. The RNA/DNA mixture was resuspended in 20 μl of Turbo DNase buffer (Ambion) with 1 U of Turbo DNase (Ambion) and incubated for 1 h at 37 °C to remove the DNA template. The remaining RNA products were hybridized overnight with the probe (5×10^5 c.p.m.) at 42 °C. Each reaction was then incubated for 45 min with 14 μg of RNase A and 0.7 μg of RNase T1 at 30 °C. The final RNA products were resolved on a 10% denaturing gel. Radioactivity was detected with a PhosphorImager.

31. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9,** 1665–1672 (1990).
32. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254,** 51–58 (1991).
33. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
34. Weeks, C. M. & Miller, R. The design and implementation of SnB v2.0. *J. Appl. Cryst.* **32,** 120–124 (1999).
35. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **374,** 22–37 (2003).
36. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
37. Emsley, P. & Cowtan, K. D. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132 (2004).

38. Brunger, A. T. *et al.* Crystallography & NMR System: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54,** 905–921 (1998).
39. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255 (1997).
40. Jogl, G., Tao, X., Xu, Y. & Tong, L. COMO: A program for combined molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **57,** 1127–1134 (2001).
41. Hirose, Y. & Manley, J. L. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395,** 93–96 (1998).
42. Chapman, R. D. *et al.* Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318,** 1780–1782 (2007).

doi:10.1038/nature09406

Structure of a fucose transporter in an outward-open conformation

Shangyu Dang[1], Linfeng Sun[1]*, Yongjian Huang[1]*, Feiran Lu[1], Yufeng Liu[1], Haipeng Gong[1], Jiawei Wang[1] & Nieng Yan[1]

The major facilitator superfamily (MFS) transporters are an ancient and widespread family of secondary active transporters[1]. In *Escherichia coli*, the uptake of L-fucose, a source of carbon for microorganisms, is mediated by an MFS proton symporter, FucP[2,3]. Despite intensive study of the MFS transporters, atomic structure information is only available on three proteins and the outward-open conformation has yet to be captured[4–6]. Here we report the crystal structure of FucP at 3.1 Å resolution, which shows that it contains an outward-open, amphipathic cavity. The similarly folded amino and carboxyl domains of FucP have contrasting surface features along the transport path, with negative electrostatic potential on the N domain and hydrophobic surface on the C domain. FucP only contains two acidic residues along the transport path, Asp 46 and Glu 135, which can undergo cycles of protonation and deprotonation. Their essential role in active transport is supported by both *in vivo* and *in vitro* experiments. Structure-based biochemical analyses provide insights into energy coupling, substrate recognition and the transport mechanism of FucP.

L-fucose is a major constituent of N-linked glycans on the cell surfaces of microbes, plants and animals[7]. It can serve as the sole carbon source for some bacteria. In *E. coli*, the uptake of L-fucose is mediated by FucP, a fucose/H^+ symporter. FucP homologues include the glucose/H^+ or mannose/H^+ symporter GlcP[8], the 2-deoxy-D-ribose transporter DeoP[9], the Na^+-dependent sugar transporter HP1174[10], and the Na^+-dependent methyl α-glucoside transporter NaGLT1[11] (Supplementary Fig. 1). FucP and its homologues belong to the MFS family, the members of which exploit the electrochemical potential to shuttle substrates across cell membranes[1,12]. The MFS transporters are thought to use an alternating-access mechanism to upload and download substrate[13,14]. The best-characterized MFS transporter is LacY, the *E. coli* lactose/H^+ symporter[13,15,16], for which structures of the substrate-free and substrate-bound states are known[4,17,18]. These LacY structures have a nearly identical, inward-open conformation. Similarly, the MFS transporter GlpT (a glycerol-3-phosphate/phosphate antiporter) also has an inward-open conformation[5], whereas the multidrug resistance antiporter EmrD is in an intermediate state[6].

Although FucP and LacY are both MFS sugar/H^+ symporters, they share limited sequence homology (Supplementary Fig. 2a). It is therefore difficult to understand the functional mechanism of FucP using available structural and biochemical information on LacY. We sought to determine the crystal structure of FucP. We overexpressed full-length FucP in *E. coli* and purified it to homogeneity. Consistent with the *in vivo* observations[2,3], the recombinant FucP transported L-fucose in a pH-dependent manner (Supplementary Fig. 2b). Additional characterization confirmed that FucP selectively permeated L-fucose[2] (Fig. 1a). The FucP structure was determined by mercury-based, single-wavelength anomalous dispersion and refined at 3.14 Å resolution (Supplementary Table 1 and Supplementary Figs 3 and 4).

Consistent with an earlier prediction[19], FucP contains 12 transmembrane segments, with both the N and the C termini located on the cytoplasmic side (Fig. 1b, left). Similar to known structures of MFS transporters[4–6], the 12 transmembrane segments are arranged into two halves, the N and C domains, which can be superimposed with a root mean squared deviation (r.m.s.d.) of 2.97 Å over 138 Cα atoms and are related by a pseudo-two-fold symmetry axis that is perpendicular to the membrane bilayer (Supplementary Fig. 5a, b). The N and C domains each comprise a pair of internal structural repeats, which are related by an approximate 180° rotation around an axis parallel to the membrane bilayer (Supplementary Fig. 5c, d). Unlike any known MFS transporter structure[4–6], that of FucP has an outward-open conformation. A central cavity, approximately 20 Å in depth and 10 Å in diameter at the periplasmic side, is surrounded by transmembrane segments 1, 2, 4 and 5 of the N domain and transmembrane segments 7, 8, 10 and 11 of the C domain (Fig. 1b and Supplementary Fig. 5a). This cavity is probably an important part of the transport path for substrate molecules.

Despite limited sequence similarity between FucP and LacY (Supplementary Fig. 2), the N domains of the two proteins can be superimposed with an r.m.s.d. of 2.86 Å over 153 Cα atoms, whereas their C domains have an r.m.s.d. of 2.90 Å over 173 Cα atoms (Fig. 2 and Supplementary Fig. 6). Thus, the overall conformations of the N and C domains seem to be 'rigid'. To switch from outward-open to a LacY-like inward-open conformation, the N and C domains of FucP need to undergo a rotation of approximately 38° around an axis parallel to the membrane bilayer (Fig. 2). The long, flexible linker between transmembrane segment 6 (TM6) and TM7 (Supplementary Fig. 2), which demarcates the N and C domains, allows the proposed rotation. Thus, we propose that the N and C domains of FucP undergo concentric, rigid-body rotations to achieve the two essential conformations required for alternating access: outward-open and inward-open.

In the outward-open FucP, TM4 and TM10 interact with each other at the centre, where Glu 135 of TM4 accepts a hydrogen bond from Tyr 365 of TM10 (Supplementary Fig. 7a, b). To achieve the inward-open conformation, the interdomain contacts between the N and C halves must undergo considerable rearrangements. Notably, TM1 and TM7 do not directly contact each other in the outward-open conformation of FucP, whereas they are at the centre of interdomain packing in the inward-open structures of LacY[4] and GlpT[5]. Therefore, a rigid-body rotation of the N and C domains is likely to bring together TM1 and TM7 in the inward-open FucP (Supplementary Fig. 7c).

The N and C domains of FucP have contrasting surface features, giving rise to an amphipathic cavity (Figs 1b and 3a). The N domain has a strip of negative electrostatic potential along the central cavity; whereas the C domain contains a hydrophobic patch in the cavity, capped by positively charged residues on the periplasmic side as well as the cytoplasmic side (Fig. 3a). The cavity-facing side of the N domain is enriched by Asn and Gln residues, with 14 Asn or Gln and four Asp or Glu residues (Fig. 3b, left). Only two of the four Asp or Glu residues are highly conserved among FucP homologues and located in the transport path. In contrast, residues of the C domain that line the central cavity are mostly hydrophobic, with only a few polar residues and two

[1]State Key Laboratory of Bio-membrane and Membrane Biotechnology, Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China.
*These authors contributed equally to this work.

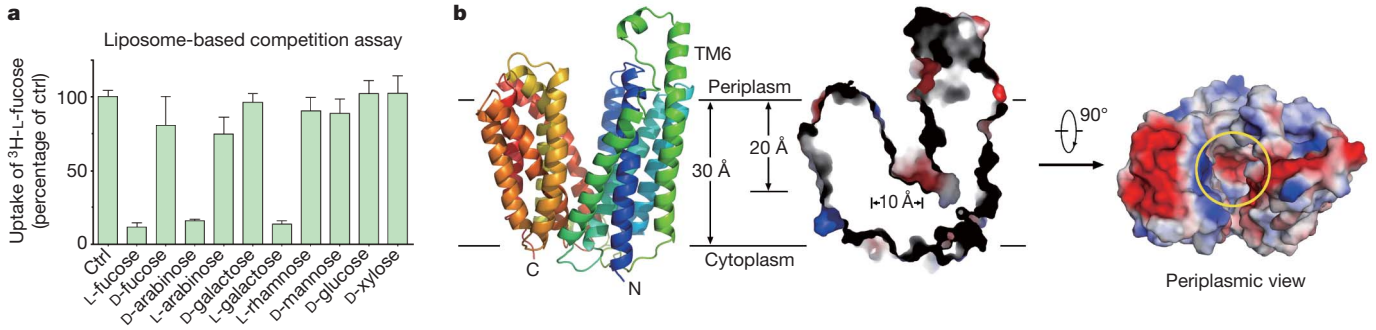


Figure 1 | FucP has an outward-open conformation. **a**, FucP selectively permeates L-fucose; this can be inhibited by D-arabinose and L-galactose but not by the other sugars tested. 'Ctrl' refers to the condition where only 0.17 μM ^{3}H-L-fucose was present in the external buffer, without additional cold sugars. In other, parallel, experiments, cold competitors were added at a final concentration of 5 mM in the external solution. Error bars, s.d. of three independent experiments. Details of the experiments can be found in Supplementary Methods. **b**, Overall structure of FucP. The central image shows a slab of FucP in surface electrostatic potential to highlight the central cavity. The right-hand image shows FucP in surface electrostatic potential from a periplasmic view. All structure figures were prepared with PYMOL[30].

charged residues, Arg 283 and Arg 312 (Fig. 3b, right), which are unlikely to be deprotonated during transport. Therefore, the only residues that can undergo cycles of protonation and deprotonation along the transport path are Asp 46 and Glu 135, which are located on the third helical turn of the corresponding TM1 and TM4 in the two internal structural repeats of the N domain (Supplementary Fig. 5c).

The essential roles of the two acid residues are manifested by both *in vivo* uptake and *in vitro* counterflow experiments. Replacement of Asp 46 by Ala or Asn, or replacement of Glu 135 by Ala, Gln or Asp, resulted in complete abrogation of FucP-mediated active transport of L-fucose (Fig. 3c). In the liposome-based counterflow assay, which does not rely on proton gradient for substrate exchange, Asp 46 Ala and Asp 46 Asn showed a transport activity that is two to three times higher than that of the wild-type FucP, whereas Glu 135 mutants had severely compromised activities (Fig. 3d). These observations indicate that although both Asp 46 and Glu 135 are essential for the active transport activity of FucP, they have distinct biochemical roles. In addition, neutralization of Asp 46 by Asn or Ala rendered FucP insensitive to pH changes in the counterflow assay (Fig. 3e). The behaviour of Asp 46 is reminiscent of Glu 325 of LacY[13,20]. These observations suggest that Asp 46 is essential in proton-dependent active transport, whereas Glu 135 might be involved in additional processes such as substrate recognition.

The binding affinity between L-fucose and FucP, measured in the presence of the detergent Cymal-7 by isothermal titration calorimetry (ITC), was approximately 0.47 ± 0.02 mM (Fig. 4a and Supplementary Fig. 8a), which is similar to that between lactose and LacY[21]. We included 2 mM L-fucose in the crystallization solution. During structure refinement, a disc-shaped electron density, which is reminiscent of a sugar ring, was observed at the bottom of the central cavity, in proximity to Glu 135 and Asn 162 (Supplementary Fig. 9a). However, the same electron density persisted in crystals grown in the absence of L-fucose (Supplementary Fig. 9b), indicating that the bound ligand might be the sugar moiety of *n*-nonyl-β-D-glucopyranoside (β-NG), which is a component of the crystallization solution (Fig. 4b and Supplementary Fig. 9c). Notably, FucP had no detectable interaction with L-fucose in the presence of β-NG, suggesting that β-NG might have blocked L-fucose binding (Supplementary Fig. 8b). Finally, we determined the crystal structure of the FucP mutant Asn 162 Ala (Supplementary Table 2), which abolished L-fucose binding (Supplementary Fig. 8c). The overall structure was nearly identical to that of wild-type FucP, and the electron density for β-NG was clearly visible (Fig. 4c and Supplementary Fig. 10).

These experimental data demonstrate that β-NG is present in the crystal structure of FucP, blocking L-fucose binding. The sugar moiety of β-NG is stereochemically different from L-fucose, so it is not clear why β-NG can coincide with L-fucose binding. It is probably possible because of the presence of a hydrophobic pocket next to the potential L-fucose-binding site, which may strengthen β-NG binding by accommodating its aliphatic tail group (Supplementary Fig. 9c). Notably, it is not uncommon for a detergent molecule to occupy the substrate-binding site[22]. We tentatively modelled an L-fucose molecule in the position of the sugar moiety of β-NG, where L-fucose is within hydrogen-bond distances of Glu 135 and Asn 162 (Supplementary

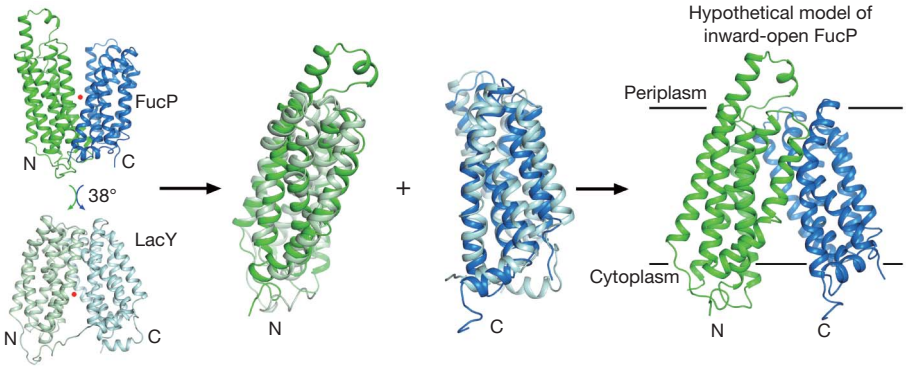


Figure 2 | Alternating access achieved by concentric, rigid-body rotation of the N and C domains of FucP. Structural comparison of FucP and LacY shows that the N (green) and C (blue) domains of the outward-open FucP and the inward-open LacY are structurally similar, suggesting that a conformational switch from outward-open to inward-open may involve only rigid-body rotation of the N and C domains. An inward-open FucP model was constructed on the basis of structural superimposition with LacY[4] (PDB ID, 1PV7). The red dots in the left-most image indicate the axes for concentric movement of the N and C domains.

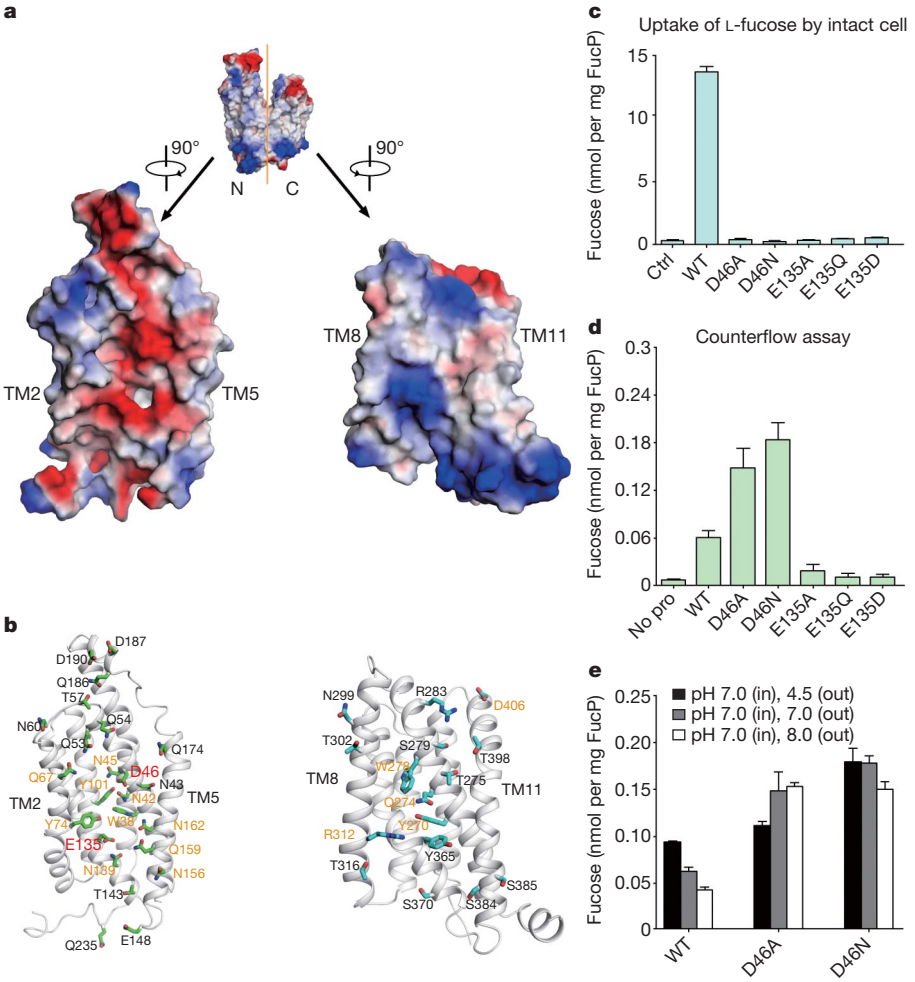


Figure 3 | The cavity-facing sides of the N and C domains have contrasting surface electrostatic potentials. **a**, The surface electrostatic potentials calculated for the N and C domains. **b**, The polar or charged residues lining the cavity from the N and C domains. Residues that are conserved in the five L-fucose transporters (Supplementary Fig. 1) are labelled in orange. Asp 46 and Glu 135 are highlighted in red. **c**, The uptake of L-fucose by *E. coli* BL21(DE3) transformed with FucP variants. Cells transformed with the formate channel FocA in pET21b vector were used as control (ctrl). Details of the experiments can be found in Supplementary Methods. WT, wild type. **d**, Transport activity of FucP variants in the counterflow assay. 'No pro' indicates control experiments with FucP-free liposomes. **e**, Neutralization of Asp 46 led to abrogation of pH dependence in the counterflow assays. Error bars, s.d. of three independent experiments.

Fig. 9d). Supporting this model, mutation of Glu 135 or Asn 162 led to abrogation of interaction between L-fucose and FucP (Supplementary Fig. 8c, d).

We next examined how proton translocation is coupled to L-fucose transport. The presence of only two residues that are capable of protonation or deprotonation along the transport path makes FucP a unique model system in which to study proton coupling. As discussed above, Asp 46 is essential for FucP-mediated active transport (Fig. 3c). Asp 46 Asn, which mimics the irreversibly protonated form of Asp, abolished active transport of L-fucose, probably owing to its loss of ability to be deprotonated. By contrast, Asp 46Asn facilitated substrate exchange in the counterflow assays (Fig. 3d). One potential explanation is that elimination of the negative charge may reduce the energetic barrier to L-fucose entry and/or transport.

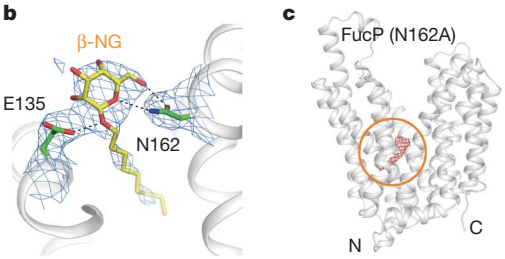
To explore the potential impact of Glu 135 protonation and deprotonation, we performed two parallel molecular dynamics simulations, with Glu 135 either deprotonated or protonated. This analysis revealed that, consistent with our structural observation (Supplementary Fig. 7b), deprotonated Glu 135 formed a stable hydrogen bond with Tyr 365. Once protonated, however, Glu 135 preferred to adopt a distinct rotamer, which is unlikely to maintain the hydrogen bond (Supplementary Fig. 11).

Together, our structural, biochemical and computational analyses suggest a working model of how FucP may mediate the symport of L-fucose and a proton (Fig. 4d). In this model, Asp 46 and Glu 135 have an essential role, coupling proton translocation and substrate recognition. L-fucose can only bind to FucP following the protonation of Asp 46, where the proton neutralizes the negative charge of Asp 46 and lowers the energetic barrier for L-fucose entry and/or transport. Given the locations of Asp 46 and Glu 135, translocation of the proton presumably involves a proton relay mediated by structural water molecules and other polar and/or charged amino acids in FucP (Supplementary Fig. 12a). The presence of the substrate L-fucose may also facilitate the proton relay by coordinating water molecules (Supplementary Fig. 12b). Substrate binding and protonation of Glu 135 is likely to trigger a conformational switch that results in the inward-open conformation of FucP. The transport is complete with the release of the substrate and the deprotonation of Glu 135.

Despite its speculative nature, the proposed model explains the biochemical observations and is supported by structural analyses. Confirmation or disproof of this model requires many more experiments. Nonetheless, the structural and biochemical studies reported here provide insights into the function of FucP and serve as a framework for future investigation.

a

FucP variants	K_d (mM)
WT (Cymal-7)	0.47 ± 0.02
WT (β-NG)	Undetectable
E135Q (Cymal-7)	Undetectable
N162A (Cymal-7)	Undetectable



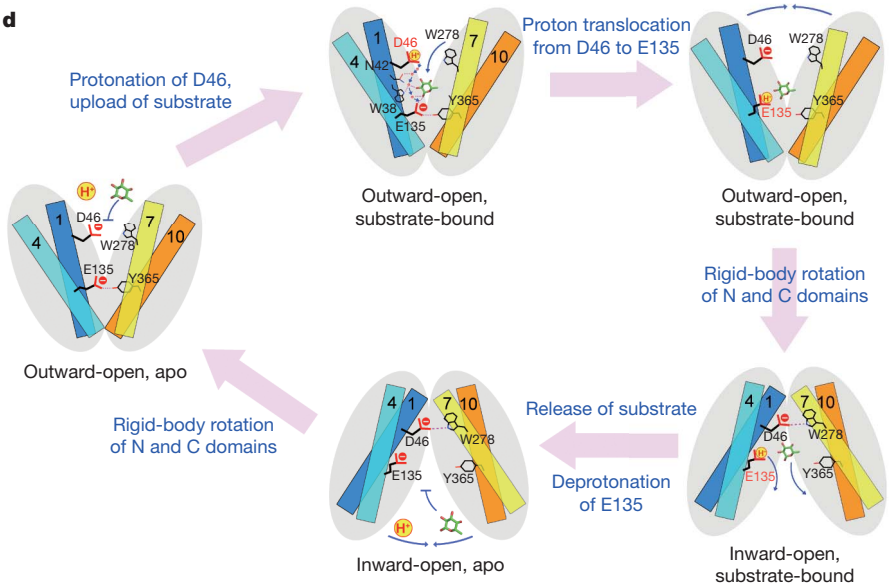


Figure 4 | A working model for the L-fucose/H$^+$ symport by FucP. **a**, Summary of ITC measurement of the binding affinity between L-fucose and FucP variants in the presence of different detergents. The raw data were included in Supplementary Fig. 8. K_d, dissociation constant. **b**, β-NG molecule modelled into a disc-shaped density and refined. The $2F_o - F_c$ electron density map, contoured at 1σ, is shown for Glu 135, Asn 162 and β-NG. **c**, The structure of FucP (Asn 162 Ala) was nearly identical to that of wild-type FucP, except that the electron density for β-NG was better defined. The 'omit' electron density map, shown in red mesh, is contoured at 3σ. **d**, Working model depicting the mechanism of FucP mediated L-fucose/H$^+$ symport. The details of the model are discussed in the main text.

METHODS SUMMARY

The genome of *E. coli* strain BL21(DE3) does not encode FucP. We overexpressed full-length FucP from *E. coli* strain O157:H7 in *E. coli* BL21(DE3) and purified it to homogeneity by affinity chromatography and gel filtration. FucP purified in 0.4% β-NG was crystallized by the hanging-drop vapour-diffusion method. The data sets were collected at Spring-8 beamline BL41XU and Shanghai Synchrotron Radiation Facility beamline BL17U, and processed with HKL2000[23] and the CCP4 suite[24]. We generated the experimental phases by mercury-based, single-wavelength anomalous dispersion using SHELXD[25] and PHASER[26] and improved them using DM[27] and DMMulti[27]. The model was built with COOT[28]. The final model was refined using PHENIX[29]. We determined the structure of FucP (Asn 162 Ala) by molecular replacement with PHASER and refined it with PHENIX. The binding affinities between L-fucose and FucP variants were measured by ITC. The *in vivo* uptake assay and *in vitro* counterflow assays were performed as described in Supplementary Methods. All the reactions were 30 s in duration if not otherwise indicated.

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

Received 12 February; accepted 12 August 2010.
Published online 26 September 2010.

1. Pao, S. S., Paulsen, I. T. & Saier, M. H. Jr. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* **62,** 1–34 (1998).
2. Bradley, S. A., Tinsley, C. R., Muiry, J. A. & Henderson, P. J. Proton-linked L-fucose transport in *Escherichia coli. Biochem. J.* **248,** 495–500 (1987).
3. Gunn, F. J., Tate, C. G. & Henderson, P. J. Identification of a novel sugar-H$^+$ symport protein, FucP, for transport of L-fucose into *Escherichia coli. Mol. Microbiol.* **12,** 799–809 (1994).
4. Abramson, J. *et al.* Structure and mechanism of the lactose permease of *Escherichia coli. Science* **301,** 610–615 (2003).
5. Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N. Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli. Science* **301,** 616–620 (2003).
6. Yin, Y., He, X., Szewczyk, P., Nguyen, T. & Chang, G. Structure of the multidrug transporter EmrD from *Escherichia coli. Science* **312,** 741–744 (2006).
7. Becker, D. J. & Lowe, J. B. Fucose: biosynthesis and biological function in mammals. *Glycobiology* **13,** 41R–53R (2003).
8. Paulsen, I. T., Chauvaux, S., Choi, P. & Saier, M. H. Jr. Characterization of glucose-specific catabolite repression-resistant mutants of *Bacillus subtilis*: identification of a novel hexose:H$^+$ symporter. *J. Bacteriol.* **180,** 498–504 (1998).
9. Christensen, M. *et al.* Regulation of expression of the 2-deoxy-D-ribose utilization regulon, *deoQKPX*, from *Salmonella enterica* serovar typhimurium. *J. Bacteriol.* **185,** 6042–6050 (2003).
10. Psakis, G. *et al.* The sodium-dependent D-glucose transport protein of *Helicobacter pylori. Mol. Microbiol.* **71,** 391–403 (2009).
11. Horiba, N. *et al.* Cloning and characterization of a novel Na$^+$-dependent glucose transporter (NaGLT1) in rat kidney. *J. Biol. Chem.* **278,** 14669–14676 (2003).
12. Law, C. J., Maloney, P. C. & Wang, D. N. Ins and outs of major facilitator superfamily antiporters. *Annu. Rev. Microbiol.* **62,** 289–305 (2008).
13. Guan, L. & Kaback, H. R. Lessons from lactose permease. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 67–91 (2006).
14. Smirnova, I., Kasho, V., Sugihara, J. & Kaback, H. R. Probing of the rates of alternating access in LacY with Trp fluorescence. *Proc. Natl Acad. Sci. USA* **106,** 21561–21566 (2009).
15. Kaback, H. R., Sahin-Toth, M. & Weinglass, A. B. The kamikaze approach to membrane transport. *Nature Rev. Mol. Cell Biol.* **2,** 610–620 (2001).
16. Kaback, H. R. Structure and mechanism of the lactose permease. *C. R. Biol.* **328,** 557–567 (2005).
17. Guan, L., Mirza, O., Verner, G., Iwata, S. & Kaback, H. R. Structural determination of wild-type lactose permease. *Proc. Natl Acad. Sci. USA* **104,** 15294–15298 (2007).

18. Mirza, O., Guan, L., Verner, G., Iwata, S. & Kaback, H. R. Structural evidence for induced fit and a mechanism for sugar/H^+ symport in LacY. *EMBO J.* **25,** 1177–1183 (2006).
19. Gunn, F. J., Tate, C. G., Sansom, C. E. & Henderson, P. J. Topological analyses of the L-fucose-H^+ symport protein, FucP, from *Escherichia coli. Mol. Microbiol.* **15,** 771–783 (1995).
20. Franco, P. J. & Brooker, R. J. Functional roles of Glu-269 and Glu-325 within the lactose permease of *Escherichia coli. J. Biol. Chem.* **269,** 7379–7386 (1994).
21. Guan, L. & Kaback, H. R. Binding affinity of lactose permease is not altered by the H^+ electrochemical gradient. *Proc. Natl Acad. Sci. USA* **101,** 12148–12152 (2004).
22. Quick, M. *et al.* Binding of an octylglucoside detergent molecule in the second substrate (S2) site of LeuT establishes an inhibitor-bound conformation. *Proc. Natl Acad. Sci. USA* **106,** 5563–5568 (2009).
23. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
24. Collaborative Computational Project, Number 4. The *CCP4* suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
25. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58,** 1772–1779 (2002).
26. McCoy, A. J. *et al. Phaser* crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
27. Cowtan, K. DM: an automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **31,** 34–38 (1994).
28. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
29. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
30. DeLano, W. L. *PyMOL Molecular Viewer* ⟨http://www.pymol.org⟩ (2002).

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

Acknowledgements We thank N. Shimizu and T. Kumasaka at Spring-8 beamline BL41XU in Japan, and J. He and S. Huang at the Shanghai Synchrotron Radiation Facility (SSRF) for on-site assistance. This work was supported by funds from the Ministry of Science and Technology (grant number 2009CB918802) and by Tsinghua University 985 Phase II funds. N.Y. acknowledges support from the Yuyuan Foundation and the Li Foundation.

Author Contributions N.Y. designed the experiments. S.D., L.S., Y.H., F.L. and Y.L. performed the experiments and analysed data. H.G., J.W. and N.Y. analysed data. N.Y. wrote the manuscript.

Author Information The atomic coordinates and structure factors of wild-type FucP and FucP (Asn 162 Ala) have been deposited in the Protein Data Bank under the accession codes 3O7Q and 3O7P, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.W. (jwwang@tsinghua.edu.cn) and N.Y. (nyan@tsinghua.edu.cn).

METHODS

Protein preparation. The complementary DNA of full-length FucP from *E. coli* strain O157:H7 was subcloned into pET21b (Novagen). Overexpression of FucP was induced in *E. coli* BL21(DE3) by 0.2 mM isopropyl β-D-thiogalactoside (IPTG) when the cell density reached OD_{600} 0.8. After growth at 37 °C for 4 h, the cells were collected, homogenized in the buffer containing 25 mM Tris-HCl (pH 8.0) and 150 mM NaCl, and disrupted using a French press with two passes at 10,000–15,000 p.s.i. Cell debris was removed by low-speed centrifugation for 10 min. The supernatant was collected and ultracentrifuged at 150,000*g* for 1 h. Membrane fraction was collected and incubated with 1.5% (w/v) dodecyl-β-D-maltopyranoside (DDM, Anatrace) for 1 h at 4 °C. After another ultracentrifugation step at 150,000*g* for 30 min, the supernatant was collected and loaded on Ni^{2+}-nitrilotriacetate affinity resin (Ni-NTA, Qiagen) and washed with 25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 20 mM imidazole and 0.02% DDM. The protein was eluted from the affinity resin with 25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 250 mM imidazole and 0.02% DDM, and concentrated to about 10 mg ml^{-1} before further purification by gel filtration (Superdex-200 10/30, GE Healthcare). The buffer for gel filtration contained 25 mM Tris-HCl (pH 8.0), 150 mM NaCl and detergents. The peak fractions were collected.

The FucP mutants were generated with a standard PCR-based strategy and were subcloned, overexpressed and purified in the same way as the wild-type protein.

Crystallization. We performed extensive crystallization trials for FucP proteins purified in more than 20 different detergents. Crystals were grown at 18 °C by the hanging-drop vapour-diffusion method. The full-length wild-type FucP with C-terminal His_6 tag purified in 0.4% *n*-nonyl-β-D-glucopyranoside (β-NG, Anatrace) gave rise to crystals in multiple polyethylene glycol (PEG) conditions. The crystals tested at an X-ray home source diffracted to about 8–10 Å. After optimization, FucP crystals appeared overnight in the well buffer containing 2 mM L-fucose, 30% (w/v) PEG400, 0.1 M MES (pH 6.0) and 0.1 M $MgCl_2$, and grew into tetragonal rods of typically 50 μm × 50 μm × 500 μm in about two weeks. The crystals diffracted to 6 Å at an X-ray home source and 4 Å in a synchrotron. To further improve the resolution, we performed extensive screening for small molecule additives. Finally, 50 mM NaF was shown to be the key to improving the diffraction to about 3.1 Å at Spring-8 beamline BL41XU. The crystals belong to space group $P3_221$, with unit-cell dimensions of $a = b = 119.46$ Å and $c = 98.36$ Å. Crystals were also obtained in the absence of L-fucose under similar conditions. However, these crystals consistently diffracted X-rays to lower resolutions and the best diffraction was obtained at 3.3 Å resolution after more than 300 crystals were tested at SSRF beamline BL17U. Derivative crystals were obtained by soaking crystals for 10 min in mother liquor containing 1 mg ml^{-1} $HgCN_2$, and then back-soaking for 1 min in well buffer plus β-NG. Both native and heavy-atom-derived crystals were directly flash frozen in a cold nitrogen stream at 100 K. The crystals of FucP (N162A) were obtained in similar conditions to wild-type FucP in the absence of L-fucose.

Data collection and structure determination. The data sets were collected at Spring-8 beamline BL41XU or SSRF beamline BL17U, and processed with HKL2000[23]. Further processing was carried out using programs from the CCP4 suite[24]. Data collection statistics are summarized in Supplementary Tables 1 and 2.

The mercury sites were located using SHELXD[25] from the Bijvoet differences in the mercury-based, single-wavelength anomalous dispersion (Hg-SAD) data. The identified positions were refined and phases were calculated using the PHASER SAD experimental phasing module[27]. The real-space constraints, including solvent flattening and histogram matching, were applied to the electron density map in DM[27]. Cross-crystal averaging in DMMulti[27] gave rise to electron density maps of sufficient quality for model building, using the Hg-SAD and native data. The initial model was built in COOT[28] manually. The structure of FucP (N162A) was determined by molecular replacement with PHASER[26]. The final model was refined using PHENIX[29].

For wild-type FucP, 89.0% and 11.0% of the amino acids in the final atomic model reside in the most favourable and the allowed regions of the Ramachandran plots, respectively. For FucP (N162A), 86.0% and 14.0% of the amino acids reside in the most favourable and allowed regions of the Ramachandran plots, respectively.

Cell-based L-fucose uptake assay. FucP is not encoded by the genome of *E. coli* strain BL21(DE3). The cells transformed with pET21b-FucP plasmids were grown in LB medium at 37 °C. Cells were collected at OD_{600} ~1.0. After being rinsed with MK buffer (5 mM MES (pH 6.5) and 150 mM KCl) three times, the cells were resuspended in the same buffer at a density of OD_{600} ~2.0. Cells were energized by the addition of glycerol at a final concentration of 20 mM before the uptake assay. After incubation at 25 °C for 3 min, L-[5,6-^{3}H]-fucose (American Radiolabelled Chemicals) was added at a final concentration of 0.17 μM. All the reactions lasted 30 s and were stopped by diluting the cell suspension with 2 ml ice-cold MK buffer. Cells were collected by filtering through a 0.22-μm cellulose acetate filter (Sartorius).The filter membranes were subsequently washed with another 4 ml ice-cold MK buffer, dried and then taken for liquid scintillation counting. All the experiments were repeated at least three times.

The expression levels of FucP variants were monitored by western blot using an antibody against the histidine tag. The amount of FucP protein was estimated by comparing the intensity of the protein against a serial dilution of FucP with known concentration on the same western blot. Cells transformed with the formate channel FocA in pET21b vector were used as a control to exclude the possibility that the uptake of L-[5,6-^{3}H]-fucose was caused by change of cell membrane due to the overexpression of non-specific membrane proteins.

Preparation of liposomes and proteoliposomes. Liposomes of *E. coli* polar lipid (Avanti) were prepared using a standard protocol as described previously[31]. Proteoliposomes for counterflow assay were prepared in a solution containing KPM buffer (50 mM potassium phosphate, 2 mM $MgSO_4$, pH adjusted as indicated), 20 mg ml^{-1} pre-extruded liposome, 1.25% β-OG (Anatrace), 20 mM L-fucose, and wild-type or mutant FucP proteins at a concentration of 10 μg protein per milligram of lipids. β-OG was removed by incubation with 400 mg ml^{-1} Bio-Beads SM2 (Bio-Rad) overnight. After the removal of β-OG, the proteoliposomes were frozen in liquid nitrogen and thawed at 22 °C for five to ten cycles. After another extrusion through a 400-nm membrane filter, the proteoliposomes were collected by ultracentrifugation at 100,000*g* for 1 h and washed twice with ice-cold KPM buffer to remove the extra L-fucose. The proteoliposomes were resuspended in ice-cold KPM buffer to a final concentration of 100 mg ml^{-1} and immediately underwent counterflow assay.

Counterflow assay. All counterflow assays were carried out at 25 °C. Concentrated proteoliposomes (2 μl) were diluted into 100 μl assay buffer containing 1 μCi [^{3}H]-L-fucose (specific activity, 60 Ci $mmol^{-1}$). The final concentration of [^{3}H]-L-fucose in the external buffer is about 0.17 μM. The uptake of [^{3}H]-L-fucose was stopped at indicated time points by diluting the reaction solution with 2 ml ice-cold buffer and rapidly filtering through 0.22-μm filters (Millipore). After filtration, the filter membranes were washed with another 4 ml ice-cold reaction buffer without fucose. The filter was then taken for liquid scintillation counting. All the experiments were repeated for at least three times. The reactions lasted 30 s if not otherwise stated.

Molecular dynamics simulation. The crystal structure of FucP bound with L-fucose was inserted into a POPE bilayer to mimic the *E. coli* inner membrane. The system was solvated with TIP3P water molecules and the periplasmic vestibule was filled with water molecules. NaCl (100 mM) was then added to neutralize the system. Asp 46 and Glu 135 were designed to be protonated or deprotonated as mentioned in the main text. The other ionizable residues, except histidines, were modelled in their charged forms. The whole system was then simulated using the molecular dynamics program NAMD[32] at a constant temperature of 310 K and pressure of 1 atm, maintained with a Langevin thermostat and barostat, using the CHARMM27 force field including proteins and lipids[33]. Parameters for L-fucose were provided by O. Guvench and A. D. MacKerell, Jr. The electrostatic interaction was calculated using the particle-mesh Ewald method. The system was simulated first with constraints on the protein atoms. After releasing the constraints, the system was equilibrated for 1 ns and simulated for another 5 ns for productive analyses.

Isothermal titration calorimetry. The binding affinity between L-fucose and FucP variants was measured by ITC. Full-length wild-type or mutant FucP was extracted and purified through Ni-NTA resin in buffer containing 0.02% DDM. The eluted FucP was concentrated and underwent size-exclusion chromatography (Superdex 200 10/30, GE Healthcare) in the buffer containing either 0.02% Cymal-7 (Anatrace) or 0.4% β-NG. The peak fraction was pooled for ITC titration. The protein was adjusted to about 0.1 mM and was titrated with 10 mM L-fucose dissolved in the same buffer as used in the size-exclusion chromatography. All experiments were performed with a VP-ITC microcalorimeter (MicroCal) at 22 °C. The data were fitted using the software ORIGIN 7.0 (Origin Lab).

31. Veenhoff, L. M. & Poolman, B. Substrate recognition at the cytoplasmic and extracellular binding site of the lactose transport protein of Streptococcus thermophilus. *J. Biol. Chem.* **274,** 33244–33250 (1999).
32. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).
33. Brooks, B. R. *et al.* CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30,** 1545–1614 (2009).

CAREERS

POSTDOCS US Congress recognizes National Postdoc Appreciation Week **p.741**

FUNDING Howard Hughes Medical Institute launches plant-science award **p.741**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

L. HENDERSON/IMAGES.COM/CORBIS

The spread of postdoc unions

Unionizing has become more common at US campuses, but postdocs should weigh up the pros and cons of membership.

BY VIRGINIA GEWIN

Bruce Adams, a postdoctoral researcher at the Diabetes Center at the University of California, San Francisco, isn't anti-union. But like many postdocs in the University of California (UC) system, he voiced concerns about the United Auto Workers (UAW) union during its initial attempt to become the representative for the roughly 6,400 postdocs at ten UC campuses in 2006. Adams, who helped to form a campus organization to air postdocs' trepidations about unionization, says dozens of people formally complained that they were given inaccurate or confusing information when approached to sign a card authorizing the UAW's representation.

Matthew O'Connor, then a UAW spokesman at UC Berkeley, blames a lack of communication for the confusion that was rampant four years ago. "We weren't prepared to deal with the misinformation," he says. That first unionization attempt failed, largely because the university pushed to reclassify many postdocs so that they were not technically 'employees'. O'Connor credits the UAW's successful unionization attempt in 2008 to an informative website and a more coordinated canvassing effort.

Two years later in August, the UC postdoc union, now known as Postdoctoral Researchers Organize/UAW (PRO/UAW), negotiated a contract with salary increases, retirement contributions and safeguards governing terms of employment for UC postdocs. The agreement, celebrated by many UC postdocs, sent shock waves through research institutions around North America — raising the profile of postdoc grievances, spreading a fear of unionization, and, in some universities, prompting administrations to proactively address postdoc inequities.

Despite the benefits unionization provides, many postdocs, such as Adams, remain sceptical. Some are concerned about possible tension between mentors and mentees that could result from unionization. Others worry about whether the unions will be able to deliver contract promises, and the possibility that mandated higher salaries will, ultimately, strain research grants, forcing funders or principal investigators to make tough choices.

"In this environment, it was pretty natural to say unionization is the way to have a voice."

Others, particularly university administrators, question the model itself. Traditionally designed to protect career positions, unions might not be appropriate for temporary apprentices who are tied more to a specific laboratory than to an institution. Regardless, postdocs should be careful to balance the benefits with the downsides — and understand what they're signing up for (see 'Membership pros and cons').

STATE OF THE UNION

It was perhaps inevitable that postdocs would explore the advantages of unionization. On many campuses, postdocs are effectively a voiceless group of highly trained researchers, working alongside unionized graduate teaching and research assistants. In view of postdocs' variable pay — due to the disparate funding streams used to pay the surging number of postdocs (the number in the United States has almost tripled in the past 30 years to an estimated 90,000) — the interest in unions should not come as a surprise.

Meanwhile, the postdoc position worldwide is changing. What was once a 1–2 ▶

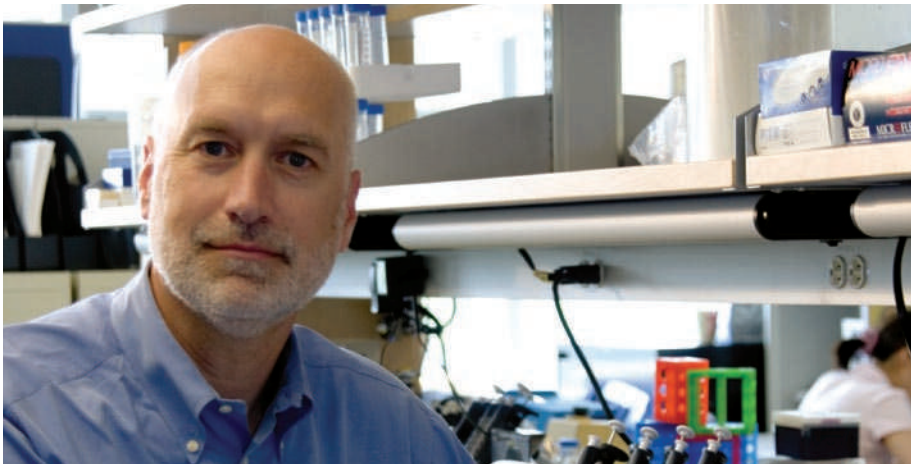
year training experience is increasingly morphing into a bona fide job lasting several years. Many postdocs are starting families, buying homes and initiating retirement funds during the same 2–5 years that they are toiling away — while competing with each other to become one of the lucky ones to secure a tenure-track job. Vulnerable, they want compensation and protection.

Thus far, postdoc unions have come about because of existing university union organizations, and the efforts of dedicated and determined postdocs. At the University of Massachusetts campuses at Amherst, postdocs were the only group of employees not unionized, and university management is used to responding to grievances from union representatives, so conditions were favourable for unionization. Last February, most of the approximately 300 postdocs at Amherst agreed to form a union with the UAW, and are now beginning contract negotiations. "In this environment, it was pretty natural to say unionization is the way to have a voice," says James Staros, provost and vice-chancellor for academic affairs at the University of Massachusetts, Amherst. "It'll be a success here because it fits our culture."

On other campuses, postdocs avoid pursuing unionization as it poses a culture clash. Instead, some appeal directly to university administrators to get their needs met. Daniel Boutz, a postdoc and volunteer organizer of the biological sciences' postdoctoral association at the University of Austin at Texas, has been working for years, with modest success, to get postdoc inequities addressed by the administration. Boutz contends that their best bet is to present requested changes as a wise investment that will benefit university research.

It is too early to tell whether unionization will be a success at Rutgers University in New Brunswick, New Jersey. Although postdocs there have agreed to form a union, they and the administration have yet to hammer out a contract. Faculty members and graduate students were already unionized, but administrators were caught off guard when postdocs formed a union with the American Federation of Teachers (AFT) in July 2009. "The administration was gradually moving towards forming an office dedicated to postdoc issues, so the move to unionize caught us by surprise," says Jerome Kukor, a microbiologist and dean of academic programmes and research at Rutgers University.

Former UC graduate students sparked the California movement after finding they had fewer protections and benefits when they became postdocs. The marriage with the UAW may have been one of convenience, but it was also one of the few unions able to execute a large operation over several campuses. "They were already on UC campuses representing graduate students and they were one of the only unions that had the resources — money and know-how — to do it," says O'Connor.



L. GOINGS, UNIV. MICHIGAN

Peter Hitchcock says university administrators should proactively address postdocs' concerns.

Postdoc status can have a big impact on gathering union support. After postdocs at the University of Western Ontario, Canada, agreed in 2008 to unionize following three years of unsuccessfully lobbying for improved benefits and services, the university required that all postdocs be classified either as fellows (independent contractors) or associates (university employees). Most were classified as fellows, which, at the time, meant their salaries would not be taxed. This meant that the number of postdocs labelled as employees and thus eligible to join a union decreased from 236 to 30 — effectively stalling negotiations.

In an unfortunate twist a year later, the Canadian Revenue Agency decided that all postdocs should be subjected to taxation, just like employees (see *Nature* doi:10.1038/news.2010.429; 2010). As a result, some postdocs now take home less pay than graduate students who are not taxed — a fact many postdocs lament. "It's insulting to get paid less than the graduate students we teach," says Marianne Stanford, chair of the Canadian Association of Postdoctoral Scholars and a postdoc at the University of Ottawa's Health Research Institute. With talks stalled, postdocs at the University of Western Ontario are waiting until the Ontario Labour Relations Board rules whether all postdocs, both fellows and associates, will be considered part of the union.

In the case of the PRO/UAW, motivated postdocs pushed the university system to reach an agreement. It had been more than a year, and they still had no contract. Postdocs on all ten UC campuses picketed, saying that the university should not use the state budget crisis as an excuse to delay an agreement. The uproar led to a state congressional hearing in April. The August agreement includes many of the postdocs' demands, although they had to accept a graduated pay increase to avoid any potential cash-flow problems in individual research labs — something that could, inadvertently, have led to postdoc layoffs, says O'Connor.

CASE STUDY

Membership pros and cons

Potential advantages

- Pay increase.
- Equal access to health insurance.
- Defined, standardized terms of holiday and maternity leave.
- A formal mechanism to air grievances.

Potential disadvantages

- Union dues — in the case of the University of California PRO/UAW union, dues represent one-third of the pay rise that most will receive (1% of the 3%).
- Existing research grant budgets may have to be altered to pay for salary increases.
- Strained mentor–mentee relations.
- Possibility of a strike that could disrupt postdoc priorities.

UNION UPS AND DOWNS

Pay increases, sought by many pro-union postdocs, may have complex implications. Under the new contract, UC postdocs earning less than US$47,000, the majority, will receive a 3% raise; those earning more than $47,000 will receive a 1.5% raise. In exchange, the union receives 0.85% and 1.15% of the salaries of non-union and union members, respectively. Paying dues was the main downside postdocs raised about unionization. For their efforts, unions will receive roughly one-third of the salary increases.

The pay increase itself will cost the UC $8.4 million for 6,420 postdocs, according to the UC Office of the President — money that will come from principal investigator grants or department funds, not from university overheads, according to a 21 September e-mail by Graham Fleming, vice-chancellor for

research at UC Berkeley. Even before that e-mail, some faculty members were concerned that unionization would, ultimately, lead to fewer postdoc posts and fewer opportunities for graduate students.

As salaries increase, in theory, less money is available for other grant expenditures. This could hit some grant budgets harder than others. For example, the average US National Science Foundation award is significantly less than that of the National Institutes of Health. The National Science Foundation intends to reassess its grant proposal guidelines if the need arises, according to Dana Topousis, an agency spokeswoman. She adds that grantees do have some flexibility to transfer funds from one budget category to another, which could help.

Principal investigators might have to make difficult staffing changes to meet budgets. "If you make the postdoc stipend prohibitively high, principal investigators will opt to hire more graduate students," says Carlito Lebrilla, a chemist at UC Davis who is tracking union developments. "Some people see unionization as devastating to the research endeavour." There is also the question of whether the more senior, successful principal investigators will be better equipped to absorb extra postdoc costs than newer investigators.

"Some people see unionization as devastating to the research endeavour."
Carlito Lebrilla

Others fear that union-sanctioned activities will interfere with their postdoc responsibilities — being asked to strike being the most obvious example. But union supporters and non-supporters alike agree that striking — the ultimate bargaining chip of unions — isn't realistic for postdocs. Postdocs need to publish papers to have a chance of obtaining a faculty job; few postdocs are likely to adhere to a strike if the union called for one. The final UC postdoc contract stipulates that postdocs will not strike, taking a risky strategy off the table. "We didn't want to have postdocs put that kind of pressure on themselves," says Norval Hickman, a union supporter and clinical psychology postdoc at UC San Francisco. "We need a good letter of recommendation to get that first independent academic position — and there were definitely concerns that a strike would create an environment that might make that difficult."

Union contracts could present other problems. Kukor hopes that Rutgers' negotiations don't affect the flexible way in which many faculty staff mentor postdocs. "Each postdoc I've worked with over the past 25 years is unique, and I work to accommodate their individual needs," he says. "I'm concerned that a standardized contract may remove some of the degrees of freedom I use to tailor career development."

SOFT POWER

The prospect of unionization has prompted many academic institutions to actively address postdoc concerns — potentially obviating unionization completely. Many institutions that train postdocs hope to avoid unions by recognizing postdocs' value and treating them fairly, says Peter Hitchcock, director of the office of postdoctoral studies at the University of Michigan Medical School in Ann Arbor. At his university, that includes providing a range of career development services — including workshops on how to set up a laboratory and opportunities to learn about career options outside academia — intended to help postdocs to realize their career goals. Megan Ballinger, president of the University of Michigan postdoctoral association, notes that unionization isn't an issue at the medical school, despite Michigan's automotive and union legacy.

Administrators don't necessarily need the pressure of union provisions to spur extraordinary measures. Consider the case of Traci Lyons. In 2007, she was only seven days into a postdoc appointment at the University of Colorado in Denver when she discovered she was pregnant. Because the school had no maternity-leave policy in place, her mentor worked with the legal department to establish a 45-day maternity leave specific to Lyons. A year later, Lyons was asked to help the administration draft a much-needed set of postdoc policies — including guidelines for minimum salaries, terms of employment and maternity leave. John Freed, dean of the graduate school of the Anschutz Medical Campus of the University of Denver, oversaw the creation of an office of postdoctoral affairs and helped secure the $250,000 necessary to ensure that all postdocs on campus received heath care and disability benefits. O'Connor, however, notes that without a union-backed contract, a university could reverse its decisions at any time.

Not every campus will be proactive and take such steps. Budding scientists would be wise to gauge the changing landscape of postdoc unionization. And institutions would be wise to anticipate postdoc needs (see *Nature* **467**, 624; 2010). O'Connor says he has received many requests from postdocs and postdoc associations for information about how the UC unionized. "Postdocs are essential to an institution's research efforts," says Hitchcock. "Institutions that do not acknowledge that do so at their own risk." ■

Virginia Gewin *is a freelance writer in Portland, Oregon.*

POSTDOCS

Official appreciation

National Postdoc Appreciation Week was officially recognized by the US Congress on 23 September. The move confers no funding and creates no laws. But it is vital for building awareness of postdocs' research contributions as well as of their compensation issues and other woes, says Cathee Johnson Phillips, executive director of the National Postdoc Association (NPA) in Washington DC. The NPA, along with institutions nationwide, began the week in 2009. This year, it was 20–24 September and more than 70 institutions took part. In the past, most legislators knew little about postdocs, says Phillips. "Now, when our members talk to them about compensation and benefits, they'll say, 'Let's talk,' not 'What's a postdoc?'" she says.

FUNDING

Plant grants announced

To boost interest in plant research, the Howard Hughes Medical Institute (HHMI) in Chevy Chase, Maryland, will for the first time fund plant-science investigator awards. The HHMI and its co-funder, the Gordon and Betty Moore Foundation in Palo Alto, California, will allocate up to 15 awards worth a total of US$75 million. Plant science is underfunded in the United States, says Robert Tjian, the institute's president, adding that plants serve as useful experimental models for biomedical research. Awardees must have run their own lab for at least four years. They will receive a five-year appointment that may be renewed indefinitely for further five-year terms. The deadline for applications is 9 November.

AWARDS

Travel for collaboration

To promote cross-disciplinary research and collaboration, the Burroughs Wellcome Fund in Research Triangle Park, North Carolina, has created travel awards of up to US$15,000 each for US and Canadian citizens or permanent residents. Applicants should hold or be pursuing a PhD in chemistry, physics, mathematics, computer science, statistics or engineering, and want to investigate biological questions with another lab or through attending a course. Biologists looking to collaborate with, or learn from, a physical scientist can also apply. The funds are for travel expenses anywhere in the world by the end of 2012. The deadline is 1 December 2010; winners will be announced by 1 March 2011.

WAR OF THE ROSES

The high price of fashion.

BY POLENTH BLAKE

This year, it was cottages. Chrome cube houses had been all the rage last year, but after the success of the Back-to-Nature gene mod range, everyone wanted an idyllic cottage.

Dave McKillen was one of the few who managed to get one. The building was a timber-framed construction and came with a rose garden. It was even on the top level of the city, so Dave could sit in the garden and photosynthesize on clear days. Most people with Back-to-Nature leaf skin had to rely on UV lamps. Sunlight was real nature.

Dave stood on the garden path and breathed in the scent of roses. He was fashionable again.

"Can I play?" Little D tugged at Dave's jeans.

Dave smiled down at his young clone. "Sure. Mind the thorns."

Little D tottered towards the rose bushes. Dave turned back to the workers, who were lurking around the removal van. "Get a move on."

Most of the furniture was inside when Little D came back. "They is plain."

"Are plain," Dave corrected. Sometimes he wondered if the lab had stiffed him on the intelligence upgrades for Little D. A one-year-old shouldn't be making those kinds of mistakes.

"No glowies either." The child looked at his shoes sadly.

"How about we log on to the market and find some?"

Little D brightened immediately.

The garden transformed over the next few weeks. There weren't many roses on the market the first time Dave looked. Hickly Labs produced a few patterned varieties, such as the zany zebra and pink polka, but that was all.

J&D Genetics had their rose out the next day, claiming their furry flower was 'like planting a kitten in your garden'.

The moonglow rose, 'the rose that glows', was Mythic Gene's first attack on the market. The adverts went out hours after J&D's announcement.

The market war had begun.

Dave collected seeds from the old roses before digging them up. He didn't get rich by being wasteful — traditional varieties might be in next year. Once the bushes were gone, the garden was ready for its makeover.

Hundreds of new varieties were out by the time the garden was clear. Dave chose the trendiest, although he did make some concessions for Little D. Jingle jigglers weren't high fashion, but the boy loved the way they trembled whenever someone was near. Once the flowers opened, each bloom jingled like a bell.

"Fairies live inside," said Little D, prodding the jiggler bloom to hear it ring.

"Fairies aren't real yet." Dave made a note to get Little D a fairy, if Mythic Gene ever managed to get them approved. After the unfortunate incident with the first dragon pet, the government wasn't too enthusiastic about licensing any more mythical creatures.

Dave settled on a sun lounger to doze as Little D explored.

He was woken by Little D tapping him. "I find something."

"Daddy's photosynthesizing."

"It's important," said Little D.

Dave followed the boy around the side of the house. Little D crouched down by a small zany zebra bush. He prodded a flower and it jingled.

It took a moment for Dave to realize what had happened: the roses must have reproduced the old-fashioned way.

"That's some find. Well done."

Little D beamed. "Can I name him?"

"Sure."

"I call him Stinkyface."

Stinkyface was soon joined by other hybrids. Jilly was a moonglow rose with jiggling flowers. Plop had zany zebra stripes and polka spots. A new hybrid seemed to pop up every day, until Little D found Sally.

She was an odd-looking rose. Her petals were shrivelled and she smelt faintly of rotten fish. Dave wouldn't have thought she was a rose, apart from a hint of moonglow light after dark.

The other roses were dead by the end of the week.

"Nothing wrong with the soil," the plant doctor said.

Dave looked at the rows of dead bushes. "Maybe the roses are faulty?"

"No one else's have died." There was an edge to the doctor's voice. Dave knew who paid her wages.

"What if they're dangerous …"

"Sir, moonglow roses have been fed to mice with no ill effects. Jingle jigglers were crossed with 23 cultivated varieties of rose to ensure they wouldn't harm the wild populations.

JACEY

They're safe for everyone, plant, animal or otherwise."

She gave a pointed look at his green skin at 'otherwise'.

"There must be something," he said.

"I've taken a sample from the survivor. We'll let you know in a week."

She packed her testing kit away in a briefcase and strode out of the garden.

The lab report's only caution was legal. They'd found 32 patented genes in Sally's genome, so it wouldn't be legal to resell her. Other than that, she was healthy and completely harmless.

Within a month, there were no roses left on the top level, apart from a few Sally bushes here and there.

The news feeds exploded. The gene companies blamed it on an unknown disease. Dave knew better. He'd been watching the bees. Normal bees wouldn't go near Sally, but the modded city bees weren't so picky about scent. Every time a non-Sally flowered, it was only a matter of time before the poisoned pollen reached it.

Dave went out and joined Little D in the garden. The child sat crying by a dead jingle jiggler.

"Hey," said Dave. "I've got something for you."

Little D looked up through his tears.

Dave handed him the box of old rose seeds.

"I plant them?" said Little D.

"Not yet, but one day."

The day Sally died. ■

Polenth Blake *(http://www.polenthblake.com) enjoys a spot of gardening, as long as the plants leave each other alone.*