# THIS WEEK

EDITORIALS

**WARFARE** Science and conflict a century after the First World War **p.132**

**WORLD VIEW** Antarctic scientist Chris Turney defends expedition **p.133**

**SMOKE SIGNAL** Nicotine breath helps hornworm repel predators **p.135**

# Data sharing will pay dividends

*As public pressure builds for drug companies to make more results available from clinical trials, the industry should not forget that it relies on collective goodwill to test new therapies.*

Readers of *Nature* who are familiar with recent controversies surrounding clinical trials and medical practice may find it bizarre that anyone could be "surprised and concerned to discover that information is routinely withheld from doctors and researchers about the methods and results of clinical trials", as stated in a UK government report last week.

After all, campaigning doctors have warned for years that pharmaceutical companies have in the past concealed data that reflected poorly on their drugs. Regulators — notably the London-based European Medicines Agency — have pushed for more information to be released into the public domain. And the drug industry itself has moved to open its private data vaults, albeit not by as much or as quickly as campaigners would like.

The 'surprise' and 'concern' at this well-chronicled behaviour of big pharma comes in a report from the UK Parliament's Committee of Public Accounts, a cross-party group with the remit to scrutinize whatever public-expenditure topic catches its eye. The report focuses on the antiviral drug Tamiflu (oseltamivir) — used to treat influenza — which the UK government and others have stockpiled at great expense owing to concerns about pandemic flu.

Independent scientists who want to investigate whether Tamiflu works have struggled to find all the information they need, in part because its manufacturer Roche, headquartered in Basel, Switzerland, held back some details of trials. The company says that it has now released all its Tamiflu data, and researchers are working their way through them, but the case has become a high-profile example of the need for greater openness in biomedical science.

The report made headlines in the United Kingdom and, according to campaigners, vindicated their position. Its call for the UK government to improve the availability of clinical-trial information is also in line with initiatives elsewhere.

For instance, the European Medicines Agency is already pushing forward plans to release the clinical-trial data that drug companies submit to it when seeking approval for their products. Legal action taken against the agency by two aggrieved companies has slowed progress. But in December, the body again stated its "firm commitment to pursuing the objective of full transparency regarding clinical trial data".

Meanwhile, the European Union is overhauling its clinical-trial legislation. Final agreement is some months away, but it seems likely that pharmaceutical firms will eventually be required to upload at least a summary of all their trials to a publicly accessible website.

The two major drug-industry associations in Europe and North America have also moved towards openness. Their joint policy on access to clinical-trial data came into force at the start of this year, adding to promises made by individual companies such as GlaxoSmithKline, Sanofi and Roche to share more data on their drugs.

Yet critics — notably the AllTrials campaign group — say that the industry is not going far enough. Their main complaints are that transparency initiatives are not retrospective and do not require the inclusion of data from older trials, and that some in industry want to act as information gatekeepers, determining which researchers have a genuine need for their expensively assembled data sets. These campaigners deserve credit for raising the issue and for their perseverance in pushing for change.

The industry is at a crossroads. As the UK committee's report shows, concern over the behaviour of pharmaceutical firms is growing. Any anger over the industry's perceived underhand tactics when it comes to data transparency could spread from the vociferous — but small — community of politically active medics and policy campaigners to the wider public.

***"Recent history is full of examples of the public turning against businesses with essential roles."***

Drug companies perform a vital function for society in fighting disease and preserving public health. But recent history is full of examples of the public turning against businesses with essential roles. From banks and energy firms to the oil industry, an increasingly networked and ethically aware public is now capable of dramatic and damaging pushbacks against disliked companies.

It may be true that the worst practices of big pharma are behind it, but shades of that bad attitude linger. And past misdeeds have a habit of coming back to bite. Drug companies must remember that they need the public and its goodwill to test their medicines in the first place. They may have to release more information than they would like, but if they do, it will safeguard the trust and support of the people on whom they ultimately rely. ■

# Risk management

*Teams aimed at preventing violence on campus can offer a lifeline to those in crisis.*

How safe is your workplace? It is nearly four years since biologist Amy Bishop walked into a faculty meeting with a loaded pistol and shot six of her colleagues, killing three. Acts of violence involving multiple victims are extremely rare, especially on college campuses, which tend to be safer than the areas that surround them. But highly publicized events such as Bishop's rampage and the shooting at Virginia Polytechnic Institute and State University in Blacksburg in 2007 — which had one of the highest death tolls of any attack on a college campus — have spurred rapid growth in what is known as threat assessment and management. Developed by behavioural psychologists

working with agencies such as the US Secret Service, threat assessment aims to identify concerning behaviour and situations, and to take pre-emptive action to stop them escalating into violence.

This can involve simply confronting an individual about inappropriate behaviour — aggression towards colleagues, for example — and working with them to correct it. Or it can include maintaining continual contact with an individual and putting them in touch with any help they might need, such as mental-health services.

It is a challenging goal. Universities are big, complex environments where many students, staff and members of the community interact, not always peacefully. But existing networks that organize and monitor housing, health, grades and social activities do offer ways for universities to identify aberrant or shifting behaviour, as well as a robust support structure to get people back on track.

A News Feature on page 150 explores the growth of such programmes and teams, particularly in the United States, where easy access to guns and several high-profile shootings have put the public on high alert. There seem to be some clear benefits, but the spread of these interdisciplinary teams, which often include law-enforcement officials and representatives of university mental-health services, also presents several risks.

One risk that team members often worry about is how to balance individuals' civil liberties with the need to protect others. In an age in which privacy is increasingly illusory, life within the boundaries of a college campus can be put under close scrutiny with little effort. And freedoms of speech and expression must be maintained if institutions of higher education are to continue to nurture ideas.

Another risk of the focus on threat assessment is more subtle, and relates to the all-too-easy assumption that people who commit unthinkable acts of violence are driven by mental illness. It is true that mental illness is implicated in many high-profile cases of targeted violence and that many behaviours that would initiate a call to a threat-assessment team are related to a deteriorating mental state. But the links between violence and mental illness are complex and hardly correlative. Most violence is perpetrated by people who are not mentally ill, and most people with mental-health problems do not commit violent crimes.

The rhetoric of threat assessment therefore runs the risk of further ostracizing people who already face stigma. Many cases managed by a threat-assessment team — there are several hundred referrals per year at an institution such as Virginia Tech — are for students or staff going through a crisis in their personal or professional life. Practitioners are quick to point out that theirs is a support-focused process, more about putting individuals in touch with the help they need to weather that crisis than punishing them, banishing them or branding them as potential threats.

*"For students these services can be extraordinarily helpful, even life-saving."*

Such nuances can be hard for an individual to remember when facing a threat-assessment investigation. And the leading part played by law-enforcement officials in proceedings adds an air of presupposed criminality.

All of this is not to devalue the efforts of these teams. They can be among the first to recognize and the most eager to serve those struggling with mental illness. And they often partner with other student-service organizations whose goals are not focused on averting the next mass shooting. If a case is not deemed particularly risky, threat-assessment teams may pass it over to these groups. For students, who are often facing unfamiliar challenges, these services can be extraordinarily helpful, even life-saving. Many referrals to threat-assessment teams are prompted by threats of suicide, for example.

The politics at play here are sadly familiar in the United States. A highly publicized mass shooting is followed by calls for stricter gun control, followed by pressure from gun supporters to maintain the status quo or even loosen restrictions on firearms. Somewhere along the line, fingers are pointed to the role mental illness had in the attack and attention shifts to the dismal state of mental health care in the country. Accusations are made, as are promises, but little is done. Threat assessment may not be a solution to violence, but if it means that more people get the help they need, irrespective of whether it staves off the next attack, then, to some people at least, it is a success. ■

# Conflict of interest

*How two world wars affected scientific research, and vice versa.*

This year marks the anniversary of two significant events from the last century, perhaps the most significant of any century: 100 years since the outbreak of the First World War and 75 years since the start of the Second World War. It is natural for specialist publications to search out a 'local' angle on major news events, and *Nature* is no different. When it comes to modern warfare, however, the task is easier than with most events, for science is not a tangential topic in armed conflict. It lies, for both good and evil, at its heart.

We live, said Martin Luther King, in an age of guided missiles and misguided men. Scientists can do little about the latter (although we must still try), whereas the former shows the contradictions of military research in all its shades of grey. If we are to kill people, then is it a good thing that we are able to target them more precisely? The death of one becomes more likely; the deaths of others less so.

In times of war, such ethical tongue-twisters tend to give way to the pragmatism of national politics. In 1943, James Collip, one of the 'Toronto group' of scientists that isolated insulin, observed that: "Today, with total war upon the world, there can be no doubt that more than ever before in history this war is a contest between the brains, imagination, inventiveness and teamwork of the scientists and production workers of one group of nations pitted against those of another group." Whereas the first three of those attributes were always common in science, teamwork, as Collip pointed out, came less naturally.

There are two ways to address the topic of science and war. The first, and the most conventional route, is to assess the impact that research has on conflict. Science in the First World War marked a turning point in tactics; no longer was a speedy and resourceful attacker likely to win. With machine guns and barbed wire at the front line, and behind them railroads for resupply, a well-dug-in defender became the favourite. (The US Civil War had demonstrated this too, but European generals were slow learners.) Technology made warfare asymmetric, and it has remained that way — the dreadful stalemate of mutually assured destruction by nuclear weapons notwithstanding.

The second route is to look at the reverse of the equation: how has conflict influenced research? What lessons are there for peacetime science in the panicked scramble of work that aims not to understand how the world works and to improve quality of life, but to ensure that it remains at all?

*Nature* intends to address both topics in several articles this year. And we kick off this week with a good example of each. On page 156, Sharon Weinberger reviews two books that analyse the wartime role of physics and psychology. And on page 153, David Kaiser explores how practical ways of getting US physicists to work together during the Second World War had an enduring impact on the organization and funding of science. For one thing, Kaiser writes, it turned on a "fire hose" of federal funds for research, a model that continues. The teamwork continues too, and if the stakes for winning and losing are lower now than when the original collaborations were forged, that can only be a good thing. ■

➲ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

ANDREW PEACOCK/REUTERS/CORBIS

# This was no Antarctic pleasure cruise

*After his polar vessel became trapped in shifting sea ice,* **Chris Turney** *defends the scientific basis of the expedition.*

Sitting in the ship's lounge of the Australian icebreaker *Aurora Australis*, safe with friends and colleagues and heading back to civilization, I can say it has been a remarkable journey.

For the past six weeks on board the Russian icebreaker *MV Akademik Shokalskiy*, my colleague Chris Fogwill and I have led a team of scientists, science communicators and volunteers on a voyage from the New Zealand subantarctic islands to the East Antarctic Ice Sheet. The aim was to study various aspects of this vast, remote region to better understand its role in the Earth system, and communicate these results directly to the public. Yet most people only became aware of our work when we got stuck and had to be rescued.

That is the reality of polar science. It is difficult. Almost every season, ships get caught in sea ice, teams lose communications and planes are sometimes tragically lost. Signatories to the Antarctic Treaty understand that one science programme supports another, so national and non-government vessels routinely assist each other.

What went wrong for the *Shokalskiy*? Contrary to some reports, the ship was not frozen in but was pinned by remobilized sea ice that had been blown by fierce winds. Most importantly, the team is safe and we are incredibly grateful to the international effort to help us.

Could this have been avoided? The satellite data leading up to our arrival in Antarctica's Commonwealth Bay indicated open clear water, and the area seemed to have been that way for some time. As the *Shokalskiy* attempted to leave, however, we found ourselves surrounded by a mass breakout of multiyear ice. This was a major event, with the vessel surrounded by blocks of sea ice more than three metres thick, apparently arriving from the other side of the Mertz Glacier. Despite the best efforts of our captain, we could not find a route out. It was deeply frustrating. We had been caught just 2–4 nautical miles (3.7–7.4 kilometres) from the edge of the sea ice. And with pervasive southeasterly winds battering our location, this distance increased to 20 nautical miles within 48 hours.

The extreme nature of the conditions is shown by the fate of the Chinese icebreaker that came to our rescue: as I write this, that vessel is also now trapped, and is awaiting the arrival of the huge US ship *Polar Star* to smash a route to open water.

Since news of our plight raced around the world, I have been surprised by the level of criticism our scientific expedition has received. This was no pleasure cruise. The science case took two years to develop, and was approved by the New Zealand Department of Conservation, the Tasmanian Parks and Wildlife Service and the Australian Antarctic Division.

Major scientific questions remain about the region we sailed into. A southward shift of westerly winds is influencing the Antarctic Circumpolar Current, increasing transitory upwelling of Circumpolar Deep Water onto the Antarctic continental shelf. At the same time, extensive sea ice has formed in Commonwealth Bay after a huge iceberg called B09B collided with and destroyed the tongue of the Mertz Glacier in 2010. This is adjacent to the Mertz polynya — a stretch of open water surrounded by ice and a major source of Antarctic bottom water formation. We wanted to gather data on the effects of both these events on circulation, ocean properties, biodiversity and stability of the East Antarctic Ice Sheet.

Never before has a science expedition reached out live to so many people from such a remote location. Public engagement was always a core theme. Well before we ran into trouble, we posted daily online reports of our research and aspects of life on the vessel and in the field. In recent weeks, this extended to reassuring those at home about the well-being of all on board. When the number of television and radio interviews increased, so did our mentions of the science. This encouraged people to follow our work, as seen by the number of hits received on the expedition website. In the past six weeks, www.spiritofmawson.com received 60,000 visits, driving traffic to our social media sites.

NEVER BEFORE HAS A SCIENCE EXPEDITION **REACHED OUT** LIVE TO SO MANY PEOPLE FROM SUCH A **REMOTE LOCATION.**

Our findings include many firsts for the region: detailed marine and terrestrial ecological studies, glaciological reconstructions and high-resolution palaeoclimate analysis of tree rings, peats and ocean cores from the subantarctic islands. Guided by real-time satellite information, the team undertook an experiment across the Antarctic Convergence — a natural boundary between cold Antarctic and warmer subantarctic waters. By combining surface drifters with Argo floats (for measuring salinity and temperature), we have gained a unique snapshot of this important frontier.

Reaching Commonwealth Bay, we crossed some 65 kilometres of sea ice to deliver scientists and conservators to the historic base established by scientist and explorer Douglas Mawson a century ago. We surveyed an airstrip for future visits, serviced and collected data from the automatic weather station, and obtained valuable Global Positioning System data for monitoring land-mass uplift as ice sheets retreat.

Our rescue has caused disruption, but fortunately we hear that the next voyage of the *Aurora Australis* is likely to leave Hobart as scheduled. Science will continue in the south: a great relief. In the meantime, the value of our expedition must be judged by the quality of the research it always intended to produce, and the remarkable rekindling of public interest in science and exploration that has come with it. ■

➲ NATURE.COM
Discuss this article online at:
go.nature.com/wbm6d6

**Chris Turney** *is a scientist in the Climate Change Research Centre at the University of New South Wales, Australia.*
*e-mail: c.turney@unsw.edu.au*

# RESEARCH HIGHLIGHTS

*Selections from the scientific literature*

NEUROSCIENCE

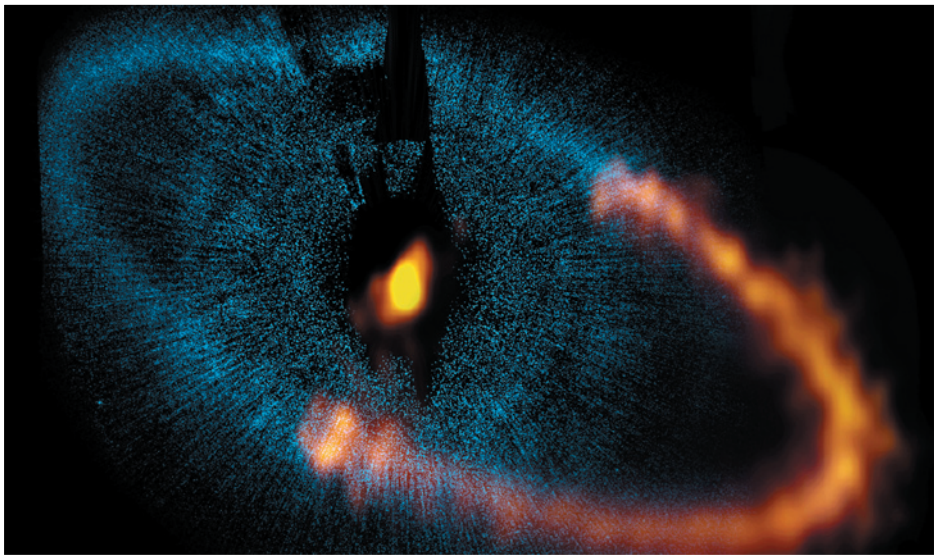
## *In vivo* cell switch for brain repair

By reprogramming one type of brain cell into another *in vivo*, researchers have opened the door to new ways of repairing damaged brains.

Gong Chen and his colleagues at Pennsylvania State University in University Park converted reactive glial cells — the cells that flood sites of brain injury — into neurons in the brains of mice. They injected a retrovirus carrying the gene encoding a protein called NeuroD1 into the cortex of normal mice and those engineered to model Alzheimer's disease. The virus delivered the gene to two types of glial cell, resulting in the reprogramming of these cells into functional excitatory or inhibitory neurons.

NeuroD1 also turned human astrocytes, a type of glial cell, into functional neurons *in vitro*. The authors suggest that the approach could be used to replace neurons lost to injury or disease in humans.

***Cell Stem Cell*** **http://doi.org/qq7 (2013)**



ALMA (ESO/NAOJ/NRAO); VISIBLE LIGHT IMAGE: NASA/ESA/HST

ASTROPHYSICS

## Comets hint at cosmic encounter

Researchers have discovered a second comet belt in Fomalhaut, a bright, triple-star system that is already known to host an exoplanet and a bright comet belt around its primary star, Fomalhaut A (**pictured**).

Grant Kennedy at the University of Cambridge, UK, and his colleagues used data from the Herschel Space Observatory to find the second belt, which surrounds the system's least-massive star, Fomalhaut C. The discovery of two such bright comet belts around stars in the same system is rare, and suggests that the two stars had a close encounter that increased collision rates within each debris disk: colliding comets generate large amounts of dust and ice that would make the comet belts look brighter. Such a stellar interaction could also explain the elliptical orbits of Fomalhaut A's exoplanet and comet ring.

***Mon. Not. R. Astron. Soc.*** 437, **2686–2701 (2014)**

MATERIALS

## Extra-stretchy graphene gloves

Graphene-based sensors that measure strain, or deformation, can be stretched to twice their normal length. These could be useful for the development of wearable interactive electronics.

Previous such devices struggled to stretch by even 30%, making them too stiff to detect the full range of motion of human joints, for instance. Pooi See Lee and her colleagues at Nanyang Technological University in Singapore made their sensors out of nanopaper: crumpled graphene (atom-thick sheets of carbon) and tiny cellulose fibrils embedded in a stretchy material.

A glove developed by the researchers has strain sensors on each finger (**pictured**) for measuring the bending and stretching of separate digits. It could one day be used to perform surgery remotely and in other applications.

***Adv. Mater.*** **http://doi.org/qrh (2013)**

CHAOYI YAN & POOI SEE LEE/WILEY

IMMUNOLOGY

## Dietary fibre dampens asthma

A high-fibre diet curbs allergic inflammation in mouse lungs by shifting the composition of microbes in the gut.

Benjamin Marsland at the University of Lausanne in Switzerland and his colleagues raised mice on diets containing different levels of fibre and exposed the animals to extracts of house-dust mite, a cause of asthma. The resulting lung inflammation was less in mice consuming high levels of fermentable fibre than in those on a low-fibre diet, and the animals also harboured a community of intestinal microbes that generated higher levels of short-chain fatty acids when metabolizing fibre. These fatty-acid molecules boosted

the generation of immune cells called dendritic cells that were less able to trigger allergic inflammation in the lungs.

The results provide a possible link between the rising incidence of asthma in developed countries and decreasing dietary-fibre intake.
***Nature Med.*** **http://dx.doi.org/10.1038/nm.3444 (2014)**

ZOOLOGY

## Tobacco breath aids defence

The tobacco hornworm, which feeds on tobacco plants, exhales some of the ingested nicotine to repel predators.

Ian Baldwin and his colleagues at the Max Planck Institute for Chemical Ecology in Jena, Germany, glued tiny sensors to the mouths of tobacco hornworm larvae (*Manduca sexta*; **pictured**) to measure the levels of nicotine in their breath. They found that larvae that fed on engineered, nicotine-free tobacco plants (*Nicotiana attenuata*) exhaled less nicotine than larvae fed on normal plants, as did larvae in which an enzyme that transfers nicotine from the gut into the circulation was silenced. Wolf spiders (*Camptocosa parallela*) preferred to prey on larvae with less nicotine on their breath.

The study shows how ingested toxic chemicals can be used for predator defence.
***Proc. Natl Acad. Sci. USA*** **http://doi.org/qq3 (2013)**

HEATHER MCCLELLAND/AP/PA

PAVAN KUMAR

GENOMICS

## CRISPR screen identifies genes

Two teams show how a genome-editing system can be used to screen human cells for genes of interest.

The CRISPR system allows biologists to edit specific genes using 'guide' RNA molecules that target them. Feng Zhang at the Broad Institute in Cambridge, Massachusetts, and his colleagues created a library of 64,751 guide RNA sequences that target 18,080 genes in human cells. Using this library, the researchers pinpointed genes that are required by cancer and stem cells to survive. They also teased out genes that, when lost, allow cancer cells to fend off the melanoma drug vemurafenib.

A separate team led by Eric Lander at the Broad Institute and David Sabatini at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, used a library of 73,000 guide RNAs to screen for several genes, including those involved in resistance to the chemotherapy drug etoposide.
***Science*** 343, **80–84; 84–87 (2014)**

STRUCTURAL BIOLOGY

## Better pictures of protein structures

A modified method for determining the three-dimensional structure of large proteins seems to show them in a more natural pose than conventional techniques do.

Proteins called G protein-coupled receptors (GPCRs) are important drug targets, but researchers struggle to figure out their structures. Vadim Cherezov of the Scripps Research Institute in La Jolla, California, and his colleagues modified the standard X-ray crystallography technique by using an X-ray free-electron laser to capture serial images of the structure of a GPCR for the neurotransmitter serotonin.

The team used the technique on tiny serotonin-receptor crystals kept at room temperature, and obtained structures that differed from those determined using conventional approaches with larger crystals kept at cool temperatures. The results suggest that the room-temperature free-electron-laser approach may better capture the protein's conformation in its native environment.
***Science*** 342, **1521–1524 (2013)**

## COMMUNITY CHOICE

*The most viewed papers in science*

CANCER

## How drugs boost resistance

HIGHLY READ on jem.rupress.org in December.

Chemotherapy kills cancer cells, but in colorectal cancer it can also stimulate their growth by activating cells called fibroblasts in the connective tissue.

Matthew Kalady and Jeremy Rich at the Cleveland Clinic in Ohio and their colleagues analysed tumours from patients with colorectal cancer before and after chemotherapy. The researchers found that the abundance of cancer-associated fibroblasts increased after treatment, and that these cells enhanced the ability of a subset of cancer cells to initiate tumour growth. The fibroblasts seem to do this by secreting signalling proteins, including one called IL-17A.

The findings suggest that chemotherapy can trigger drug resistance by changing the tumour's microenvironment. Disrupting this mechanism could be a way of improving cancer therapies, the authors say.
***J. Exp. Med.*** 210, **2851–2872 (2013)**

GEOLOGY

## Radar signals sinkhole to come

Radar measurements taken more than a month before a giant sinkhole (**pictured**) opened up in 2012 in Bayou Corne, Louisiana, reveal that nearby ground shifted horizontally towards the pit's location.

Cathleen Jones and Ronald Blom of NASA's Jet Propulsion Laboratory in Pasadena, California, looked at radar data gathered by an unmanned aircraft as part of a Mississippi River delta study. By comparing data from flight passes in June 2011 and July 2012, the team saw that surface material had moved by as much as 26 centimetres towards where the 110-metre-wide sinkhole appeared in August 2012.

Radar remote sensing could be a way of predicting the formation of these potentially catastrophic sinkholes and their growth rate, the authors say.
***Geology*** **http://doi.org/qnr (2013)**

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

# SEVEN DAYS *The news in brief*

POLICY

## Seeds of change

The US Department of Agriculture (USDA) on 3 January proposed removing restrictions on the use of maize (corn) and soya bean seeds that are genetically engineered to tolerate 2,4-D, a weedkiller that is commonly used on other crops. More widespread use of the genetically modified seeds, which are made by Dow AgroSciences in Indianapolis, Indiana, along with 2,4-D could drive evolutionary selection for weeds that are resistant to the chemical, the agency cautioned. But the USDA noted that the move would provide a much-needed tool for farmers to manage fields that are already plagued by weeds resistant to another weedkiller, glyphosate (Roundup).

## Clinical data

A lack of access to clinical-trial data is hindering research and medical care, according to a British government report released on 3 January. In the report, which also examined the United Kingdom's stockpiling programme for the influenza drug Tamiflu (oseltamivir), the authors say that evaluation of the efficacy of Tamiflu and other medicines has been hampered by drug manufacturers withholding data. The report follows recent European and US initiatives to increase data sharing and transparency in clinical trials. See page 131 and go.nature.com/9lgbd6 for more.

## Gun controls

The US Department of Health and Human Services proposed on 3 January that patient-privacy exemptions should be created so that relevant mental-health records can be submitted to the national databases used to screen potential gun buyers. So far, background checks have prevented the sale of more than 2 million firearms, according to the White House. But some researchers who study firearms violence have called for better safeguards against gun ownership by those who are mentally ill (see *Nature* **496,** 412–415; 2013).

## Cannabis commerce

The world's first legal market for recreational marijuana opened in Colorado on 1 January. Colorado is one of only two US states to have approved non-medicinal use of marijuana by adults (see go.nature.com/rtr3ou). Although the drug remains illegal under national laws, the US government has said that it will not interfere with state industries kept under strict local controls. Last month, Uruguay became the first country to approve a national market for legal marijuana, which has yet to be implemented.

## New drug approvals

The US Food and Drug Administration approved 27 new drugs in 2013, down from a 15-year high of 39 drugs the year before, which some analysts had taken as a signal of revived fortunes in the pharmaceutical industry. The agency told reporters that it had received fewer drug applications for review in 2013 and that the number of approvals was in line with the average of 28 new drugs approved annually over the past five years.

EVENTS

## Asteroid ahoy!

A small asteroid disintegrated above the Atlantic Ocean on 2 January, becoming only the second space rock to be spotted hours before it hit Earth. The Catalina Sky Survey near

LI XIN/XINHUA PRESS/CORBIS

## China joins ivory-crushing campaign

China's government crushed more than six tonnes of seized ivory in Dongguan on 6 January, as part of a global effort to crack down on illegal trading in smuggled tusks and carvings. The move was China's first public destruction of ivory, showing the country's intention to thwart a worrying rise in elephant poaching (see *Nature* **503,** 452; 2013). The United States crushed a similar weight of ivory last November. See go.nature.com/ib2fpa for more.

NUMBER CRUNCH

# 72%

**The increase in female speakers at the American Society for Microbiology general meeting in sessions organized by teams that included at least one woman, compared with those set up by all-male teams, according to a study of conferences from 2011 to 2013.**

Source: Casadevall, A. & Handelsman, J. *mBio* http://dx.doi.org/10.1128/mBio.00846-13 (2014).

US ICE

Tucson, Arizona, discovered asteroid 2014 AA in the small hours of 1 January. The rock, some 2 to 3 metres across, burned up on hitting Earth's atmosphere. In October 2008, researchers tracked asteroid 2008 $TC_3$ all the way from space to the desert wastes of northern Sudan, where fragments of it were recovered as meteorites (see *Nature* **458,** 401–403; 2009).

## Cold comfort

Scientists, journalists and tourists were rescued from the Russian ship *Akademik Shokalskiy* in the Antarctic on 2 January. The vessel had been on a research voyage when it became trapped by ice near Commonwealth Bay on 24 December. Chinese icebreaker *Xue Long* transferred the stranded passengers to an Australian icebreaker, but later reported that it, too, had become stuck. A US icebreaker was dispatched on 5 January to assist the Russian and Chinese vessels. See page 133 for more.

PEOPLE

## Fossil felony

A fossil retailer from Eagle, Colorado, pleaded guilty on 2 January to conspiracy to smuggle dinosaur bones and other fossils into the United States from China and Mongolia. John Richard

Rolater agreed to surrender any claims to the illegally obtained goods, which include a fossilized skull of a juvenile *Tyrannosaurus bataar* (**pictured**) that is estimated to be worth US$1,875,000. Other items include a sabre-toothed cat skull and dinosaur eggs. Rolater has also agreed to pay a $25,000 fine and submit to two years of supervised probation.

## Turing pardon

British mathematician Alan Turing has received a posthumous royal pardon. In 1952, Turing was convicted of 'gross indecency' under anti-homosexuality legislation, and later took his own life. Turing's work in the Second World War helped to break the German Enigma cipher, and his concept of a universal 'Turing machine', a programmable system that stores and processes information, is considered a cornerstone of computer science (see *Nature* **482,** 441; 2012).

RESEARCH

## Falsified research

The US Office of Research Integrity has sanctioned two biomedical researchers in seven days. On 30 December, the agency reported that Baoyan Xu, a former postdoctoral fellow at the US National Institutes of Health (NIH) in Bethesda, Maryland, had published falsified data on the immune responses of patients with hepatitis to a newly discovered virus. A week earlier, Dong-Pyou Han, a former research assistant professor at Iowa State University of Science and Technology in Ames, was found to have falsified results when researching a vaccine against human immunodeficiency virus 1 (HIV-1) by spiking rabbit blood samples with antibodies. The false results had been reported widely at national and international meetings, and in NIH grant applications.

## Open access

An international open-access effort kicks off this month to make all particle-physics research articles freely available to readers. The Sponsoring Consortium for Open Access Publishing in Particle Physics (SCOAP$^3$) is led by CERN, Europe's high-energy physics laboratory near Geneva, Switzerland. The project has already experienced a few hiccups, with some major journals and universities opting not to participate. See page 141 for more.

## COMING UP

**10–15 JANUARY**
Understanding how cells' nuclear receptors regulate gene expression is the focus of a Keystone Symposium on Molecular and Cellular Biology in Taos, New Mexico. Hot topics include the roles of nuclear receptors in wound healing and cancer progression.
go.nature.com/rs9oyb

**15–17 JANUARY**
The 8th Human Amyloid Imaging meeting in Miami, Florida, will discuss the latest research on measuring and interpreting changes in amyloid protein in the brain, as well as other biomarkers linked to Alzheimer's disease.
go.nature.com/oi5wkp

FUNDING

## Cancer donation

Six US research centres have received a combined donation of US$540 million from the estate of late shipping magnate Daniel Ludwig. The gift will boost funding for the Ludwig Centers at Johns Hopkins University, Harvard University, the Massachusetts Institute of Technology, Memorial Sloan-Kettering Cancer Center, Stanford University and the University of Chicago. Announced on 6 January, the donation brings the total contribution to cancer research by Ludwig and his estate to $2.5 billion.

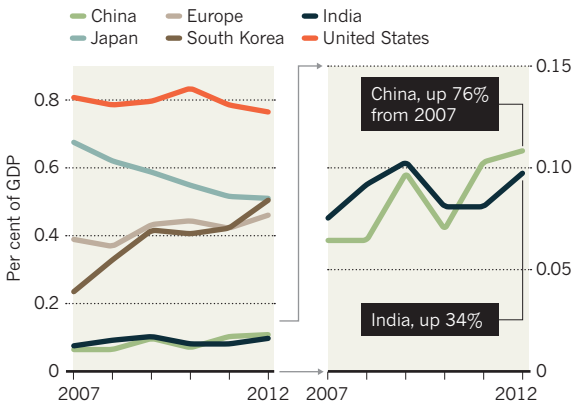
## TREND WATCH

Recent years have seen a declining share of the US economy spent on biomedical research and development (R&D), while several Asian nations are boasting growing investments (see chart), according to a study published on 2 January (J. Chakma *et al. N. Engl. J. Med.* **370,** 3–6; 2014). The shifting trends mark changes in spending by the biomedical industry, perhaps reflecting lower labour costs and more government subsidies for commercial R&D in Asia, the authors suggest.

SOURCE: *N. ENGL. J. MED.*

### ASIAN R&D BOOM

Countries including China, India and South Korea boost biomedical spending as US dominance slips.




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

# NEWS IN FOCUS



MASSIMO VALICCHIA/NURPHOTO/CORBIS


**Potential patients have offered vocal support for Stamina's stem-cell treatment in Italy.**

REGENERATIVE MEDICINE

# Leaked files slam stem-cell therapy

*Disclosures and resignations reveal scientific concerns over methods of Italy's Stamina Foundation.*

**BY ALISON ABBOTT**

A series of damning documents seen by *Nature* expose deep concerns over the safety and efficacy of the controversial stem-cell therapy promoted by Italy's Stamina Foundation. The leaked papers reveal the true nature of the processes involved, long withheld by Stamina's president, Davide Vannoni. Other disclosures show that the successes claimed by Stamina for its treatments have been overstated. And, in an unexpected twist, top Italian scientists are dissociating themselves from an influential Miami-based clinician over his apparent support for the foundation.

Stamina, based in Brescia, claims that it successfully treated more than 80 patients, mostly children, for a wide range of conditions, from Parkinson's disease to muscular dystrophy, before the health authorities halted its operations in August 2012. A clinical trial to assess the treatment formally was approved by the Italian government last May, and an expert committee was convened by the health ministry to study Stamina's method and to recommend which illnesses the trial should target.

Stamina says that its technique involves extracting mesenchymal stem cells from a patient's bone marrow, culturing them so that they turn into nerve cells, and then injecting them back into the same patient. But full details of the method have never been revealed, and Vannoni provided the full protocol to the expert committee only in August.

In October, the committee's report prompted health minister Beatrice Lorenzin to halt plans for the clinical trial. That led to public protests in support of Stamina, and, after an appeal by Vannoni, a court ruled in early December that the expert committee was unlawfully biased. Some members had previously expressed negative opinions of the method, the ruling said. As a result, Lorenzin appointed a new committee on 28 December, reopening the possibility of a clinical trial.

Stamina's protocol, together with the original committee's report, was leaked to the press on 20 December (*Nature* has also been shown transcripts of the committee's deliberations). The leaked papers reveal that the original expert committee identified serious flaws and omissions in Stamina's clinical protocol. It did not apply legally required Good Manufacturing Practice standards, the committee says. The protocol exposed an apparent ignorance of stem-cell biology and relevant clinical expertise, the report argues, as well as flawed methods and therapeutic rationale (see 'Protocol opinion').

A week after the leaks, the health ministry revealed that the condition of 36 patients treated with Stamina's stem-cell therapy had not improved, contrary to Vannoni's claims that more patients had shown improvement.

The leaked documents also convey the original committee members' disquiet over the unusually strict confidentiality agreement that they had to sign. This prevented them from ever divulging details of the protocol — and each committee member received an individualized copy of the method from Stamina to aid in the identification of any leak. The committee argued that such secrecy was ▶

## PROTOCOL OPINION

### *What the expert committee said on Stamina's methods.*

The report of the original expert committee tasked with looking at Stamina's clinical protocol includes the following opinions:

The protocol contains no method for screening for pathogens such as prions or viruses, even though the culture medium used could contain them.

The method it describes for producing mesenchymal stem cells, say the experts, would generate a mixture of different cell types that could include blood-cell precursors and bone fragments.

The method it describes for checking the biological identity of the cells uses inappropriate cell-surface markers and no functional assay. And the protocol does not include a method for making mesenchymal stem cells differentiate into neural cells, the rationale Stamina provided, along with the protocol, in support of the clinical value of the method in three disorders it proposed that the clinical trials address.

Even if the Stamina method did generate the desired cells, the experts' report notes, there would be too few of them for the treatments Stamina proposes. The experts also criticized an 'emergency' measure in the protocol that would culture a patient's sample once more just before injection if at the last minute numbers of stem cells appeared sparse. This would mean that treatments across patients in the clinical trial would not be standardized, they say.

The clinical rationales provided by Stamina also contain conceptual errors, the experts say, as it is broadly accepted by the scientific community that the stem cells can differentiate into only bone, fat or cartilage. Moreover, the committee notes that sections of the protocol are copied from Wikipedia. **A.A.**

▶ unnecessary because no intellectual property or commercial interests were involved. Vannoni insists that Stamina will not make a profit. He has also said that, in the opinion of the court, the committee "had neither the right nor competence" to comment on the protocol.

Other events have further dented Stamina's credibility. Leading researchers on the scientific advisory boards of two independent stem-cell initiatives headed by clinician Camillo Ricordi have resigned in protest over the apparent public support offered to Stamina by Ricordi. Ricordi, who works on diabetes at the University of Miami in Florida, has in the past called Stamina's method "safe" and "promising".

On 23 December, Carlo Croce, a cancer researcher at the Ohio State University in Columbus, resigned from the scientific committee of one of the initiatives, the Ri.MED Foundation, a publicly funded regenerative-medicine institute being built in Palermo, Italy. Croce has called for Ricordi to be removed as Ri.MED's president. Other committee members told *Nature* that they are also considering resigning from the Ri.MED committee.

And at the end of December, cell biologist Carlo Redi of the University of Pavia, Italy, stem-cell biologist Giulio Cossu at University College London and Francesca Pasinelli, director-general of the Italian grant-giving charity Telethon, all resigned from the Cure Alliance, a lobby group for speeding up translational medicine that Ricordi launched.

The scientists who resigned say that they were dismayed by Ricordi's insistence that the value of Stamina's therapy had not yet been proved or disproved, as well as his offer to test and possibly improve it in his Miami facilities.

Ricordi has told *Nature* that he is "not at all in favour of Stamina, but in favour of a verification process that is in the interest of all". He also points out that new members have joined the committees of the two initiatives. In response to the resignations, Vannoni told *Nature*: "I think that if a scientist resigns from a research centre because a colleague of his decides to study a new method, he/she has not got a correct approach to science."

Ricordi has featured regularly in the long battle over Stamina's methods. For example, at the July 2013 meeting of the Congress of the Cell Transplant Society in Milan, he announced that he had seen the then-secret Stamina protocol and witnessed how patients in Brescia had improved. In a press statement that day, he said that he believed it to be a "safe" procedure. "The results and data I was able to see appeared promising," he said.

On 7 January, he added that: "If someone has a protocol that results in clinical benefits — to be validated — I wouldn't question their level of stem-cell biology knowledge." He also said he had seen certificates confirming that Stamina's cell samples were sterile. But Paolo Bianco, an expert in mesenchymal stem cells from the University of Rome La Sapienza, notes that these certificates do not cover the presence of prions or viruses in the sample.

Ricordi backs a controversial proposal that cell therapies should not be regulated as medicines, as US and European regulators insist, but as transplants. He argues that because transplants are not subject to strict regulation, novel stem-cell therapies could be introduced more quickly.

As editor-in-chief, Ricordi launched the official journal of the Cure Alliance, *CellR*$^4$, in 2013. He has written editorials in the journal attacking critics of his views on transplant regulation, including Bianco and *Nature*.

Ricordi began lobbying the former Italian health minister Renato Balduzzi over his views on transplants at the end of March last year. On 16 April, Balduzzi nominated him as president of Ri.MED. And in an e-mail to *Nature* on 2 January this year, Ricordi said that Lorenzin — the current health minister — "has recently asked me to help in resetting regulation for cellular therapies". But Lorenzin told *Nature* that she had met Ricordi only once at a social occasion and that they had discussed a different topic.

In December, Ricordi reached an agreement with Stamina to begin testing its patients' cell samples in his facilities in Miami. He denies being in favour of Stamina, but says that he is an advocate of patients who need to know whether or not the treatment works. "Testing is the only way to bring clarity," he says.

But Ruggero De Maria, science director of the Regina Elena National Tumour Institute in Rome, says: "Tests on samples have already been carried out independently at the University of Modena in Italy. I feel offended when I see Ricordi praising Stamina and attacking experts."

Ricordi told *Nature* that "a mud machine is being orchestrated" against him. Vannoni says that Ricordi is objective and open to new ideas but is not "supportive" of him. ■

## MORE ONLINE

**TOP STORY**

DE AGOSTINI PICTURE LIBRARY/GETTY

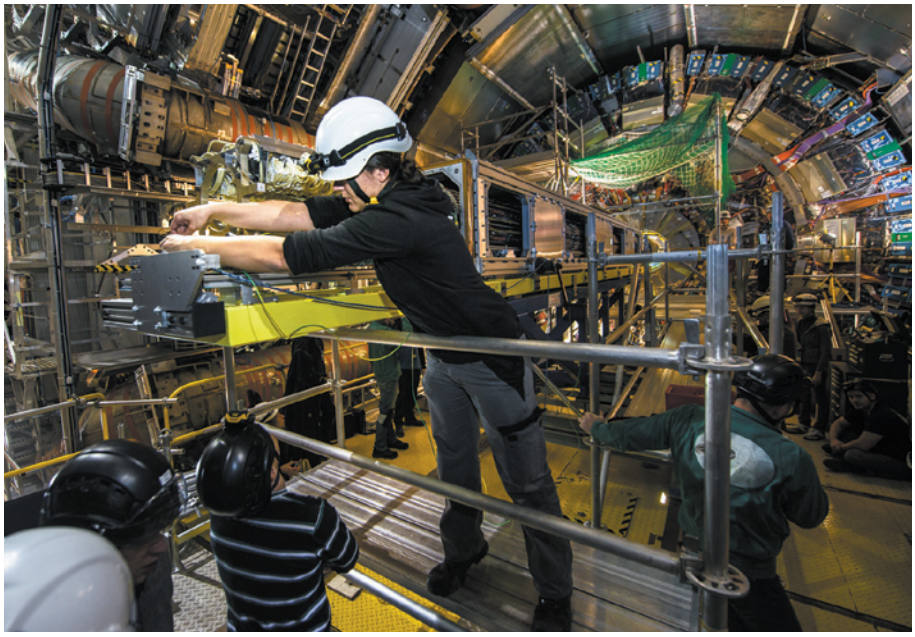
Mysterious peoples from the north left legacy in ancient European genomes **go.nature.com/fipboo**

**MORE NEWS**

- Supervolcano eruptions are powered by buoyancy **go.nature.com/xf4pt5**
- 'Earthquake lights' concentrate at seismic faults **go.nature.com/dta4mf**
- New Zealand's oldest shipwreck was secret expedition **go.nature.com/thhbtc**

**NATURE PODCAST**


Better batteries, managing threats on campus and colourful marine fossils **nature.com/nature/podcast**



MAXIMILIAN BRICE/CERN

Particle-physics research is at the centre of a global push for open-access publishing.

PUBLISHING

# Particle-physics papers set free

*Tensions as open-access initiative goes live — without the field's leading journal.*

BY RICHARD VAN NOORDEN

January sees the start of what has been billed as the largest-scale open-access initiative ever built: an international effort to switch the entire field of particle physics to open-access publishing.

But the initiative, organized by CERN, Europe's high-energy physics laboratory near Geneva in Switzerland, has not yet fulfilled its dream — it currently covers only a little more than half of published particle-physics papers.

The scheme's scope was slashed in the summer when the field's largest journal, *Physical Review D*, pulled out, although its publisher, the American Physical Society (APS), did agree to publish papers on experiments at CERN's Large Hadron Collider on an open-access basis without charging author fees.

And as the starting gun sounded, a number of US libraries, including those of Stanford University in California and Yale University in New Haven, Connecticut, declined to pay into the system. Robert Schwarzwalder, associate university librarian for Stanford's science and engineering libraries, wants to see open-access research but is not sure that CERN's initiative is needed, given that versions of almost all high-energy physics articles already appear online for free on the preprint server arXiv and the repository INSPIRE-HEP. Yale did not return *Nature*'s calls.

In the global scheme, called the Sponsoring Consortium for Open Access Publishing in Particle Physics (SCOAP$^3$), libraries either pay reduced subscription fees for participating journals, or stop paying them altogether. The cash saved goes into a central SCOAP$^3$ fund, used to pay publishers up front to publish open-access articles.

***"Being able to flip four out of the five large journals to open-access publishing is a remarkable result."***

Instead of hiding articles behind paywalls, publishers will make them immediately available on their own websites, with generous rights for reuse. Authors will retain the copyright. Libraries will not necessarily save money, because the average fee for publishing a paper — €1,150 (US$1,570) — has been set roughly to match publishers' lost subscription revenues. But in three years' time, contracts may be renegotiated, so open-access fees might go down.

SCOAP$^3$ was designed to flip journals to open access without disrupting researchers' funding arrangements: scientists would not have to pay fees from their own research grants. In 2012, after more than 6 years of discussion with over 1,000 libraries, library consortia and research organizations in 25 countries, 12 journals signed up to the scheme. Six agreed to switch entirely to open-access models and six to publish particle-physics articles in an open-access format. It was "the most systematic attempt to convert all the journals in a given field to open access", says Peter Suber, director of the Harvard Open Access Project in Cambridge, Massachusetts.

Then, in June 2013, the APS announced that it would pull out of the scheme, taking with it two journals, including *Physical Review D*, which publishes more than 3,000 high-energy physics articles each year behind a paywall, although it also allows authors to pay to make individual articles open access. The APS says that it is committed to "sustainable open access", and notes that it allows authors to post the final manuscripts of APS papers on their own websites. But the society wanted to maintain its "long-term financial stability", it said. "The data available to us did not allay our long-standing concerns about the stability of SCOAP$^3$ and about the risks to *Physical Review D* from participating," says Joseph Serene, treasurer and publisher for the APS in College Park, Maryland.

Salvatore Mele, who leads SCOAP$^3$ from CERN, says he regrets the APS decision, which has sparked a minor stand-off. CERN has decreed that all articles based on its research must be open access, but says it will not pay open-access fees to journals that either withdrew from or remained outside the SCOAP$^3$ process. The APS says that — apart from papers relating to the Large Hadron Collider — it will not waive open-access fees for CERN, so it will be up to individual authors to find the money.

Schwarzwalder is also concerned about the long-term viability of the scheme. The danger, he says, is that libraries may be tempted to renege on their pledges in order to save money — because they will have access to the papers whether or not they pay. Libraries that do not pay into SCOAP$^3$'s €5-million pot will also effectively be freeloading on those that have. This could rapidly make SCOAP$^3$'s publishing economically unsustainable. Nevertheless, Schwarzwalder says Stanford is now reconsidering its decision.

Despite all the teething pains, Mele considers SCOAP$^3$ a success. It has created a worldwide community of funding agencies and libraries that believe it is possible to convert subscription fees into open-access fees. "Being able to flip four out of the five large journals to open-access publishing — which was considered to be impossible — is a remarkable result." ■

MOHAMED NURELDIN ABDALLAH/MÉDECINS SANS FRONTIÈRES/REUTERS/CORBIS

This woman's rash is symptomatic of kala-azar, a parasitic disease spread by sandfly bites in the tropics.

PHARMACEUTICALS

# Projects set to tackle neglected diseases

*But they do little to alter the process of drug development.*

BY ERIKA CHECK HAYDEN

Kala-azar, the most deadly parasitic disease after malaria, afflicts hundreds of thousands of the world's poorest people in tropical countries such as India, Brazil and Sudan. Spread by sandfly bites, the disease can be fought with existing treatments — but these are expensive and inconvenient, and sometimes have toxic side effects.

Yet commercial work aimed at finding better drugs for kala-azar has largely been abandoned. Pharmaceutical companies say that poor customers cannot afford to pay the high prices needed to recoup development costs. Critics say that eight proposals, endorsed last month by reviewers for the World Health Organization (WHO) to break the stalemate for this and other neglected diseases, are noble, but no solution. The measures will do little, they say, to solve a broader problem: the disparity in spending on research and development for diseases of the rich and those of the poor.

The proposal to combat kala-azar (also known as visceral leishmaniasis, or VL) would combine groups already working on drugs for the disease into a single organization, the VL Global R&D & Access Initiative. This would seek to develop durable oral drugs that do not require cold storage or intravenous delivery. The non-profit plan will be considered by the WHO executive board at a meeting on 20–25 January at the organization's headquarters in Geneva, Switzerland.

But critics are upset that novel and more risky ideas that would have helped to unlink the cost of drug development from prices were eschewed in favour of the eight shortlisted proposals, which were seen as more viable because they build on existing efforts and focus on specific diseases. "The proposals that were brought forward were not as strong as we had hoped in identifying alternative pathways to traditional research and development through commercial channels," says Nils Daulaire, assistant secretary for global affairs at the US Department of Health and Human Services. "That was, frankly, disappointing."

In response to this criticism, the WHO has asked the backers of the eight projects — five of which focus on developing vaccines or medicines for specific neglected diseases, one on fever diagnostics and two on basic research — to explain this month how they will test methods for funding the work. The responses will help the executive board to decide which projects to endorse. Then, at a World Health Assembly meeting in Geneva in May, countries will be asked to commit funds for the schemes.

The projects are part of an attempt to salvage a decade-long effort to create new funding mechanisms for neglected diseases. Despite campaigning from advocacy organizations such as the Drugs for Neglected Diseases Initiative in Geneva, and hefty donations from groups such as the Bill & Melinda Gates Foundation in Seattle, Washington, drug development is still disproportionately focused on diseases of the rich, such as heart disease and cancer.

Three times in the past decade, countries have failed to sign treaties that would commit them to fund drug development for neglected diseases. When the latest attempt was quashed in November 2012, diplomats agreed instead to back a series of demonstration projects that would test new funding mechanisms and be reviewed in 2016.

But critics worry that the eight shortlisted pilot projects are not actually testing new ways of funding, and that more innovative ones have been shelved. One proposal, rejected last month, would have used two tools — milestone payments and patent pools — to spur the development of tuberculosis medicines. Milestone payments would reward early-stage successes of potential drugs, such as proof of activity in humans. Recipients of the payments would then place intellectual property on these potential drugs into a patent pool. Drug developers could license these patents at low cost and would agree to put further patents back in the pool. Another rejected proposal involved taxing antibiotic use to fund the development of antimicrobials.

In their deliberations, reviewers were asked to score the projects' public-health impact and scientific merit ahead of their novelty. Some neglected-disease advocates say that those priorities should have been reversed. Now that the more innovative projects have been dropped, "we're not going to get to the place in two years' time where we can say how well a completely different approach to research and development can work", says Katy Athersuch, an advocate for affordable medicines with the non-profit organization Médecins Sans Frontières (also known as Doctors Without Borders), based in Geneva.

Paying for the projects in the traditional way — by garnering direct support from donor nations — may be difficult enough. A 2012 WHO report recommended that all countries spend 0.01% of their annual gross domestic product on neglected diseases, which would roughly double spending on these illnesses to US$6 billion per year. Only the United States is currently doing this, and emerging economies such as China, Brazil and India have yet to increase their spending.

If the WHO endorses some of the eight projects later this month, it will be a critical time to see if nations step up to pay for them, says John-Arne Roettingen, a global-health researcher at Harvard School of Public Health in Boston, Massachusetts. He says: "This will be the first test of whether countries are willing to put their money on the table." ■

EARTH SCIENCE

# Many eyes on Earth

*Swarms of small satellites set to deliver close to real-time imagery of swathes of the planet.*

BY DECLAN BUTLER

MATT MCDONALD/SKYBOX IMAGING

**Skybox Imaging plans to launch 24 high-resolution SkySats in the next few years.**

Imagine using Google Earth or other online mapping tools to zoom in on high-resolution satellite images of the planet taken just hours or days ago. Navigating backwards and forwards in time, one could track changes in everything from crops, forests and wildlife movement to urban sprawl and natural disasters, all with unrivalled temporal precision.

This is the vision of two Californian start-up companies that are set to launch swarms of small imaging satellites, which, by virtue of their sheer numbers, will be able to revisit and photograph huge swathes of the planet as often as several times each day — a frequency much higher than that achieved by current Earth-observing satellites.

San Francisco-based Planet Labs, founded in 2010 by three former NASA scientists, is scheduled to launch 28 of its 'Doves' on 8 January. Each toaster-sized device weighs about 5 kilograms and can take images at a resolution of 3–5 metres.

At Skybox Imaging in nearby Palo Alto, plans are afoot for a swarm of 24 satellites, each weighing about 100 kilograms, which will take images of 1 metre resolution or better. Skybox launched its first satellite on 21 November and plans to launch another this year, followed by the remainder between 2015 and 2017.

In a first — at least for civilian satellites — Skybox's devices will also stream short segments of near-live high-resolution video footage of the planet. So, too, will UrtheCast, a start-up based in Vancouver, Canada, whose cameras will hitch a ride on the International Space Station (see go.nature.com/cebdkb).

The efforts could herald a sea change for imaging. Conventional imaging satellites, which are the size of a van and weigh tonnes, cost hundreds of millions of dollars to build and launch (see 'The swarm cometh'). As a result, there are only a handful of operators, and the commercial world fleet comprises fewer than 20 satellites. Commercial satellites also tend to take pictures mainly when their operators receive orders from customers.

By contrast, the swarm satellites' cameras will always be on, photographing everything in their path and, owing to their numbers, will pass over the same points on Earth with a frequency of hours to a few days, depending on latitude.

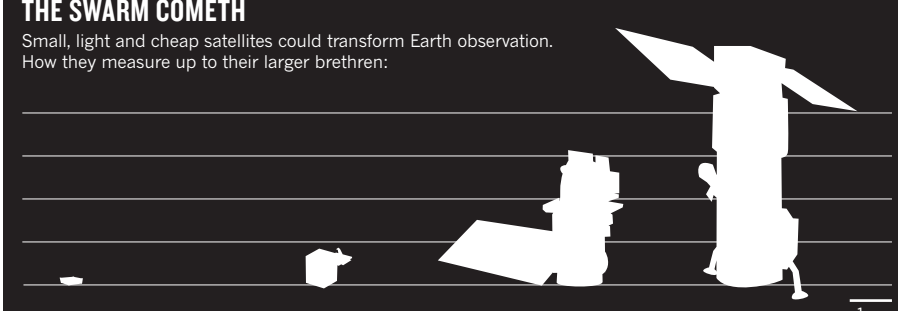
The biggest customers of conventional commercial imaging satellites are governments, in particular intelligence agencies and the military. Prices can be prohibitive for many other potential users, including researchers, in areas as diverse as farming, forest carbon management, regional and local planning, and environmental stewardship. By making their images cheaper, the new entrants into the marketplace hope to spur a proliferation of innovative uses. They also hope to offer heavy discounts or even make imagery free to academics and non-governmental organizations.

"This sector has for so long been driven by government requirements and, to a lesser extent, big industry players, that the mass-market consumer — the long tail — has been almost completely neglected," says Scott Larson, chief executive of UrtheCast. Cheaper imagery, he says, will lead to "the democratization of near-real-time Earth-observation data".

To slash costs, Planet Labs and Skybox Imaging use off-the-shelf technologies from the ▶

## THE SWARM COMETH

Small, light and cheap satellites could transform Earth observation. How they measure up to their larger brethren:




| DOVE | SKYSAT | LANDSAT 8 | WORLDVIEW-3 |
|---|---|---|---|
| **Operator:** Planet Labs | Skybox Imaging | NASA | DigitalGlobe |
| **Number of satellites*:** 32 | 24 | N/A | N/A |
| **Weight:** ~5 kg | ~100 kg | 2,071 kg† | 2,800 kg |
| **Instruments:** Optical and near-infrared spectral bands | Optical and near-infrared spectral bands | Multiple spectral bands | Multiple spectral bands |
| **Spatial resolution:** 3–5 m | ~1 m | 15–100 m‡ | 0.3–30 m‡ |

*When fully operational † Without instruments ‡ Depending on spectral frequency

automotive, smartphone and other consumer industries — including low-cost electronics, and sensors from high-end digital cameras. Using the latest technologies from these fast-paced industries also allows the rapid, continuous development of better and better satellites, says Will Marshall, chief executive of Planet Labs. And miniaturizing satellites reduces launch costs.

Because the swarms are still to be launched, scientists have yet to fully assess the quality of the imagery. But the satellites' spatial resolutions of 1–5 metres are much higher than those of most scientific satellites. Landsat, NASA's Earth-observation workhorse, for example, has a resolution of 15–100 metres depending on the spectral frequency, with 30 metres in the visible-light range.

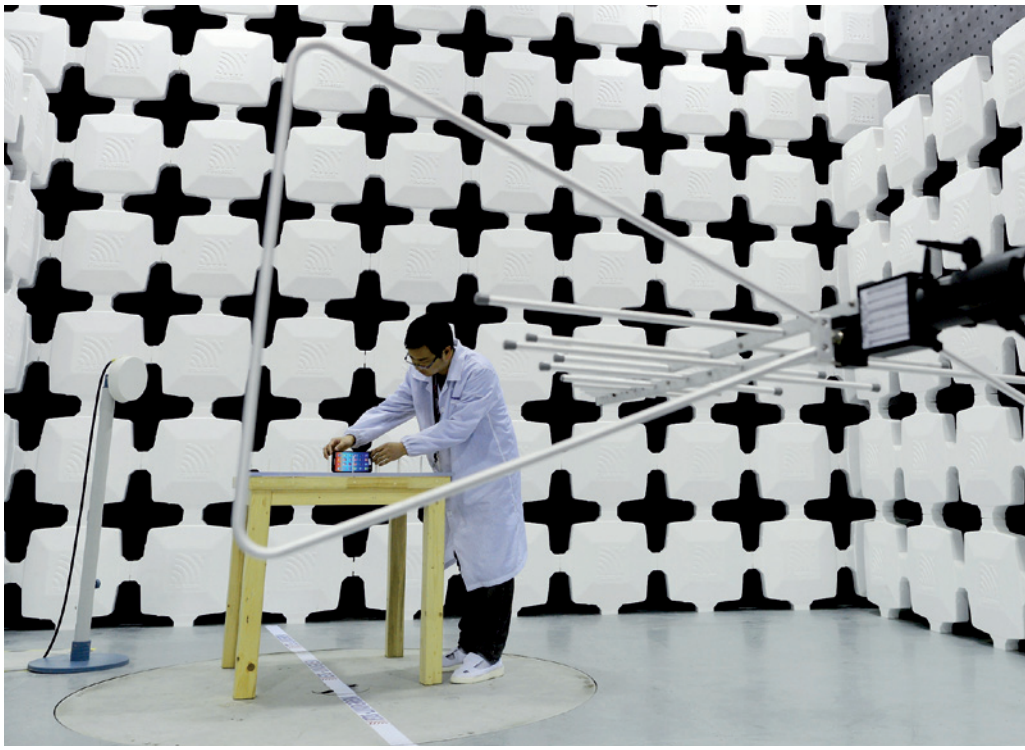
Such medium-resolution imagery is adequate for many purposes, but higher resolution can have benefits, says Dan Berkenstock, co-founder and chief product officer of Skybox Imaging. He points to a study published in November that found that the use of moderate-resolution Landsat imagery greatly underestimated forest loss in the Democratic Republic of the Congo (A. Tyukavina *et al. Environ. Res. Lett.* **8,** 044039; 2013).

Precision agriculture, a method that uses remote sensing to aid farm management, will also benefit from swarms, says Berkenstock, because the technology will be able to provide timely crop-yield and health estimates down to the level of rows of plants. Such detail could inform decisions on fertilizer and irrigation use, but is currently out of reach of most farmers.

However, spatial resolution is only part of the picture, says Mike Wulder, a researcher at the Canadian Forest Service in Victoria and a member of the Landsat science team. He uses remote sensing to study forests, and notes that good spectral and radiometric resolution (detection of small differences in wavelength and radiation, respectively), are essential for quantitative scientific analyses. "These very small satellites should not be expected to provide data that are similar or in competition with full-blown Earth-observing satellites," he says. "They occupy a different niche."

The scientific value of the swarm data will be "radically dependent" on quality issues, says Greg Asner, an Earth scientist at the Carnegie Institution for Science in Stanford, California. Stitching together such frequent-repeat imagery from so many satellites will be challenging, because performance will probably differ between satellites and vary over time, he argues.

But he is nonetheless excited at the prospect of constantly updated fresh imagery. "It will almost be like updating Google Earth each day," he says. ■



Research at computer firm Lenovo is helping to drive China's rising R&D spending.

CHINAFOTOPRESS VIA GETTY

RESEARCH FUNDING

# China tops Europe in R&D intensity

*Reforms to commercial and academic research systems still needed despite reaching spending milestone, say scientists.*

BY RICHARD VAN NOORDEN

By pouring cash into science and technology faster than its economy has expanded, China has for the first time overtaken Europe on a key measure of innovation: the share of its economy devoted to research and development (R&D).

In 2012, China invested 1.98% of its gross domestic product (GDP) into R&D — just edging out the 28 member states of the European Union (EU), which together managed 1.96%, according to the latest estimates of research intensity, to be released this month by the Paris-based Organisation for Economic Co-operation and Development (OECD).

The figures show that China's research intensity has tripled since 1998, whereas Europe's has barely increased (see 'Shooting star'). The numbers are dominated by business spending, reflecting China's push in the manufacturing and information- and communication-technology industries.

James Wilsdon, a science-policy analyst at the University of Sussex in Brighton, UK, says that China's R&D juggernaut is "astonishing", considering that the entire system emerged only after the end of the Cultural Revolution in 1976. In absolute terms, China's R&D spending is still almost one-third lower than that of Europe, but the new figures are "a significant milestone", says Wilsdon.

The reorientation of China's economy displays its soaring ambition. However, money does not buy innovation. Despite success in some areas, notably high-speed rail, solar energy, supercomputing and space exploration, leaders in China are concerned that innovation is lacking, say science-policy analysts. "Chinese leaders would like something equivalent to a Nobel prize, or a world-class product similar to an iPhone," says Denis Simon, an expert on Chinese science and innovation at Arizona State University

in Tempe. "But there is a lot of risk aversion within the Chinese R&D system that doesn't allow for entrepreneurial behaviour."

China's leaders recognize the issues: the government is now reviewing a 2006 long-term plan on science and technology, and will be taking advice from international experts in Beijing this month. Lan Xue, director of the China Institute of Science and Technology Policy at Tsinghua University in Beijing, expects some changes at the level of academic science. "I'm relatively optimistic that there will be improvement in how R&D programmes are managed and peer-reviewed."

In contrast to China's rapid rise, Europe's R&D spending has remained stagnant. The continent has made little headway in the past decade on a long-term target to reach 3% of GDP by 2020. "The European Commission has long warned that China is catching up in terms of R&D intensity," says Michael Jennings, a spokesman for research at the commission. "The EU needs a real push now to increase R&D spending in the public sector, but especially in the private sector."

One problem is that the commission cannot dictate business spending for individual member states. Another is the expansion of the EU, which has brought down average research intensity. OECD figures show the stark contrast between nations such as Germany, at 2.92% of GDP, and newer EU members such as Croatia, at 0.75%. Jennings adds, however, that an almost 30% boost to Horizon 2020, the EU research programme, is a good sign.

Some analysts argue that Europe does not need to be too worried by the stasis in research intensity. The number is an increasingly poor indicator of innovative activity, argues Kieron Flanagan, a science-policy analyst at the University of Manchester, UK. For example, it fails to pick up on innovation in the service-oriented industries that dominate many Western economies. An architectural or advertising firm could innovate while meeting the demands of a contract — making advances that could be widely copied and meaningfully affect an economy. Yet they would not count as R&D spending.

In China, meanwhile, "a great stodgy mass" of state-owned enterprises dominates commercial R&D spending — and they might actually suppress innovation, says Wilsdon. According to a study co-authored by Wilsdon and published in October 2013 by the innovation charity Nesta, based in London, the state companies might block more-inventive small and medium-sized enterprises. China, the study argues, is an "absorptive state": one that adopts and adapts incoming technologies from overseas but does little breakthrough research. However, Wilsdon points to a few eye-catching bright spots: privately held, globally minded companies that include the telecommunications firms Huawei Technologies and ZTE, the e-commerce giant Alibaba and the computer firm Lenovo.

***"Chinese leaders would like something equivalent to a Nobel prize, or a world-class product similar to an iPhone."***

China's emphasis on applied and product-development research means that funding for basic science remains low: only 5% of the country's total R&D is devoted to this, compared with 15–20% in other major OECD nations. That money has to support a larger number of researchers who are already poorly paid, says Xue. Many academics, he says, complement their salaries by taking on short-term projects for industry — work that can distract their focus from fundamental science problems.

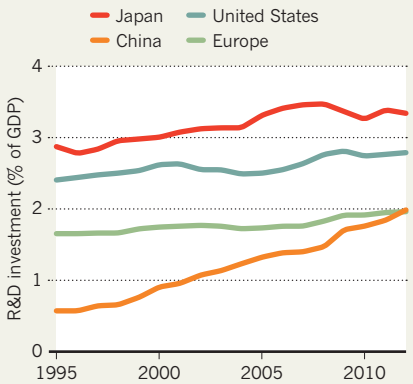
Funding and evaluation systems suffer other distortions, says Cong Cao, a science-policy analyst at the University of Nottingham, UK. Grant money is not disbursed transparently, and basic-research funding tends to go to eminent scientists and safe projects, he says, with academics judged mechanically on the number of publications that they author. A staggering rise in scientific output has not yet been matched by an equivalent rise in highly cited articles; swathes of patents are filed but rarely used. Wilsdon says that world-class research occurs at the country's top 30 universities and at Chinese Academy of Science institutes. "But it is still very patchy, and a lot of it is reliant on a relatively small number of outstanding scientists lured back from overseas," he says.

Simon adds that China's scientists need more independence and freedom to work on risky projects. Such changes might be on the way: Cao expects that at the forthcoming review of China's 2006 science plan, funding agencies will be told to be more transparent about their grants and grantees, and Chinese researchers will be allowed to use more of their funding to boost the salaries of research staff.

One of the plan's paramount goals seems to be right on target, however: China, unlike Europe, looks set to boost its research spending to 2.5% of GDP by 2020. ■

**SHOOTING STAR**

China has passed Europe in its investment in research and development (R&D) as a percentage of its gross domestic product (GDP).




SOURCE: OECD

# THE LEARNING MACHINES

*Using massive amounts of data to recognize photos and speech, deep-learning computers are taking a big step towards true artificial intelligence.*

**BY NICOLA JONES**

Three years ago, researchers at the secretive Google X lab in Mountain View, California, extracted some 10 million still images from YouTube videos and fed them into Google Brain — a network of 1,000 computers programmed to soak up the world much as a human toddler does. After three days looking for recurring patterns, Google Brain decided, all on its own, that there were certain repeating categories it could identify: human faces, human bodies and … cats[1].

Google Brain's discovery that the Internet is full of cat videos provoked a flurry of jokes from journalists. But it was also a landmark in the resurgence of deep learning: a three-decade-old technique in which massive amounts of data and processing power help computers to crack messy problems that humans solve almost intuitively, from recognizing faces to understanding language.

Deep learning itself is a revival of an even older idea for computing: neural networks. These systems, loosely inspired by the densely interconnected neurons of the brain, mimic human learning by changing the strength of simulated neural connections on the basis of experience. Google Brain, with about 1 million simulated neurons and 1 billion simulated connections, was ten times larger than any deep neural network before it. Project founder Andrew Ng, now director of the Artificial Intelligence Laboratory at Stanford University in California, has gone on to make deep-learning systems ten times larger again.

Such advances make for exciting times in

BRUCE ROLFF/SHUTTERSTOCK

artificial intelligence (AI) — the often-frustrating attempt to get computers to think like humans. In the past few years, companies such as Google, Apple and IBM have been aggressively snapping up start-up companies and researchers with deep-learning expertise. For everyday consumers, the results include software better able to sort through photos, understand spoken commands and translate text from foreign languages. For scientists and industry, deep-learning computers can search for potential drug candidates, map real neural networks in the brain or predict the functions of proteins.

"AI has gone from failure to failure, with bits of progress. This could be another leapfrog," says Yann LeCun, director of the Center for Data Science at New York University and a deep-learning pioneer.

"Over the next few years we'll see a feeding frenzy. Lots of people will jump on the deep-learning bandwagon," agrees Jitendra Malik, who studies computer image recognition at the University of California, Berkeley. But in the long term, deep learning may not win the day; some researchers are pursuing other techniques that show promise. "I'm agnostic," says Malik. "Over time people will decide what works best in different domains."

## INSPIRED BY THE BRAIN

Back in the 1950s, when computers were new, the first generation of AI researchers eagerly predicted that fully fledged AI was right around the corner. But that optimism faded as researchers began to grasp the vast complexity of real-world knowledge — particularly when it came to perceptual problems such as what makes a face a human face, rather than a mask or a monkey face. Hundreds of researchers and graduate students spent decades hand-coding rules about all the different features that computers needed to identify objects. "Coming up with features is difficult, time consuming and requires expert knowledge," says Ng. "You have to ask if there's a better way."

In the 1980s, one better way seemed to be deep learning in neural networks. These systems promised to learn their own rules from scratch, and offered the pleasing symmetry of using brain-inspired mechanics to achieve brain-like function. The strategy called for simulated neurons to be organized into several layers. Give such a system a picture and the first layer of learning will simply notice all the dark and light pixels. The next layer might realize that some of these pixels form edges; the next might distinguish between horizontal and vertical lines. Eventually, a layer might recognize eyes, and might realize that two eyes are usually present in a human face (see 'Facial recognition').

➲ NATURE.COM
Learn about another approach to brain-like computers:
go.nature.com/fktnso

The first deep-learning programs did not perform any better than simpler systems, says Malik. Plus, they were tricky to work with. "Neural nets were always a delicate art to manage. There is some black magic involved," he says. The networks needed a rich stream of examples to learn from — like a baby gathering information about the world. In the 1980s and 1990s, there was not much digital information available, and it took too long for computers to crunch through what did exist. Applications were rare. One of the few was a technique — developed by LeCun — that is now used by banks to read handwritten cheques.

> "OVER THE NEXT FEW YEARS WE'LL SEE A FEEDING FRENZY. LOTS OF PEOPLE WILL JUMP ON THE DEEP-LEARNING BANDWAGON."

By the 2000s, however, advocates such as LeCun and his former supervisor, computer scientist Geoffrey Hinton of the University of Toronto in Canada, were convinced that increases in computing power and an explosion of digital data meant that it was time for a renewed push. "We wanted to show the world that these deep neural networks were really useful and could really help," says George Dahl, a current student of Hinton's.

As a start, Hinton, Dahl and several others tackled the difficult but commercially important task of speech recognition. In 2009, the researchers reported[2] that after training on a classic data set — three hours of taped and transcribed speech — their deep-learning neural network had broken the record for accuracy in turning the spoken word into typed text, a record that had not shifted much in a decade with the standard, rules-based approach. The achievement caught the attention of major players in the smartphone market, says Dahl, who took the technique to Microsoft during an internship. "In a couple of years they all switched to deep learning." For example, the iPhone's voice-activated digital assistant, Siri, relies on deep learning.

## GIANT LEAP

When Google adopted deep-learning-based speech recognition in its Android smartphone operating system, it achieved a 25% reduction in word errors. "That's the kind of drop you expect to take ten years to achieve," says Hinton — a reflection of just how difficult it has been to make progress in this area. "That's like ten breakthroughs all together."

Meanwhile, Ng had convinced Google to let him use its data and computers on what became Google Brain. The project's ability to spot cats was a compelling (but not, on its own, commercially viable) demonstration of unsupervised learning — the most difficult learning task, because the input comes without any explanatory information such as names, titles or categories. But Ng soon became troubled that few researchers outside Google had the tools to work on deep learning. "After many of my talks," he says, "depressed graduate students would come up to me and say: 'I don't have 1,000 computers lying around, can I even research this?'"

So back at Stanford, Ng started developing bigger, cheaper deep-learning networks using graphics processing units (GPUs) — the super-fast chips developed for home-computer gaming[3]. Others were doing the same. "For about US$100,000 in hardware, we can build an 11-billion-connection network, with 64 GPUs," says Ng.

## VICTORIOUS MACHINE

But winning over computer-vision scientists would take more: they wanted to see gains on standardized tests. Malik remembers that Hinton asked him: "You're a sceptic. What would convince you?" Malik replied that a victory in the internationally renowned ImageNet competition might do the trick.

In that competition, teams train computer programs on a data set of about 1 million images that have each been manually labelled with a category. After training, the programs are tested by getting them to suggest labels for similar images that they have never seen before. They are given five guesses for each test image; if the right answer is not one of those five, the test counts as an error. Past winners had typically erred about 25% of the time. In 2012, Hinton's lab entered the first ever competitor to use deep learning. It had an error rate of just 15% (ref. 4).

"Deep learning stomped on everything else," says LeCun, who was not part of that team. The win landed Hinton a part-time job at Google, and the company used the program to update its Google+ photo-search software in May 2013.

Malik was won over. "In science you have to be swayed by empirical evidence, and this was clear evidence," he says. Since then, he has adapted the technique to beat the record in another visual-recognition competition[5]. Many others have followed: in 2013, all entrants to the ImageNet competition used deep learning.

With triumphs in hand for image and speech recognition, there is now increasing interest in applying deep learning to natural-language understanding — comprehending human discourse well enough to rephrase or answer questions, for example — and to translation from one language to another. Again, these are currently done using hand-coded rules and statistical analysis of known text. The state-of-the-art of such techniques can be seen in software such as Google Translate, which can produce results that are comprehensible (if

sometimes comical) but nowhere near as good as a smooth human translation. "Deep learning will have a chance to do something much better than the current practice here," says crowd-sourcing expert Luis von Ahn, whose company Duolingo, based in Pittsburgh, Pennsylvania, relies on humans, not computers, to translate text. "The one thing everyone agrees on is that it's time to try something different."

## DEEP SCIENCE

In the meantime, deep learning has been proving useful for a variety of scientific tasks. "Deep nets are really good at finding patterns in data sets," says Hinton. In 2012, the pharmaceutical company Merck offered a prize to whoever could beat its best programs for helping to predict useful drug candidates. The task was to trawl through database entries on more than 30,000 small molecules, each of which had thousands of numerical chemical-property descriptors, and to try to predict how each one acted on 15 different target molecules. Dahl and his colleagues won $22,000 with a deep-learning system. "We improved on Merck's baseline by about 15%," he says.

Biologists and computational researchers including Sebastian Seung of the Massachusetts Institute of Technology in Cambridge are using deep learning to help them to analyse three-dimensional images of brain slices. Such images contain a tangle of lines that represent the connections between neurons; these need to be identified so they can be mapped and counted. In the past, undergraduates have been enlisted to trace out the lines, but automating the process is the only way to deal with the billions of connections that are expected to turn up as such projects continue. Deep learning seems to be the best way to automate. Seung is currently using a deep-learning program to map neurons in a large chunk of the retina, then forwarding the results to be proofread by volunteers in a crowd-sourced online game called EyeWire.

William Stafford Noble, a computer scientist at the University of Washington in Seattle, has used deep learning to teach a program to look at a string of amino acids and predict the structure of the resulting protein — whether various portions will form a helix or a loop, for example, or how easy it will be for a solvent to sneak into gaps in the structure. Noble has so far trained his program on one small data set, and over the coming months he will move on to the Protein Data Bank: a global repository that currently contains nearly 100,000 structures.

For computer scientists, deep learning could earn big profits: Dahl is thinking about start-up opportunities, and LeCun was hired last month to head a new AI department at Facebook. The technique holds the promise of practical success for AI. "Deep learning happens to have the property that if you feed it more data it gets better and better," notes Ng. "Deep-learning algorithms aren't the only ones like that, but they're arguably the best — certainly the easiest. That's why it has huge promise for the future."

> "DEEP LEARNING HAS THE PROPERTY THAT IF YOU FEED IT MORE DATA, IT GETS BETTER AND BETTER."

Not all researchers are so committed to the idea. Oren Etzioni, director of the Allen Institute for Artificial Intelligence in Seattle, which launched last September with the aim of developing AI, says he will not be using the brain for inspiration. "It's like when we invented flight," he says; the most successful designs for aeroplanes were not modelled on bird biology. Etzioni's specific goal is to invent a computer that, when given a stack of scanned textbooks, can pass standardized elementary-school science tests (ramping up eventually to pre-university exams). To pass the tests, a computer must be able to read and understand diagrams and text. How the Allen Institute will make that happen is undecided as yet — but for Etzioni, neural networks and deep learning are not at the top of the list.

One competing idea is to rely on a computer that can reason on the basis of inputted facts, rather than trying to learn its own facts from scratch. So it might be programmed with assertions such as 'all girls are people'. Then, when it is presented with a text that mentions a girl, the computer could deduce that the girl in question is a person. Thousands, if not millions, of such facts are required to cover even ordinary, common-sense knowledge about the world. But it is roughly what went into IBM's Watson computer, which famously won a match of the television game show *Jeopardy* against top human competitors in 2011. Even so, IBM's Watson Solutions has an experimental interest in deep learning for improving pattern recognition, says Rob High, chief technology officer for the company, which is based in Austin, Texas.

Google, too, is hedging its bets. Although its latest advances in picture tagging are based on Hinton's deep-learning networks, it has other departments with a wider remit. In December 2012, it hired futurist Ray Kurzweil to pursue various ways for computers to learn from experience — using techniques including but not limited to deep learning. Last May, Google acquired a quantum computer made by D-Wave in Burnaby, Canada (see *Nature* **498,** 286–288; 2013). This computer holds promise for non-AI tasks such as difficult mathematical computations — although it could, theoretically, be applied to deep learning.

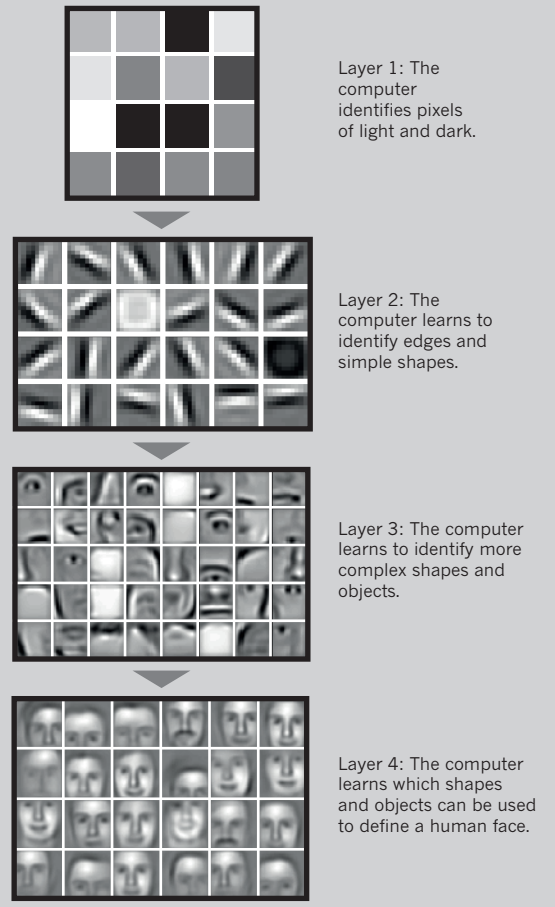
Despite its successes, deep learning is still in its infancy. "It's part of the future," says Dahl. "In a way it's amazing we've done so much with so little." And, he adds, "we've barely begun". ■

**Nicola Jones** *is a freelance reporter based near Vancouver, Canada.*


1. Le, Q. V. *et al.* Preprint at http://arxiv.org/abs/1112.6209 (2011).
2. Mohamed, A. *et al. 2011 IEEE Int. Conf. Acoustics Speech Signal Process.* http://dx.doi.org/10.1109/ICASSP.2011.5947494 (2011).
3. Coates, A. *et al. J. Machine Learn. Res. Workshop Conf. Proc.* **28,** 1337–1345 (2013).
4. Krizhevsky, A., Sutskever, I. & Hinton, G. E. In *Advances in Neural Information Processing Systems 25*; available at go.nature.com/ibace6.
5. Girshick, R., Donahue, J., Darrell, T. & Malik, J. Preprint at http://arxiv.org/abs/1311.2524 (2013).


## FACIAL RECOGNITION

Deep-learning neural networks use layers of increasingly complex rules to categorize complicated shapes such as faces.




IMAGES: ANDREW NG


# CAUGHT ON CAMPUS

*Violent incidents at academic institutions have spurred universities to adopt formal procedures designed to keep campuses safer. But do the tactics work?*

BY BRENDAN MAHER

In many pictures of her online, Kayla Bourque looks like a typical college student: there are selfies of her on a coastal holiday, or smirking mischievously after an experiment in hair colour. But in 2012, Bourque, then a criminology student at Simon Fraser University in British Columbia, told a classmate that she fantasized about killing a homeless person and that she was studying forensics so that she could get away with it. She also talked about killing her family pets and neighbourhood cats.

The classmate told a teaching assistant what Bourque had said, and the department chair called campus security. This triggered a formal process called a threat assessment, in which security, university administrators and outside consultants gathered evidence and evaluated Bourque's recent behaviour. They took the allegation seriously, says Stephen Hart, a forensic psychologist at Simon Fraser who advised on the case. "Often something like this is a cry for help," he says. But her actions on several occasions suggested that she might pose a threat to other students, so simply referring her to the university's outpatient mental-health services would not suffice. The team notified the local police, and told Bourque that she would not be

TIM SLOAN/AFP/GETTY

able to return to university without a thorough psychological evaluation.

Then, while university employees were packing up her dorm room, they found what has been described in court documents as a 'kill kit': a bag containing a kitchen knife, a razor blade, latex gloves, a syringe and plastic ties — the kind used to restrain people. "They realized that this wasn't just a call for help," says Hart. The discovery led to a search warrant for her computer, on which police found violent pornography, disturbing artwork and more selfies, including one of her standing naked next to her disembowelled dog, Molly.

Bourque spent nine months in custody in 2012 for killing Molly, as well as her cat Snowflake, and for possession of a weapon. When she was released it was with an impressive list of probationary conditions, including not using the Internet unsupervised, informing anyone she interacts with about her crimes, never owning a pet, and staying away from Simon Fraser. As horrifying as the case is, Hart sees it as a major triumph for the growing field of threat assessment.

Although they are exceedingly rare, the number of violent incidents reported on college and university campuses has been increasing. Recently, academic institutions have served as the backdrop to a series of highly publicized attacks — and sometimes scientists are the central figures. In 2010, Amy Bishop, a biology professor, gunned down three fellow faculty members at the University of Alabama at Huntsville after being denied tenure (see *Nature* **465,** 150–155; 2010). Two years later, James Holmes withdrew from his PhD studies in neuroscience at the University of Colorado Denver about a month before killing 12 people and wounding 58 at a cinema in Aurora.

In many cases, such events are preceded by an escalation in threatening or aberrant behaviour. Holmes had told a university psychiatrist that he fantasized about killing, and Bishop's behaviour allegedly prompted a dean and provost at the university to request police protection months before the attack. Both of these cases are subjects of pending wrongful-death lawsuits. But could the attacks have been prevented?

That is the goal of threat assessment, in which organizations adopt formal procedures to identify and mitigate a dangerous situation before it explodes into violence. Threat-assessment teams and plans are becoming standard at colleges and universities in the United States, and are mandatory in some states, including Virginia, Connecticut and Illinois. Other countries are following suit. "The biggest push we've seen has been in higher education," says Marisa Randazzo, a social psychologist and former US Secret Service agent who works with a threat-assessment consulting firm.

**"WE LIVE IN ONE OF THE MOST VIOLENT SOCIETIES. ANYTHING WE CAN DO TO MITIGATE RISK IS OF VALUE AND SOMETHING IMPORTANT TO CONSIDER."**

It is difficult to prove that the tactics work, and there are concerns that they may tread on civil liberties, but many see threat assessment as a necessary part of emergency preparedness. "We live in one of the most violent societies," says Reid Meloy, a forensic psychologist at the University of California, San Diego. "Anything we can do to mitigate risk is of value and something important to consider."

## ROOTS OF VIOLENCE

Talk to anyone in the threat-assessment field about violence on college campuses, and Gene Deisinger's name will inevitably come up. Deisinger was clinical director of the counselling centre at Iowa State University in Ames when the institution decided to build a threat-assessment team. A number of events influenced that decision: in 1986, a former computer-science student set fire to the house of one of his professors, killing two of the professor's children. Then, in 1991, a young physicist at the University of Iowa in Iowa City killed five people and himself, reportedly because he had been passed over for a thesis prize. In response to these and some other incidents, Loras Jaeger, then chief of Iowa State's campus police department, asked Deisinger if he would help to create a threat-management team. "I didn't know what one was, and so I started researching," Deisinger says.

**NATURE.COM**
For a podcast and interview with Gene Deisinger, see:
go.nature.com/qxjgj4

The US Secret Service — tasked with protecting the president and other public officials — had a long history of developing methods to assess threats, but that work was not publicly available at the time. Deisinger turned to research on dangerous behaviour, workplace violence and dealing with students in crisis. From that work, he developed an approach for identifying behavioural concerns and intervening in a campus setting. He had a team up and running by 1994. The duties grew quickly, he says, so Jaeger created a full-time position for him within the university police department. During the first month, Jaeger asked Deisinger to report to the state police academy for training, something the 33-year-old psychologist was not eager to do at that stage in his career. But he is glad that he relented. As "a psychologist that carries a badge and a gun, a lot of doors open up", he says.

A campus threat-assessment team is interdisciplinary, and includes law-enforcement professionals, psychologists, academic administrators, representatives from student services and human resources and legal counsel. When someone reports a suspicious behaviour, such as a threat from a student, the team often starts by confronting the person about the behaviour. They may talk to peers, advisers and teachers.

By studying past attacks through the lens of psychology, researchers have identified a range of behaviours and environmental factors that may conspire to trigger violence. Individuals may exhibit extreme or sudden changes in behaviour, alienate themselves or others, or adopt unhealthy interests in weapons or violent acts. Environmental factors may include a tolerance to aggressive interactions in a workplace, an unresolved conflict, or the existence of cliques or pecking orders. And there are often precipitating events. These could be personal conflicts or work-life pressures — such as not getting tenure or a key grant — that an individual has had trouble dealing with appropriately.

Empirical data on attacks suggest that there is a 'pathway to violence': there may be some form of grievance, the development of an intention to do harm, then research, planning and preparation. Bourque, for example, told psychologists that she had been taking a bus into town to check out potential victims, and her 'kill kit' suggested an advanced stage of preparation. But the data can tell only so much, because such attacks are rare. There are no simple checklists and no simple profile of an attacker.

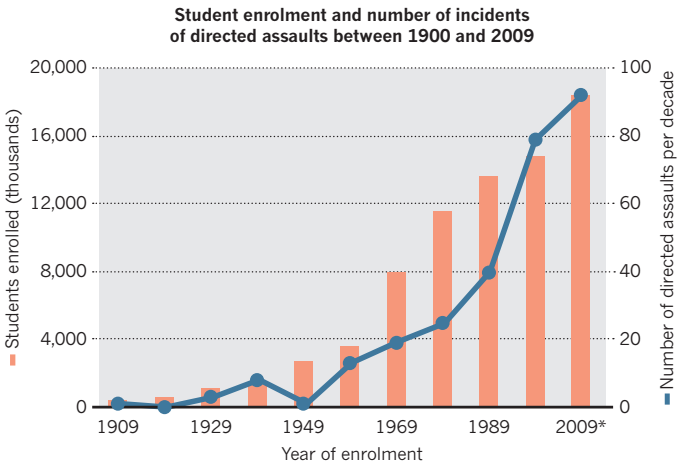
If concerns are legitimate and a threat-assessment team decides that a person may be on a trajectory towards violence, then the group works to manage the threat, often by putting the individual in touch with support or mental-health services, or by working out a way to resolve the environmental factors contributing to the situation. Team members may make regular visits, or what Deisinger describes as 'coffee dates'. These are meant to help to keep an eye on people, and unless the person has violated the law or a university rule they are considered voluntary. Most of them are cordial: "I've established a rapport with you that you would at least allow me through the front door, which would allow me to see your living circumstances," says Deisinger. "Is there evidence of psychotic deterioration? Are there weapons stacked in the corner? Are you taking care of yourself? Is there food there? Is your hygiene intact?"

It all sounds a bit Big Brother-ish, and Deisinger doesn't pretend otherwise. These individuals, he says, "know that I'm doing this. We don't play games with it, most of the time." Still, he says, it can be surprising how open people often are. Bourque "talked freely" to the threat-assessment team that evaluated her, says Hart, and even gave them permission to pack up her residence, which was where they spotted the 'kill kit'.

Threat assessment works only when people have signalled an intent to do harm. Luckily, these signals often appear. In the 1990s, the Secret

## DANGEROUS TREND

A study of attacks reported on campuses in the United States since 1900 showed that the increasing frequency of such incidents largely followed rising student numbers, but enrolment alone may not fully explain the increase.




*Data for 2009 are projected.

SOURCE: REF. 2

Service looked at 83 individuals who had attacked or come close to attacking a prominent public official or public figure. It showed that 63% had communicated some sort of threat in advance, although rarely to the intended target[1]. "The people who carry out these acts, they typically tell someone what they're planning to do," says Randazzo. "We've seen many cases where they broadcast it on social media."

### 20/20 HINDSIGHT

In November 2005, campus security at Virginia Polytechnic Institute and State University in Blacksburg received a report about Seung-Hui Cho, a South Korean-born student studying English who was allegedly harassing a female student. A room-mate added that Cho had made comments about contemplating suicide. Cho was assessed three times and said that the suicidal statements were a joke. He was briefly hospitalized.

In February and March of 2007, Cho bought two pistols; in April, he killed 32 people in what is, so far, the deadliest campus shooting in history.

The incident launched a national study on campus attacks by the Secret Service, the US Federal Bureau of Investigation and the US Department of Education. It collected information on 272 incidents in the United States between 1900 and 2008. The study showed that such events are rare — but that they are increasing in frequency[2]. For instance, the report catalogues only 25 incidents between 1970 and 1979, but 83 between 2000 and 2008 (see 'Dangerous trend'). The rise in student numbers is certainly a factor in the increase, and it is likely that more incidents are being reported than in the past. But those changes might not fully account for the trend, says Andre Simons, a behavioural scientist with the FBI who is working on a follow-up to the report.

The study also revealed the varied nature of attacks occurring in a university setting. Students or former students accounted for 60% of the perpetrators, and another 11% were employees. But the remaining 29% were not official members of the campus.

It is hard to say whether the individuals posed clear threats before their attacks. Nearly 30% of perpetrators displayed threatening behaviour, such as stalking or harassing, making verbal or written threats, or being physically aggressive towards their targets. Another 20% exhibited at least some sort of concerning behaviour to a friend, family member, work associate or police officer. But such behaviours could be vague and general, and were not always reported.

In Cho's case, there were signs of aberrant behaviour, but no process was in place to follow him in an exhaustive way, says Deisinger. After the shooting, Deisinger consulted with Virginia Tech to help build a threat-assessment team there, and he was eventually hired as the university's chief of police. The Virginia Tech shooting was a pivotal case that spurred more universities to develop such teams, says Meloy, although how many exist is unclear. A self-reported survey from 2012 found that 92% of universities and community colleges in the United States have some sort of team in place[3], but it included other kinds of behavioural-intervention teams that do not typically work directly with the police. The trend is not limited to the United States: universities in Australia have increasingly been taking an interest in threat-assessment procedures; and an estimated ten universities in German-speaking countries have established or developed plans for teams, says Jens Hoffmann, a psychologist at the Institute of Psychology and Threat Management in Darmstadt, Germany, and co-editor with Meloy of the *International Handbook of Threat Assessment* (Oxford Univ. Press, 2014).

### SAFE AND SOUND?

These systems and tactics for managing threats cannot stamp out all targeted violence. In 2009, after Virginia Tech had established its threat-assessment team but before Deisinger had arrived, a graduate student in agricultural economics beheaded a woman who had rebuffed his romantic advances. And a team was in place at the University of Colorado where Holmes made threats to a mental-health professional, but Holmes left the university shortly after, making it difficult to follow up on the case.

There are bound to be missed signals, says Deisinger. And threat assessment is only as good as the vigilance of a community, because it relies heavily on reporting. But in the wake of recent high-profile shootings, this vigilance has improved, says Randazzo. "People are reporting things that seem not right in ways that they didn't in the past." Many cite the "see something, say something" campaigns that have blanketed New York City since the terrorist attacks of 11 September 2001 as having helped to encourage people to report suspicious behaviour. But a team is useless if no one knows about it, so Deisinger and others have tried to spread the word by creating websites and fliers, as well as holding training sessions on how to deal with inappropriate behaviour.

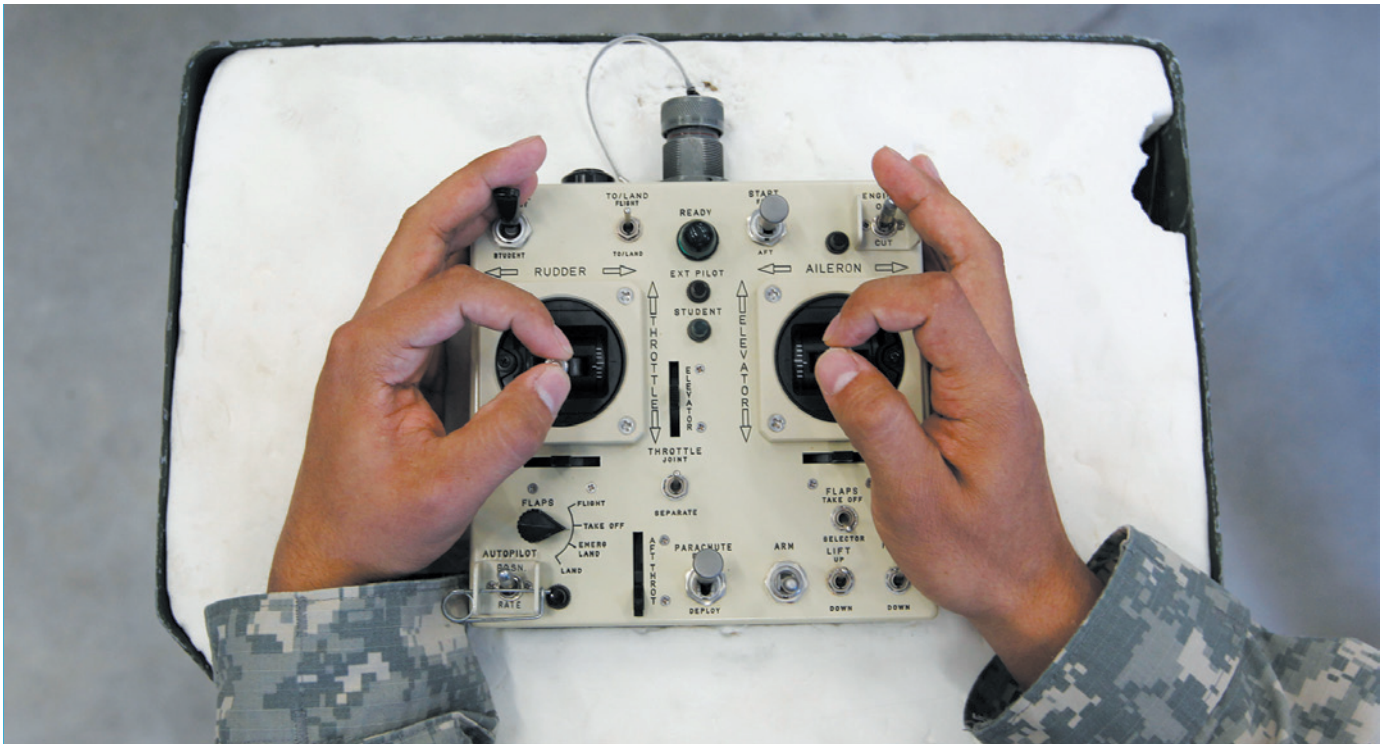
There are costs to all this watchfulness, however, says Joe Cohn, legislative and policy director for the Foundation for Individual Rights in Education in Philadelphia, Pennsylvania. "It's not unusual for universities to engage in behaviours that chill freedom of speech in the name of safety," he says. He cites recent examples in which a student was expelled for protesting over the construction of a parking garage and a professor was reported to a threat-assessment team for hanging posters with aggressive messages outside his office. He urges teams to include civil libertarians to better ensure that universities do not encroach on people's rights.

It is also difficult to prove that having a threat-assessment team makes a campus any safer. There are no standards for how to report a successful case, and privacy concerns make the sharing of data complicated. Deisinger says that his team tracks cases to see whether interventions have improved the situation for the individual and the people around him or her. "Most of them, we can resolve to a level that is akin to the day-to-day moderately inappropriate behaviour," he says. It is not ideal, "but it's liveable", he adds. The field is trying to set standards and collect data on how well the process works. Phase two of the FBI's campus-attacks study, which will focus on attacks that happened between 1985 and 2010, may fill in some of these holes.

These are all concerns that weigh heavily on Deisinger. But what worries him most is the thought that someone, somewhere, is planning something that no one can anticipate. "I'm often asked, because of the cases we work, 'How do you sleep at night?'" He does not worry so much about people who have already been identified by his team. "It's the cases I don't know about," he says, "that give me difficulty sleeping." ■ SEE EDITORIAL P. 131

**Brendan Maher** *is a features editor for* Nature *based in New York.*

1. Fein, R. A. & Vossekuil, B. *J. Forens. Sci.* **44,** 321–333 (1999).
2. Drysdale, D. A., Modzeleski, W. & Simons, A. B. *Campus Attacks: Targeted Violence Affecting Institutions of Higher Education.* Available at http://go.nature.com/cbxfhy (2010).
3. Van Brunt, B., Sokolow, B., Lewis, W. & Schuster, S. NaBITA Team Survey. Available at www.nabita.org (2012).




MICHAELA REHLE/REUTERS/CORBIS

Remote control of drones is one of the many contributions science and technology have made to war.

MILITARY SCIENCE

# The evolving science of war

**Sharon Weinberger** assesses two studies probing the roles of physics and psychology in conflicts past, present and future.

Barry Parker's chronicle of the interplay between the military and science, *The Physics of War*, is largely a record of people developing more effective ways to kill each other. So it is poignant that Parker, a physicist, begins the book with a passage on a battle that took place more than 3,000 years ago in what is now Syria, a country in the middle of a bloody civil war threatening to draw in world powers. It seems that fundamentals of warfare have not changed, but with the advent of science and the creation of more-powerful weapons, the stakes are now higher.

Physics, Parker argues, has enabled much of the killing. For thousands of years people have used its principles to build increasingly powerful weapons, even before they understood what made the devices work. Weapon by weapon, and in excruciating detail, Parker shows how a mix of tinkering, basic maths and physics — including, much later, nuclear physics — enabled the development of weapons of war, from the chariots of ancient Syria to modern thermonuclear weapons.

That is a lot of ground to cover, and Parker's book is best read as a primer for those interested in the science of weapons and their contributions to various battles. It is on less solid ground in helping us to understand when military leaders realized that advancing science as a discipline could aid warfare. At one point, for example, Parker writes that "Napoleon studied physics along with mathematics and astronomy in military school, and knew the importance of science to war". In the same paragraph, however, he states that there is "no indication" that Napoleon "took a lot of interest in physics, or science, in general".

By contrast, Michael Matthews' lively and engaging *Head Strong* makes the weighty argument that psychology is emerging as the science that will make the difference in twenty-first-century warfare. War is not just about killing, he argues; it is about understanding the enemy, and ourselves. Matthews, a military psychologist, makes a valiant case, noting how psychology has contributed to everything from selecting leaders to helping soldiers navigate foreign cultures. He predicts that it will one day help to produce drugs "capable of regulating the brain's response to combat stress", perhaps eliminating post-traumatic stress disorder.

**The Physics of War: From Arrows to Atoms**
BARRY PARKER
*Prometheus Books: 2014.*

**Head Strong: Psychology and Military Dominance in the 21st Century**
MICHAEL D. MATTHEWS
*Oxford University Press: 2013.*

When Matthews writes about his own research on the psychology of soldier performance and leadership, or his experience as a professor at the United States Military Academy at West Point in New York, the book springs to life. He shows how psychological methods have challenged some of the military leadership's entrenched beliefs about gender, citing a study he was involved in that surveyed Air Force base commanders' attitudes about women. Almost every commander told a story of how the pilot of a crashed, burning aeroplane died because a female firefighter was not strong enough to carry him out. The story, Matthews later found, was apocryphal.

His larger point is how science, particularly psychology, can inform decisions about

integration. In another example, he notes that West Point, which trains officers, targets women's enrolment at about 15% to reflect the ratio of women in the military. That sounds noble; but he notes that West Point tries (and has so far failed) to recruit African Americans at a rate reflecting their representation in the recruiting-age population. Were the same rule applied to women, he writes, they should make up half the class. West Point spokesman Francis DeMaro declined to comment on goals linked to gender or ethnicity, instead providing numbers on the most recent entering class (16% women, 10% African Americans) that seem to bolster Matthews's argument. "We strive to ensure our cadet population is representative of the soldiers they will lead," says DeMaro.

Matthews stumbles a bit when talking about the importance of psychology in understanding foreign cultures. He praises the Human Terrain System, the well-intentioned but troubled US programme that embeds social scientists into teams that deploy with the military (see *Nature* http://doi.org/bxmgsw; 2011). Matthews engages in the same kind of oversimplification of cultural knowledge that underlies the problems facing these teams. He recalls how a US military commander in Iraq learned that arriving heavily armed at meetings with community leaders was a "major social blunder" (as it might be, of course, in most cultures).

By focusing on the progression of weapons, Parker misses the point at which physics was overtaken by other fields, including psychology, as disciplines crucial to warfare. But Matthews, in focusing so closely on current and future applications of psychology, omits mention of one of most important military psychologists.

In the 1960s, the US Department of Defense's Advanced Research Projects Agency hired psychologist J. C. R. Licklider to create a behavioural sciences office. It was his unique insights into how man would interact with machine in the future that laid the foundation for ARPANET, the precursor to the Internet. Today, networked computers are as key to military command and control as they are to modern society. It could be argued that, thanks to Licklider, military psychology has already revolutionized war. Whether it will help the United States to win future wars is another matter. ■

**Sharon Weinberger** *is a Global Fellow at the Woodrow Wilson International Center for Scholars in Washington DC. Her book about the Defense Advanced Research Projects Agency will be published in 2015.*
*e-mail: sharonweinberger@gmail.com*

PSYCHOLOGY

# Feeling the fear

**David Adam** applauds the autobiography of a high-flyer confronting his own nervous suffering head-on.

Scott Stossel is, in his own words, a "quivering, quaking, neurotic wreck". He is frightened of flying, vomiting and cheese. He has thrown tennis matches from a winning position just to get off the exposed stage of the court, and struggles to control his bowels. For three decades he has been a regular in the offices and clinics of psychiatrists, psychologists and psychoanalysts, and a testing ground for whatever treatment, drug or quack therapy they thought might bring some relief.

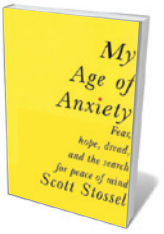
Stossel is also a married father of two and editor of *The Atlantic* magazine. His terrific book *My Age of Anxiety* is his attempt to reconcile those two worlds, and offers an unsparing and unsentimental look at a subject that many keep hidden: mental illness.

Stossel suffers from anxiety, a condition that he identifies early on as tricky to define. Is anxiety the list of symptoms offered by psychiatrists? The biological response to threat that we share with animals? The social consequence of the shared knowledge of our mortality? Or the chemical consequence of misfiring neurotransmitters and brain circuitry?

Books exploring personal experiences of mental illness tend to be either overwrought accounts of personal trauma that shed little light on the world beyond the author's nose, or the more detached observations of scientists and medics. It is rare to find works that bridge these objectives, which is one reason that the writer Andrew Solomon achieved such success with *The Noonday Demon* (Chatto & Windus, 2001), his personal and scientific account of depression. Stossel's book deserves a place on this higher shelf.

***"Stossel offers an unsparing and unsentimental look at a subject that many keep hidden."***

*My Age of Anxiety* covers all the academic ground one would expect. We get the biological idea that anxiety is an unsuited modern deployment of an atavistic fight-or-flight physiological response to threat, the psychological basis for conditioned responses — that anxiety is a learned, if inappropriate, fear — and the nascent attempts to link mind and body through brain scans and genetics. With help from some friendly neuroscientists, Stossel finds he has a variant of the *SERT* gene implicated in anxiety.




**My Age of Anxiety: Fear, Hope, Dread, and the Search for Peace of Mind**
SCOTT STOSSEL
*Knopf: 2014.*

Stossel is also aware of current controversies in psychiatry. He gives fair voice, for example, to both sides in the debate over the usefulness of pharmaceuticals, talk therapies and the shift from viewing anxiety as a social and philosophical issue to a disorder of chemical and electrical signals.

And he shows his skills as a writer with colourful and moving accounts of traumatic personal episodes. As a child and adolescent he suffered extreme separation anxiety and, aged 13, would wake the neighbours and ask them to call the police when his parents were out. The treatments were often equally grisly. Given an emetic syrup to make him vomit as exposure therapy to rid him of his phobia, he endures only hours of severe nausea and painful retching.

Stossel addresses the heterogeneous ingredients of anxiety by trying to cover them all — as if a sense of completeness alone can bind them together. His policy of full disclosure may not always be to everyone's tastes: an anecdote of a blocked toilet and a meeting with John F. Kennedy Jr, for one, feels gratuitous. But the approach also offers useful reminders of the human cost of taking strong positions on the use of drugs and other areas of scientific and medical uncertainty. Poised between a psychiatrist who puts him on drugs and a therapist who urges him to abandon them, Stossel finds himself lying to the therapist to spare her feelings when he returns to the psychiatrist.

One of Stossel's motives is the hope that the book might bring him peace. Still, he writes, "If it's relief from nervous suffering that I crave, then burrowing into the history and science of anxiety, and into my own psyche, is perhaps not the best way to achieve it."

We should all hope it works: the man is due a break. ■

**David Adam** *is* Nature*'s Editorial and Columns editor. His first book,* The Man Who Couldn't Stop: OCD and the True Story of a Life Lost in Thought, *will be published in April 2014.*

# COMMENT





MIT MUSEUM

**Physicists Edward Bowen (left), Lee DuBridge (centre) and I. I. Rabi work on a cavity magnetron in the 1940s.**

# Shut up and calculate!

Practical, interdisciplinary ways of working forged during the Second World War had a lasting impact on a generation of physicists and their findings, says **David Kaiser**.

On 17 October 1940, Karl Compton, president of the Massachusetts Institute of Technology (MIT) in Cambridge, made a hasty telephone call from Washington DC to a colleague back on campus. Could MIT spare some modest space to host an urgent, top-secret defence project? After making some quick assessments, Compton's assistant reported that MIT could shuffle some other laboratories to accommodate the facility. With that phone call, the Radiation Laboratory, or 'Rad Lab', was born. The laboratory had an enormous impact on the course of the Second World War. Arguably, its impact on science was even greater.

Within weeks of Compton's call, a skeleton staff at the Rad Lab was hard at work trying to improve on a British-designed cavity magnetron, which they hoped could become the centrepiece of a type of short-wavelength radar. When the laboratory began operation — more than a year before the United States entered the Second World War — the staff consisted of 20 physicists, three security guards, two stockroom clerks and a secretary. By the war's end, the lab had swollen to 4,000 people and was managing development contracts worth US$1.5 billion (nearly $20 billion in 2013 dollars)[1].

The Allied nuclear-weapons project, code-named the Manhattan Project, grew even faster. Again building on early insights from a British team, the Manhattan Project, coordinated from the Los Alamos laboratory in New Mexico, ballooned to encompass 125,000 people working at 31 facilities across North America. By the time the atomic bombs were dropped on Hiroshima and Nagasaki in August 1945, the project had cost $1.9 billion (about $25 billion today)[2]. Together, the radar and atomic-bomb projects amounted to about 1% of US military expenditure during the war: modest on the scale of defence appropriations, but utterly unprecedented for the academic scientists and engineers caught up in the war projects.

And it was more than just the ▶

budgets that grew. In both projects, physicists, chemists, metallurgists and their colleagues found themselves working in huge groups with larger-than-life equipment. Isotope-separation plants in Oak Ridge, Tennessee, stretched the length of a city block; the nuclear-reactor facilities in Hanford, Washington, required more than half a billion cubic metres of concrete.

After the war, many physicists dismissed their work on such sprawling wartime projects as temporary distractions: an important but limited hiatus from their 'real' scientific research. One Rad Lab veteran even composed a song soon after the war closing with the memorable line, "Oh, dammit! Engineering isn't physics, is that plain? Take, oh take, your billion dollars, let's be physicists again".[3]

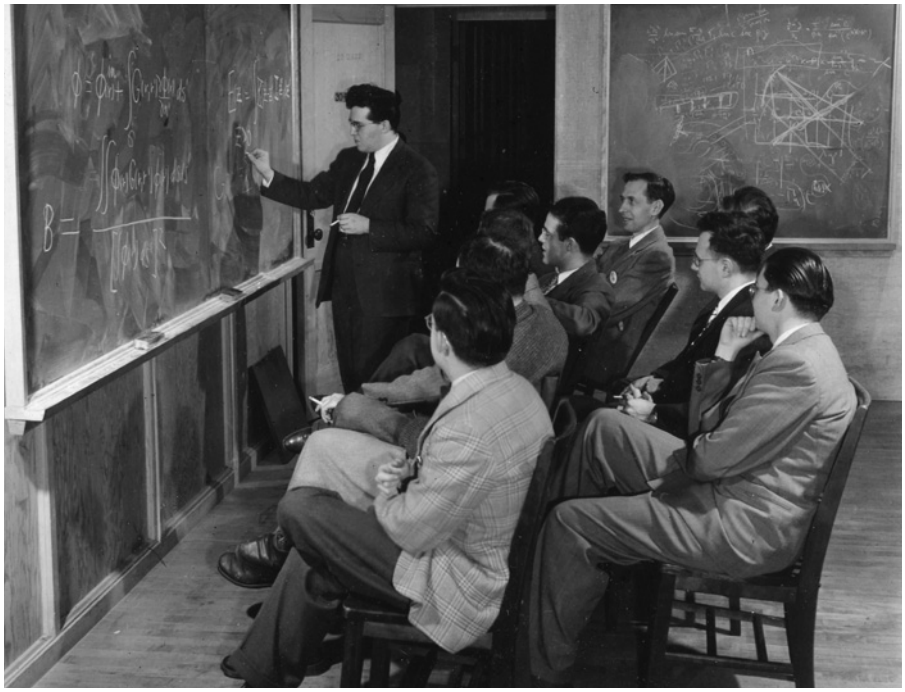
Despite the songwriter's plea, scientists did not return to the antebellum status quo. Instead, many characteristics of the wartime projects became the new normal, even in peacetime. The war cast a long shadow on how science is organized and funded, and even on the methods and questions that many scientists pursued throughout their careers.

## COMMON PURPOSE

Until the war, most scientific research in the United States had been supported by private foundations, local industries and undergraduate tuition fees. After the war, scientists experienced a continuity — even an expansion — of the wartime funding model. Almost all support for basic, unclassified research (as well as for mission-oriented defence projects) came from the federal government.

In 1949, 96% of all funding for basic research in the physical sciences in the United States came from defence-oriented federal agencies, including the Department of Defense and the then-new Atomic Energy Commission, successor to the Manhattan Project. In 1954 — four years after the establishment of the civilian US National Science Foundation — 98% of funding for basic research in the physical sciences came from federal defence agencies. And the scale of funding was unlike anything before the war. By 1953, funding for basic research in the United States had leapt to 25 times what it had been in 1938 (in constant dollars, adjusting for inflation)[4]. The fire hose of federal spending paid for all kinds of interesting research.

Much of the work was conducted in institutions modelled on wartime examples. Defence projects during the war had thrown together experts from many different fields of science and engineering to work towards common objectives, rather than grouping specialists by disciplines. The enormous time pressures and shared goals of war work forced scientists and engineers to craft



MIT MUSEUM

**Julian Schwinger (standing) with colleagues at MIT's Radiation Laboratory during the Second World War.**

effective means of communicating with each other. Mathematical rigour and abstruse theoretical derivations were worth little if colleagues from other specialities could not build on the results.

Veterans of the intense, multidisciplinary wartime projects came to speak of a new type of scientist. They touted the war-forged 'radar philosophy' and the quintessential 'Los Alamos man': a pragmatist who could collaborate with everyone from ballistics experts to metallurgists, and who had a gut feeling for the relevant phenomena without getting lost in philosophical niceties[5].

Leading scientists and policy-makers actively sought to continue the wartime spirit of collaboration across disciplines.

*"This war-forged pragmatism produced enormously impressive research."*

The Atomic Energy Commission oversaw a new network of national laboratories to pursue both civilian and defence research. The labs featured interdisciplinary teams that mixed physicists, mathematicians and chemists with engineers of many stripes[6]. A similar set-up appeared across dozens of US universities: facilities straddling several academic departments, such as the Research Laboratory for Electronics and the Laboratory for Nuclear Science and Engineering, both founded at MIT by the end of 1945 (ref. 7).

The facilities hummed with surplus equipment and know-how culled from the wartime projects. Physicist Bruno Rossi, for one, studied cosmic rays after the war by adapting the sensitive timing circuits he had built at Los Alamos to measure nuclear-fission rates[5].

Similarly, just months after the end of hostilities, self-described 'boffins' who had spent the war working on radar turned their attention to the impossibly small and the cosmically large. Some began to build radio telescopes and aimed them at the heavens. An international community coalesced, linking the Jodrell Bank telescopes near Manchester, UK, and the Parkes telescope in New South Wales, Australia, to similar instruments dotted across North America — from the California Institute of Technology in Pasadena to the National Radio Astronomy Observatory in Green Bank, West Virginia[8]. And in 1947, using repurposed microwave-frequency electronics left over from his wartime radar work, physicist Willis Lamb of Columbia University in New York measured a tiny shift — of about one part in a million — in the energy levels of an electron in the 2s and 2p orbitals of a hydrogen atom. Lamb's remarkable achievement challenged physicists' prevailing understanding of the vacuum — the mysterious state of lowest-possible energy[9].

One of the first to hear about the Lamb shift was physicist Julian Schwinger, who before the war had been a rising star in quantum theory. Like so many physicists at the Rad Lab, Schwinger had been forced to rethink his approach to calculation. Elegant derivations from first principles — which often proved tractable only when applied to idealized situations — were of little value to the many colleagues who needed to fine-tune electronics components for maximum efficiency. Instead, as Schwinger himself later recalled, he internalized from the engineers a

Students protest against military research at the Massachusetts Institute of Technology in 1969.

AP

modular, 'effective circuit' approach. Rather than calculate the total electrical resistance of a complicated component from the lofty heights of Maxwell's equations, he could 'blackbox' each component, substituting its overall resistance as determined from measurements of inputs and outputs. The niceties of how current flowed between constituent parts of a given component mattered much less to the main objective — improving radar designs — than did the effect of that component in a given circuit[5,9].

Schwinger approached the Lamb shift with his Rad Lab lessons still fresh. Since the 1930s, senior theorists had tried to calculate the effects of subtle quantum fluctuations from first principles. Maddeningly, their equations always broke down, producing unphysical infinities instead of finite answers. Schwinger rearranged his equations in terms of measurable inputs and outputs, just as his engineering colleagues at the Rad Lab had done with real-world electronics. By recasting the calculation, Schwinger managed to calculate the effects of quantum fluctuations on the electron's energy levels and obtain an answer that matched Lamb's measurement to an extraordinary precision. As it turned out, Japanese physicist Sin-Itiro Tomonaga had accomplished the same goal a few years earlier. Tomonaga's work on radar during the war had proven similarly essential to his theoretical approach[5,9].

## PHILOSOPHY RETURNS

This war-forged pragmatism produced enormously impressive research and influenced a generation of leading scientists. Their approach to basic research — and the institutions in which they pursued it — assumed an aura of inevitability. But the approach came with some trade-offs, largely unnoticed at the time. Important questions that resisted the powerful, phenomenological methods tended to get eclipsed. Anything that smacked of 'interpretation', or worse, 'philosophy', began to carry a taint for many scientists who had come through the wartime projects. Conceptual scrutiny of foundations struck many as a luxury. The wartime style was reinforced in the United States by exponentially rising university enrolments after the war. The new classroom realities left little space for informal discussion of philosophy or foundations. The Rad Lab rallying cry of "Get the numbers out" shaded into "Shut up and calculate!"[10]

By the mid-1960s, three-quarters of each year's crop of physics PhD graduates in the United States specialized in either nuclear physics or solid-state physics: two important and interesting areas, to be sure, but also those most readily funded by defence agencies (even for unclassified, basic research). They were also areas in which most physicists came to agree that a pragmatic style could yield the greatest success. During this period, for example, physicists first understood the nuclear force that causes radioactivity, and conquered strange phenomena such as superconductivity — both Nobel-prizewinning achievements.

Openly philosophical areas of physics, the intellectual roots of which stretched back before the war, became increasingly marginalized, such as grand questions about the birth and fate of the Universe, the thin border between order and disorder in chaotic systems, or the subtle foundations of quantum theory. Sometimes these were denigrated as not even being 'real physics' by influential physicists in the United States, although research in these areas advanced in other parts of the world[10].

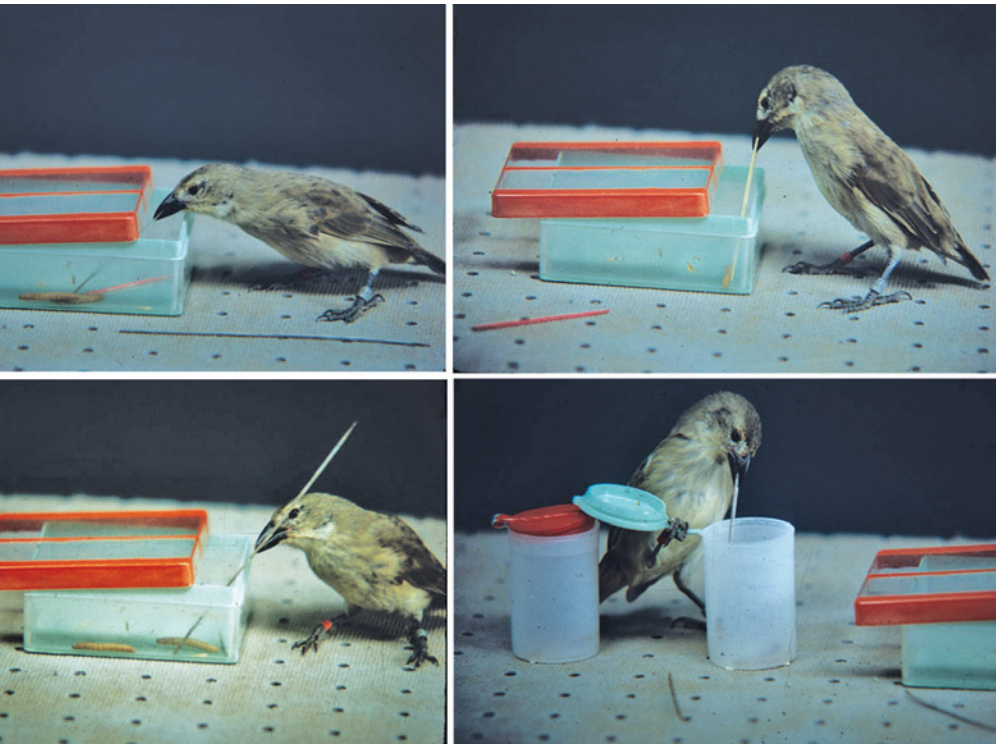
A quarter of a century after the end of the Second World War, cracks in the system began to show. The escalation of fighting in Vietnam made many people question the dominant place of military funding on university campuses, and difficult economic conditions further drove a rapid reversal of fortunes in the sciences, and in physics in particular. Job opportunities for those with science PhDs fell sharply, and university enrolments quickly followed suit, none more drastically than in physics.

The organization, funding and basic approach to research that had come to seem normal — even inevitable — after the war were no longer taken for granted. Complementary styles of research began to creep back in, and growing numbers of physicists turned to topics that had seemed beyond the pale just a few years earlier, such as cosmology, chaos theory and quantum entanglement[10].

Radar philosophy and the Los Alamos man did not disappear from view. To this day, most basic research in the United States depends on federal funding, and many of the great successes of the postwar generation — such as the standard model of particle physics — remain mainstays of research and teaching. But that legacy now sits beside more recent breakthroughs born of the era that reclaimed more openly speculative and philosophical approaches to the deep mysteries of nature. ■

**David Kaiser** *is professor of the history of science and department head for the Program in Science, Technology, and Society at the Massachusetts Institute of Technology, Cambridge, Massachusetts.*
*e-mail: dikaiser@mit.edu*

1. Guerlac, H. *Radar in World War II* (American Institute of Physics, 1987).
2. Hewlett, R. G. & Anderson, O. E. *A History of the United States Atomic Energy Commission: Vol 1 The New World* (Pennsylvania State Univ. Press, 1962).
3. Roberts, A. *Phys. Today* **1,** 17–21 (1948).
4. Forman, P. *Hist. Stud. Phys. Biol. Sci.* **18,** 149–229 (1987).
5. Galison, P. *Image and Logic: A Material Culture of Microphysics* (Univ. Chicago Press, 1997).
6. Westwick, P. *The National Labs: Science in an American System, 1947–1974* (Harvard Univ. Press, 2003).
7. Leslie, S. W. *The Cold War and American Science* (Columbia Univ. Press, 1993).
8. Munns, D. *A Single Sky: How an International Community Forged the Science of Radio Astronomy* (MIT Press, 2013).
9. Schweber, S. *QED and the Men Who Made It* (Princeton Univ. Press, 1994).
10. Kaiser, D. *How the Hippies Saved Physics: Science, Counterculture, and the Quantum Revival* (W. W. Norton, 2011).



CALIFORNIA ACADEMY OF SCIENCES

Studies of birds such as the woodpecker finch (*Camarhynchus pallidus*) have aided a range of fields.

ORNITHOLOGY

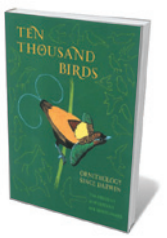
# Under their wing

**Ben Sheldon** relishes a study of the broad-ranging impact of ornithology on modern biology.

The study of wild birds has had a disproportionate impact on the birth and evolution of several branches of science. Thus argue Tim Birkhead, Jo Wimpenny and Bob Montgomerie in *Ten Thousand Birds*.

In chronicling the development of ornithology over the past 150 years, the authors face the challenge of encapsulating a broad and diffuse field. Just how broad can be seen in the questions it tackles today, which span global-scale macroecology of all birds, detailed individual-level behavioural variation, the physiology of migration and the genomics of speciation. Birkhead and his colleagues — all behavioural ecologists — eschew both the obvious chronological approach and the option of presenting potted scientific biographies of 'great ornithologists'. Instead they pick, from the 380,000 papers on birds put out since Charles Darwin published *On the Origin of Species*, 11 areas in which avian research has illuminated broader questions in biology.

➲ NATURE.COM
For more on ornithology, see:
go.nature.com/nbvras




**Ten Thousand Birds: Ornithology Since Darwin**
TIM BIRKHEAD, JO WIMPENNY AND BOB MONTGOMERIE
*Princeton University Press: 2014.*

Birkhead, Wimpenny and Montgomerie hope to stimulate debate over which areas ornithology has contributed to most, but the ones they have chosen will surely appear on most lists. They range from palaeontology, speciation and systematics to physiology and comparative anatomy, by way of ethology, behavioural ecology and conservation.

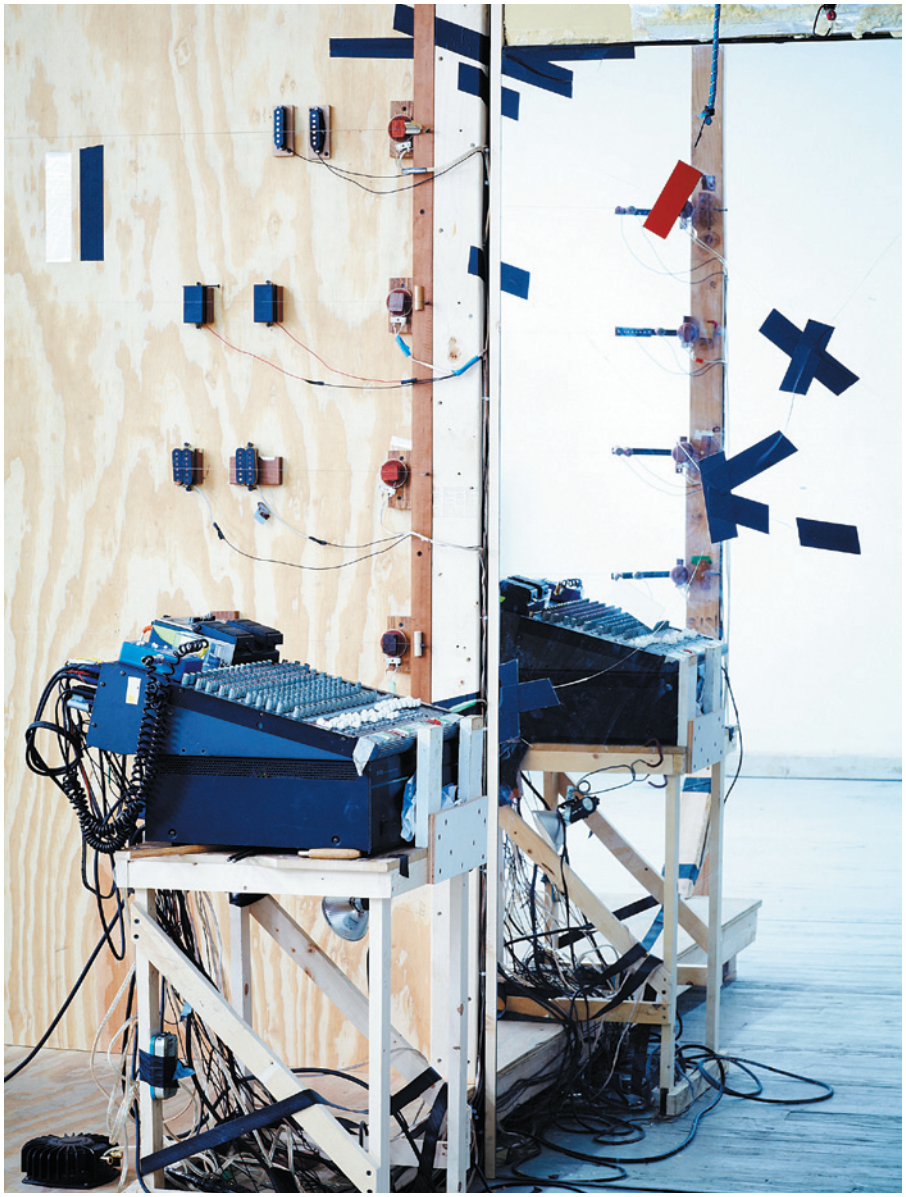
Research on wild birds has been key to understanding population dynamics, thanks to the ease of marking and identifying individuals. Swedish ecologist Malte Andersson's 1980s field experiments on the long-tailed widowbird (*Euplectes progne*) — in which longer tails in males were found to endow increased breeding success — ushered in a new phase of research into sexual selection. Similarly, birds have been inspirational subjects for studies of natural selection ever since US biologist Hermon Bumpus found a link between morphology and individual survival in a flock of sparrows caught in an 1898 Rhode Island snowstorm. Even research on phenomena found only in birds offers general scientific insights: for instance, the elucidation of the mechanism behind the magnetic sense in avian navigation serves as a model for sensory physiology.

This is a serious book that manages to be compulsively readable. The authors are at their most vivid when offering vignettes of individuals or specific events from the modern history of ornithology: pivotal moments of discovery or the presentation of new ideas. Some are familiar. We have British evolutionary biologists David Lack and then Peter and Rosemary Grant successively exploring the evolution of Darwin's finches from the 1940s to the present, and the explosion in discoveries of fossil birds in China since the 1990s. Such cases act as scene-setters in each chapter, and are supplemented by wonderful, less well-known examples. We meet, for instance, the wealthy Hungarian palaeontologist Franz Nopcsa von Felsö-Szilvás (whose theories on early bird flight were influential), who in 1907 crossed Albanian rivers disguised as a shepherd and using an inflated goat's bladder as a flotation device. There is a fine description of ornithologist Charles Sibley striding arrogantly into a 1986 conference bearing a tapestry-sized poster illustrating his revolutionary new avian phylogeny. The book is beautifully illustrated, and also contains charmingly candid autobiographical sketches contributed by more than 20 of today's leading ornithologists.

How did ornithology come to have such a large impact on other areas? The answer seems to lie in the fact that young people with an interest in the natural world are attracted to observing birds, and the ease of studying them. With this in mind, two themes could have been developed further. The first is the importance of birdwatching in the early lives of many scientists who went on to excel in other areas, such as James Watson, co-discoverer of DNA's double helix, and evolutionary biologist Ernst Mayr. The second, which could fill another book, is the huge contribution of non-professionals. From the pigeon fanciers who influenced Darwin to today's army of digitally empowered citizen scientists collecting swathes of distributional and abundance data, amateur ornithologists have built the foundations of the modern science, and enabled its impact in so many other fields. ■

**Ben Sheldon** *is Luc Hoffmann Professor of Field Ornithology and director of the Edward Grey Institute in the Department of Zoology at the University of Oxford, UK.*
*e-mail: ben.sheldon@zoo.ox.ac.uk*

OLIVER JONES



**Part of an instrument built with artist Oliver Jones to create sounds in Composing the Tinnitus Suites.**

# Q&A Daniel Fishkin

# Tinnitus tunesmith

*Sound artist Daniel Fishkin tries to convey the experience of tinnitus. As the latest incarnation of his installation series Composing the Tinnitus Suites opens in Brooklyn, New York, he talks about building a mechanical model of the inner ear.*

**How did your tinnitus begin?**
It started in 2008. I was 22, and it was the evening of my senior recital in music composition at Bard College [in Annandale-on-Hudson, New York]. The concert was not loud, but afterwards I could hear a high-pitched sound. I had my ears tested over and over. Doctors told me to get used to it, but habituation is not always possible. People with tinnitus can experience depression; some have been driven to suicide.

**What you have learned about the condition?**
The mechanics of tinnitus are still poorly understood. Mine is likely to be from noise exposure, but it also occurs as a result of skull fractures and brain tumours, and as a side effect of common drugs. No treatments I was prescribed have helped. I haven't had access to experimental treatments such as [repetitive] transcranial magnetic stimulation. Because there is currently no way to regrow damaged sensory hair cells in the human cochlea, there is no cure for me.

ILLUSTRATION BY NICK HIGGINS BASED ON PHOTO BY SAMUEL LANG BUDI

**Composing the Tinnitus Suites**
DANIEL FISHKIN
*Nothing Space, Brooklyn, New York. On 24, 25 and 31 January and 1, 8 and 9 February.*

**How did this affect your music?**
I tried to keep composing, but it was impossible. So I began to study circuitry and engineering, and built a large sculpture strung with 6-metre-long piano strings wired in a feedback loop that made gorgeous long crescendos. I didn't realize until I showed the device to my girlfriend, an auditory neuroscientist who studies hearing loss, that I had made a giant mechanical model of the inner ear. It became the first installation in my ongoing project, Composing the Tinnitus Suites (see go.nature.com/xhvyu9).

**Tell me about your next installation in this project.**
This month I will create a new musical sculpture at the Nothing Space in Brooklyn. I will line the walls of the gallery with piano wire so that when the audience enters, they will be walking inside the instrument. There will be six performances that will integrate the sustained sounds of my instrument with woodwind and drums.

**Are you trying to reproduce your experience of tinnitus?**
My sculptures generate tones that drift slowly like those in my head, but because my tinnitus is not an acoustic phenomenon — it is a perceived sound rather than actual sound waves — I have found that it is impossible to reproduce the sound. For inspiration, I have tried to enhance the ringing in my ears by listening to loud tones for long periods of time, and consuming substances such as aspirin, caffeine and quinine, which are toxic to the ear.

**How will the project continue?**
The crux of the work is human agency. I want to connect people with tinnitus, give doctors a sense of what their patients are going through, and start a conversation between musicians and scientists. This piece will go on as long my ears ring. Instead of trying to get rid of the ringing, I am now asking: what if, instead of a curse, tinnitus is a kind of superpower? ■

INTERVIEW BY JASCHA HOFFMAN

# Correspondence

## Call for ecosystem modelling data

We call on bioinformaticians, taxonomists and ecologists to collect, store and share new types of data for creating general ecosystem models (GEMs) that will eventually be used for predictive modelling of the biosphere (see G. Mace *Nature* **503,** 191–192; 2013). Funding bodies also need to recognize the importance of this work.

Crucial requirements include: a database of functional traits for different species, which would allow modelling to take advantage of existing information associated with species-based databases; biotic data that indicate how organisms interact through space and time; and census data that quantify the number or weight of organisms in an ecosystem organized by functional traits.

These comprehensive functional data will speed up the development of GEMs by enabling uncertainties to be reduced and predictions to be evaluated (see D. Purves *et al. Nature* **493,** 295–297; 2013).

We therefore appeal to taxonomists, who are usually concerned with morphological traits, to collect information on life histories and behavioural traits, especially with respect to species interactions. Ecologists can also contribute to these enhanced species descriptions by treating individuals as part of assemblages of the same species and as communities of different species, thereby providing valuable collective information.

The technological capacity to store and share trait information is being developed by the *Encyclopedia of Life* (eol.org/traitbank). This can be scaled up to provide more-comprehensive data, such as those to describe interactions between organisms.

**Mike Harfoot** *United Nations Environment Programme World Conservation Monitoring Centre, and Microsoft Research, Cambridge, UK.*
*mike.harfoot@unep-wcmc.org.*
**Dave Roberts** *Natural History Museum, London, UK.*
*scratchpads@nhm.ac.uk*

## Halt self-citation in impact measures

We can improve the gender differences in science publishing and research (see V. Larivière *et al. Nature* **504,** 211–213; 2013) by making measurements of scientific output and impact fairer.

For example, time spent on active research should be incorporated into assessments of research productivity. This would provide a fairer comparison for researchers who take parental leave or who have other caring duties or high teaching loads, and would reduce the pressure on those scientists.

It would also be useful to halt the inclusion of author self-citations in measures of research impact, because self-citation is a male-biased practice (E. Z. Cameron *et al. Trends Ecol. Evol.* **28,** 7–8; 2013). After all, genuine impact hinges on independent citation.

**Elissa Z. Cameron, Amy M. Edwards** *University of Tasmania, Hobart, Australia.*
*elissa.cameron@utas.edu.au*
**Angela M. White** *US Department of Agriculture Forest Service, Davis, California, USA.*

## Himalayas already have hazard network

Maharaj Pandit calls for the protection of the Himalayas through an international network to monitor environmental risks, develop early-warning systems to detect hazards and provide a better understanding of Himalayan geology and ecology (*Nature* **501,** 283; 2013). Such a network is in fact already in place, but it needs more international support if it is to be properly effective.

The International Centre for Integrated Mountain Development (ICIMOD) in Kathmandu, Nepal, was founded 30 years ago by the eight countries of the Hindu Kush–Himalayan region. ICIMOD's expertise is now internationally recognized (S. Sarkar *Himal. J. Sci.* **4,** 7–8; 2007).

ICIMOD forms a centre for intergovernmental knowledge and learning, as well as for regional research and development. It works for sustainable economic and environmental development of the Himalayan ecosystems, by monitoring risks from glacial lakes and providing early warning of hazards such as forest fires and flooding in many of its member countries.

Furthermore, the formation of a Himalayan University Consortium is strengthening collaboration between universities in the region. ICIMOD has been collaborating with governments, academics and non-governmental and community-based organizations from several Hindu Kush–Himalayan countries on conservation programmes to identify the most vulnerable transboundary landscape areas, which are also of global importance.

**Yadav Uprety, Ram P. Chaudhary** *Research Centre for Applied Science and Technology, Tribhuvan University, Kathmandu, Nepal.*
*yuprety@yahoo.com*
**Nakul Chettri** *ICIMOD, Kathmandu, Nepal.*

## Avoid pitfalls of consensus methods

We would like to clarify points raised in William Sutherland's criticism of the treatment of pollinators in the UK National Ecosystem Assessment (*Nature* **503,** 167; 2013).

The estimated economic costs of pollinator decline are only as robust as the natural science on which they rest, as Sutherland indicates. If we could predict with certainty the effects of changes in pollinator populations on agricultural production, then evaluating them would be trivial.

It was because of uncertainty in the underlying population ecology that we omitted estimates of pollination services from our economic analysis of the impacts of land-use change in our report, which was extensively peer-reviewed (see also I. J. Bateman *et al. Science* **341,** 45–50; 2013).

The Delphi technique — a consensus method that Sutherland mentions for synthesizing research findings — can be helpful in some situations, but should be applied with caution to environmental valuation. The rapid expansion of empirical literature in this field means that conventional beliefs can rapidly become group-think norms, with dangerous consequences.

For example, we rejected the popular consensus in favour of using survey techniques as a way of valuing biodiversity, choosing instead to estimate the costs of ensuring species conservation. We stand by our approach, which we believe conforms with Sutherland's appeal for quality over quantity.

**Ian Bateman, Matthew Agarwala, Tomáš Baďura** *University of East Anglia, Norwich, UK.*
*i.bateman@uea.ac.uk*

**CONTRIBUTIONS**
Correspondence may be sent to **correspondence@nature.com** after consulting the guidelines at **http://go.nature.com/cmchno**. Alternatively, readers may comment online: **www.nature.com/nature**.

**FORUM:** Developmental biology

# Tethered wings

**Wnt signalling molecules are thought to direct the development of an organism by spreading through tissues. But flies grow with almost normal appendages even when their main Wnt protein cannot move. Two scientists discuss the implications of this finding for our understanding of development.** SEE ARTICLE P.180

**THE PAPER IN BRIEF**

- The *Drosophila* (fruitfly) protein Wingless (Wg) is the prototype member of the Wnt family of proteins, which regulate tissue patterning and growth during development.
- Wg is thought to act as a morphogen — a protein that forms concentration gradients as it spreads from its site of synthesis and that regulates gene expression as a function of its concentration.
- On page 180 of this issue[1], Alexandre *et al.* describe wing formation in flies expressing a form of Wg that is tethered to the cell membrane, in place of the secreted protein.
- They observe normal wing morphology, although development is delayed and the final wings are smaller than those of normal flies.

## Non-essential spread

GINÉS MORATA

Morphogen regulation of target genes depends on the physical distance from the morphogen-secreting cell population, such that the levels of this molecule provide a genetic reading of position, a key issue in morphogenesis. The best examples of morphogens come from *Drosophila*: the secreted molecules Hedgehog, Decapentaplegic (Dpp) and Wingless (Wg) have been identified as morphogens[2], and for Dpp and Wg there is compelling evidence that they act at long range[3,4]. It follows from the very definition of a morphogen that the spread of the molecule is an essential component of its function. But Alexandre and colleagues' results suggest that this idea needs to be reconsidered.

A clear demonstration of long-range action by Wg came from the finding[4] that the protein activates target genes, including *vestigial* (*vg*) and *Distal-less* (*Dll*), in cells distant from Wg-secreting cells. By contrast, a Wg variant protein (Nrt–Wg) that is functional but anchored to the cell membrane, through the addition of part of the transmembrane protein Neurotactin, was found to activate these target genes only in neighbouring cells. These original experiments were performed by artificial overexpression of Nrt–Wg; Alexandre *et al.* have now used sophisticated genome-editing technology to generate flies in which the *wg* gene encodes the Nrt–Wg protein. The edited gene contains all the normal regulatory sequences and is therefore expressed normally, and the method seems to work with high efficiency, opening up the possibility of performing similar manipulations in other *Drosophila* genes of interest[5].

Considering the many functions of Wg during embryogenesis and during larval and adult life, and the essential role assigned to the protein's spread, any expert would have confidently predicted that a fly with only tethered Wg would not develop. But Alexandre and colleagues' flies survive and are normal in appearance, although they grow more slowly than normal flies and their wings are smaller (Fig. 1). The authors examined the situation only in the wing disc, but the fact that the flies survive indicates that other Wg functions are more or less normal. The implication of their findings is that, at least for *Drosophila*, the long-range diffusion of Wg may be of minor significance — bringing into question the functional value of its role as a morphogen.

How should these results be interpreted in light of the compelling evidence for long-range Wg action? Alexandre *et al.* confirm that Nrt–Wg can induce *Dll* and *vg* gene expression only in adjoining cells, so the almost-normal expression of these target genes in their mutant flies is hard to explain. The authors propose a 'cellular memory' model, in which cells initially expressing target genes continue to express them even when they no longer receive the expression-inducing signal. That implies that *Dll* and *vg* expression is perpetuated through cell divisions, but this is not supported by published evidence[6]. An alternative explanation is that, although the Nrt–Wg protein is considered to be functionally equivalent to Wg (except for its diffusion), there might be undetected differences in its expression levels or stability.

Despite the need for clarifying some aspects of these findings, the survival of flies that have only membrane-tethered Wg is telling, and the authors' results call for a reassessment of how we think about Wg function, and perhaps about that of other morphogens. Wg and Dpp are evolutionarily conserved in all animals, so their mode of function is likely to be conserved as well. The two proteins have acted as model morphogens, and understanding how they work is of major importance to biology. Wnt signalling is also of biomedical interest, because its misregulation is associated with human cancers and other diseases[7]. Although there is no question that these molecules have a crucial role in development and disease, re-examining how they work might change our picture of these processes.

**Ginés Morata** *is in the Centro de Biología Molecular, CSIC-UAM, Universidad Autónoma de Madrid, Madrid 28049, Spain.*
*e-mail: gmorata@cbm.uam.es*

## Long-range thinking

GARY STRUHL

There is compelling evidence[3,6,8] that Wg can, and normally does, act over many cell diameters to control gene expression and growth of the *Drosophila* wing. So the remarkable discovery that a membrane-tethered form of Wg can substitute for the normal protein poses the question: must morphogens move to organize development? When considering this challenge to how we think of morphogens, the devil is in the details.

The main phase of *Drosophila* wing development begins with the induction of Wg expression in all cells of the nascent wing and lasts for around two days, during which time the wing increases by about 50 times in size and cell number. On the first day, Wg is broadly

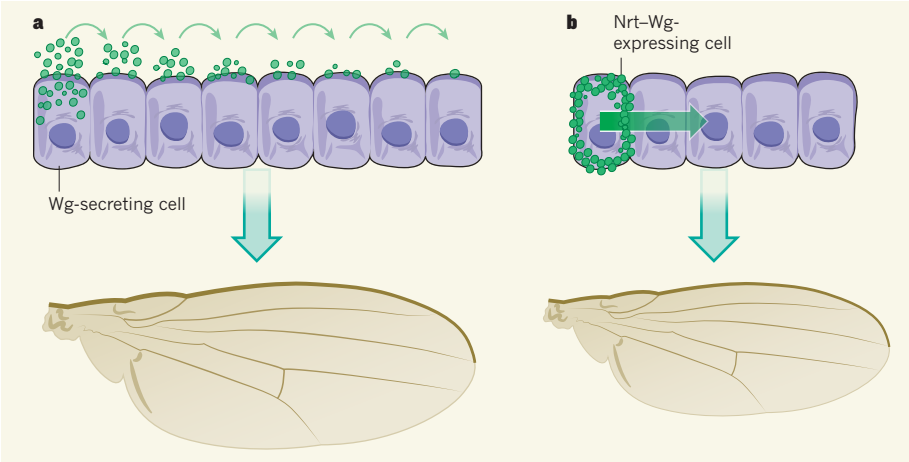



**Figure 1 | Wing development and Wingless spread.** **a**, The Wingless protein (Wg) is thought to regulate the development of *Drosophila* wings by diffusing from Wg-secreting cells, thereby activating Wg target genes in distant cells as the wing grows. **b**, Alexandre *et al.*[1] show that, in flies expressing the Wg variant Nrt–Wg, which is tethered to the cell membrane and cannot diffuse, wings with normal morphology develop, although development is delayed and the wings are smaller than normal. Previous work[3,6] has shown that Nrt–Wg can activate Wg target genes in nearby cells but not in distant cells (green arrow), so it remains unclear how the long-distance Wg signalling thought to be required for wing development is exerted in these flies.

expressed, but its expression fades progressively in the more dorsally and ventrally positioned cells, generating Wg gradients. These gradients might well suffice to initially direct normal gene expression and growth without requiring Wg to spread, and Alexandre and colleagues' results support this idea.

During the second day, Wg production becomes restricted to a central stripe of cells, but the protein continues to control gene expression and growth in cells up to 15–20 cell diameters away. The conventional view is that this is because they continue to receive Wg secreted by the central cells. Alexandre and colleagues propose instead that, once prospective wing cells receive Wg, they acquire a long-term cellular memory of Wg exposure that controls the behaviour of their descendants thereafter. According to this model, the descendants of cells exposed to Nrt–Wg should grow and express Wg target genes even when they are many cell diameters away from Nrt–Wg-expressing cells. But previous work[3,6] shows that this is not so; rather, only those cells that remain close to Nrt–Wg-expressing cells continue to express Wg target genes and grow.

Without invoking a memory model, how might tethered Wg mimic the long-range action of the normal protein? One possible answer comes from the observation[5] that Nrt–Wg accumulates in secreting cells to several times higher levels than normal Wg, indicating either that it is significantly more stable and/or that it provides a more potent signal to neighbouring cells because it is not attenuated by release and dispersion. Accordingly, Nrt–Wg that is expressed during the first day might persist and function adventitiously during the second day, providing a signal that would otherwise require spread of the protein. Another possibility is that Wg moves to some extent via cellular projections[9,10]; membrane-tethered Wg might also be able to do this, allowing it to influence cells at least a few cell diameters away. A third option is that the downstream effects of Wg signalling (including the function of proteins encoded by target genes) might persist for several hours, even after cells cease to receive Wg. All of these factors, some normal and others artefacts of the Nrt–Wg system, would extend the range over which Nrt–Wg can influence cell behaviour during the second day of wing development.

Notably, Alexandre *et al.* find that the tethered protein fails to sustain normal patterns of gene expression or support normal growth after its expression becomes restricted to the central stripe. This results in delayed wing development, and wings that never reach full size even when they have an extra day or longer to catch up.

Thus, the new results do not falsify the interpretation of Wg as a classic morphogen in the *Drosophila* wing. Instead, they highlight that Wg acts at short range during early wing development but must act at long range at later times, as its production becomes restricted and the population of cells requiring Wg input expands. The *Drosophila* wing will continue to serve as a model for understanding how morphogens act to organize the development of larger tissues (such as butterfly wings and vertebrate limbs), and further studies using the methods introduced by Alexandre *et al.* will contribute to this understanding. ■

**Gary Struhl** *is in the Department of Genetics and Development, Columbia University, New York, New York 10032, USA.*
*e-mail: gs20@columbia.edu*

1. Alexandre, C., Baena-Lopez, A. & Vincent, J.-P. *Nature* **505,** 180–185 (2014).
2. Lawrence, P. A. & Struhl, G. *Cell* **85,** 951–961 (1996).
3. Zecca, M., Basler, K. & Struhl, G. *Cell* **87,** 833–844 (1996).
4. Nellen, D., Burke, R., Struhl, G. & Basler, K. *Cell* **85,** 357–368 (1996).
5. Baena-Lopez, L. A., Alexandre, C., Mitchell, A., Pasakarnis, L. & Vincent, J.-P. *Development* **140,** 4818–4825 (2013).
6. Zecca, M. & Struhl, G. *Development* **134,** 3001–3010 (2007).
7. MacDonald, B. T., Tamai, K. & He, X. *Dev. Cell* **17,** 9–26 (2009).
8. Neumann, C. J. & Cohen, S. M. *Development* **124,** 871–880 (1997).
9. Locke, M. & Huie, P. *Tissue Cell* **13,** 787–803 (1981).
10. Gradilla, A.-C. & Guerrero, I. *Curr. Opin. Genet. Dev.* **23,** 363–373 (2013).

**This article was published online on 25 December 2013.**

ELECTROCHEMISTRY

# Metal-free energy storage

**In flow batteries, energy is produced by passing solutions of 'electroactive' materials — often, metal salts — through an electrochemical cell. A non-metallic electroactive material opens the way to large-scale energy storage.** **SEE LETTER P.195**

GRIGORII L. SOLOVEICHIK

The adoption of intermittently available renewable energy sources, such as solar energy and wind power, to more than 20% of total energy capacity will require electric-energy storage systems to be deployed[1]. For grid-scale applications and remote generation sites, cheap and flexible storage systems are needed, but presently the options are either limited to a specific geographic location (such as pumping water from a reservoir to an elevated level as a source of potential energy) or expensive (for example, conventional batteries, flywheels and superconductive electromagnetic storage)[2]. On page 195 of this issue, Huskinson *et al.*[3] report a major advance in the development of economical energy storage: a 'flow' battery that uses only water-soluble, non-metallic materials as the electrode components.


## 50 Years Ago

The web-building of spiders provides a useful test of the activity of pharmacological agents. Particular substances affect web-building in particular ways … Unfortunately, the traditional procedure of producing discrete lesions in the central nervous system to observe behavioural consequences is particularly difficult in most arthropods because of the small size and rigid cuticle of the animals … [H]ere, the intense and directional radiation of a laser beam was used to produce lesions … Transient behavioural abnormalities occurred in 4 of the exposed spiders, apparently a consequence of damage to tissue with recovery capacities. Permanent disturbance occurred only after lesions in the cephalothorax of spiders, as the presumed consequence of damage to tissue incapable of regeneration, such as nerve tissue. Histological analysis appears to confirm central nervous system lesions … [A] means seems to be at hand for objectively relating spider (and other arthropod) behaviour to the results of histological analysis.

**From *Nature* 11 January 1964**

## 100 Years Ago

'Lucretius or Kapteyn?' — *Nonne vides etiam diversis nubila ventis diversas ire in partis inferna supernis? Qui minus illa queant per magnos aetheris orbis aestibus inter se diversis sidera ferri?*
*De Rerum Naturâ*, v., 646–9.
See you not too that clouds from contrary winds pass in contrary directions, the upper in a way contrary to the lower? Why may not yon stars just as well be borne on through their great orbits in ether by currents contrary one to the other? *Munro's Translation* E.J.M.

**From *Nature* 8 January 1914**

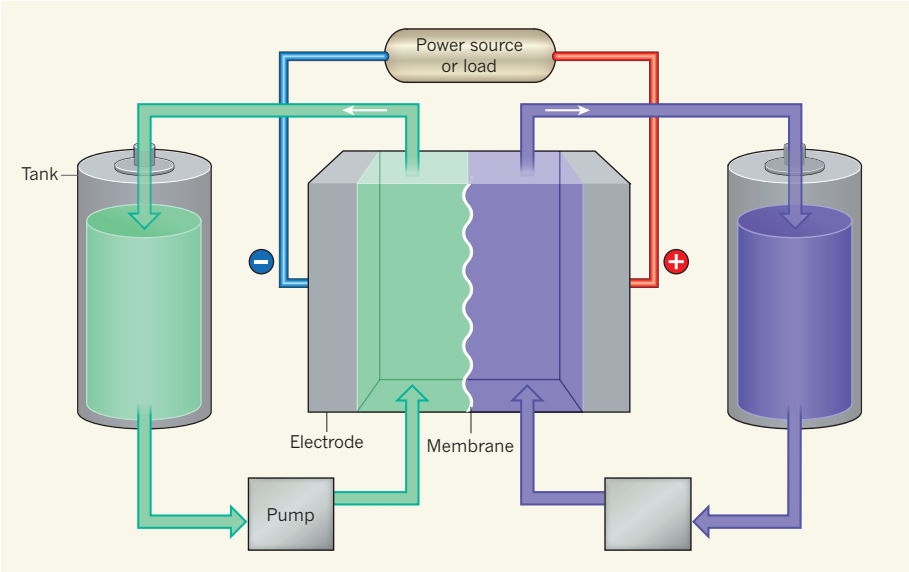



**Figure 1 | Schematic of a flow battery.** In this set-up, two solutions of electroactive materials (green and purple) are stored in external tanks and pumped to flow-through electrodes in an electrochemical cell. The materials undergo reactions at the electrodes, generating electricity when a load is connected. A membrane between the electrodes prevents the solutions from mixing, but allows the transport of charge-carrying ions, thus allowing electrical neutrality in the system to be maintained. In charging mode, the power source generates a potential difference across the cell. Huskinson *et al.*[3] report that organic compounds known as quinones can be used as the electroactive material at the anode of a flow battery. (Schematic adapted from ref. 8.)

These materials could lower the cost of flow batteries, while increasing the energy density.

Flow batteries require two soluble electroactive components — compounds that take part in an electrochemical reaction at an electrode. These components are separated by an ion-conducting membrane in an electrochemical cell, in which chemical energy is converted to electricity (and vice versa). In contrast to the stationary electroactive materials of conventional batteries, the electroactive components in flow batteries are pumped through the cell in a flow of liquid, and are stored outside the cell in separate tanks (Fig. 1). This design allows for individual optimization of the amount of energy stored (which is controlled by the size of the storage tanks) and the power generated (which depends on the size of the electrochemical cell or stack of cells).

Because the electroactive materials in flow batteries are stored separately, the possibility that they will react violently with each other is almost completely eliminated, making these devices much safer than conventional batteries. They also have more flexible layouts and are potentially cheaper. Unfortunately, the choice of electroactive materials for flow batteries is limited to a small selection of metal redox systems (with a few exceptions for cathode materials), and by the low solubility of these metal salts — typically, in water. The solubility problem prevents high energy densities from being achieved[4].

Huskinson *et al.* overcome the solubility problem by using as the electroactive components soluble, organic, redox-active materials known as quinones in place of metals. Water-insoluble quinones were proposed[5] as electrode materials in 1972, but the use of this class of compound as the energy-storage components of a flow battery is new. The authors found that the chemical reduction of their quinones to form hydroquinones in water at an electrode is very fast, which is a prerequisite for high-power battery discharge.

Redox potentials and the solubilities of metal complexes can be tuned by modifying the ligands bound to the metal atoms. With quinones, these properties can be modified by changing the chemical groups attached to the aromatic rings of the molecules. This offers much wider scope for modification than is possible for metal systems, because the chemical groups are closer to the redox centre than ligands are to metals, and so their effect is more pronounced. In addition, having negatively charged electroactive species — quinones to which negatively charged groups are attached, such as those used by Huskinson *et al.* — should help to reduce one of the major practical problems associated with flow batteries, namely the crossover of these materials through the negatively charged ion-conducting membrane. However, this approach has a downside, because any improvements in electrochemical properties and/or solubility will be associated with an increase in the molecular weight of the electroactive species, and will therefore reduce the energy density of the battery.

Huskinson and co-workers coupled the liquid quinone/hydroquinone system at the anode of their battery with a bromine/bromide

system at the cathode. This cathode system has previously been used in a zinc–bromine flow battery[6] and in a hydrogen–bromine regenerative fuel cell[7] (a variant of a flow battery). The bromine/bromide cathode provides good energy density at a reasonable cost, although it is corrosive and environmentally unfriendly. When the authors tested a small version (2 square centimetres) of their flow battery, they found that it gave a respectable power density (600 milliwatts per square centimetre) and good current efficiency (the efficiency with which charge is transferred to allow a targeted electrochemical reaction to occur).

The new findings open the way to inexpensive energy storage, but there is a long way to go to develop a practically useful flow battery. In particular, several issues must be addressed before this chemistry can be used in grid-scale energy storage. The authors studied only quinone reduction, so the reverse reaction — the oxidation of hydroquinones — should also be investigated. If the reverse reaction is as fast as quinone reduction, then quinones could potentially be used in high-power devices. The effect of the electroactive-species concentration, and of impurities in the quinones, on the cell's performance and ability to be used through many charge–discharge cycles must be evaluated. If high-purity quinones are needed, it could noticeably increase the cost.

Bromine crossover through the membrane should also be considered seriously. Even if bromine does not react with compounds in the anode system, such crossover will reduce battery capacity and energy efficiency (the ratio of electrical-energy output to input), which should be measured as a function of cycle number. Scaling up from a small single cell to an industrial-sized, multi-cell stack may be challenging, and integrating the various components of a large-scale device into a working battery might also be difficult. For stationary energy storage, a long life (more than 10,000 cycles) is key to keeping costs down, so the number of cycles demonstrated in the paper (15) is far from that needed.

Nevertheless, Huskinson and colleagues' results are promising, and may serve as the basis for a new flow-battery technology. If long-term capacity and energy-efficiency retention can be demonstrated, and if practically useful batteries can indeed be prepared cheaply, then this technology will be suitable for a wide array of energy-storage applications. ■

**Grigorii L. Soloveichik** *is at General Electric Global Research, Niskayuna, New York 12309, USA, and at the Energy Frontier Research Center for Electrocatalysis, Transport Phenomena and Materials, General Electric Global Research.*
*e-mail: soloveichik@ge.com*

1. Denholm, P., Ela, E., Kirby, B. & Milligan, M. *The Role of Energy Storage with Renewable Electricity Generation* (Natl Renewable Energy Lab, 2010).
2. Soloveichik, G. L. *Annu. Rev. Chem. Biomol. Eng.* **2,** 503–527 (2011).
3. Huskinson, B. *et al. Nature* **505,** 195–198 (2014).
4. Weber, A. Z. *et al. J. Appl. Electrochem.* **41,** 1137–1164 (2011).
5. Alt, H., Binder, H., Klempert, G., Köhling, A. & Sandstede, G. *J. Appl. Electrochem.* **2,** 193–200 (1972).
6. Leung, P. *et al. RSC Adv.* **2,** 10125–10156 (2012).
7. Cho, K. T. *J. Electrochem. Soc.* **159,** A1806–A1815 (2012).
8. Wang, W. *et al. Adv. Funct. Mater.* **23,** 970–986 (2013).

THERAPEUTICS

# Detective work on drug dosage

**Patients differ in their requirement for, and response to, various drug doses. A general platform that allows continuous monitoring of drug levels in the blood of rats may open the door to patient-specific dosing.**

RICHARD M. CROOKS

Choosing the right drug dose for a particular patient is more of an art than a science. For example, the dosage of most drugs is simply based on patients' age: "Adults and children 12 years and over, take 2 tablespoons every 6 hours," for example. In reality, there is patient-to-patient variability in drug metabolism and excretion, highlighting the need for accurate and patient-specific approaches to monitor drug concentration after administration. Taking a big step towards this goal, Ferguson *et al.*[1] describe, in a paper published in *Science Translational Medicine*, a detection system that allows real-time tracking of drugs in the blood.

For a few relatively toxic drugs, several factors, including gender, body mass or body surface area, are taken into account to better estimate the effective dose. For drugs with a particularly narrow window between the minimum effective and toxic doses, clinicians often opt for 'peaks and troughs' measurements. In this approach, blood is drawn half an hour after a drug dose is given, when the drug's concentration in the blood is likely to be highest; a second sample is then drawn immediately before the next dose is due. From these two isolated data points the drug's pharmacokinetics (the rate at which its concentration in the blood falls) is inferred, and from that the optimal dosing regimen for that specific patient is determined.

Even the most advanced methods used to estimate an appropriate drug dose are rather crude and imprecise. Existing measurements of pharmacokinetics typically involve drawing blood and sending it to a central lab for analysis. The ability to monitor drug levels in the blood continuously in real time in the clinic would vastly improve the precision of such measurements. This, in turn, would greatly improve the ability to tailor

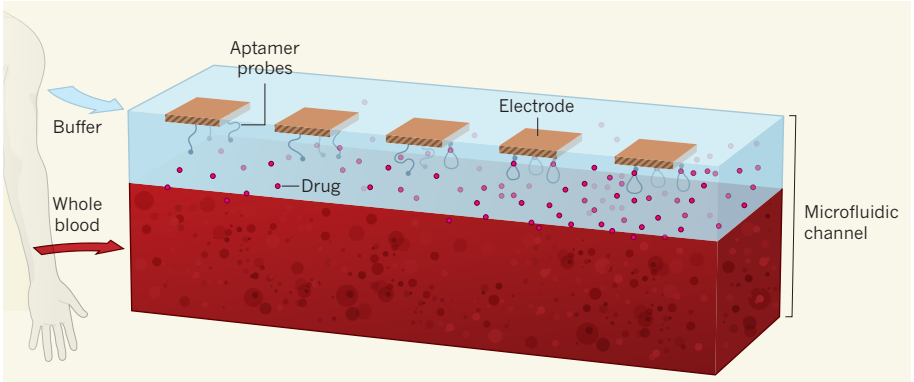



**Figure 1 | MEDIC in action.** Ferguson *et al.*[1] describe a detector they call MEDIC, which could potentially be used as follows. A patient's blood enters the multilayered device, where it flows along a microfluidic channel and encounters, but does not mix with, a separate stream of saline buffer. The buffer layer acts as a continuous flow filter: small-molecule drugs rapidly diffuse to the surface, whereas more slowly diffusing large proteins and blood cells cannot reach it. Aptamer probes, which are immobilized on multiple electrodes along the top of the channel, act as molecular switches: on binding to a specific drug, they change shape and induce an electrical current. The current is directly related to the concentration of the target drug, allowing accurate quantification. (Figure adapted from ref. 1.)

drug dosing to a patient's specific needs.

Measuring drug levels continuously requires a technology that is reversible, so that the sensor's response rises and falls in concert with fluctuating drug concentrations. Moreover, the technology must be continuous, of course, and so should not rely on wash steps or other batch processes. Finally, it must be sufficiently selective to be used on whole blood. Unfortunately, although conventional analytical methods, including chromatography, spectroscopy and immunochemistry, often have one or more of these attributes, no general approach has achieved all these goals simultaneously. There are ways to measure a few specific molecules in the body in real time (for instance, blood glucose levels in patients with diabetes), but these are single-analyte sensors that are not easily generalizable to the detection of other molecules.

Ferguson *et al.* describe a sensor that cleverly links the above three technologies; they call it microfluidic electrochemical detector for *in vivo* continuous monitoring (MEDIC).

The sensing technology underlying this platform is a reagent-free electrochemical device[2,3] that uses the binding-induced folding of aptamers[4] (artificially selected nucleic acids that bind specific molecular targets) to signal the presence of a given analyte. This reagent-free, wash-free, sensing architecture has previously been shown[5] by some of the same authors to support continuous measurements in flowing, undiluted blood serum. The approach fails, however, when the sensor is challenged with whole blood, owing to the nonspecific adsorption of molecules onto the electrode surface, which progressively deactivates the associated aptamers — thereby leading to baseline drift in the output signal.

To eliminate this drift, Ferguson and colleagues took a two-pronged approach. The first was to place the sensors in a microfluidic device that insulates them with a micrometres-thick stream of buffer. Blood continuously collected from the subject (by a cannula) is drawn into the device, where it forms a laminar flow over this buffer (Fig. 1). Because the drug molecules are small, they quickly diffuse through the buffer layer to reach the sensor surface. The much larger blood cells and other large interfering agents diffuse too slowly to reach the buffer stream, so sensor fouling is essentially eliminated.

The authors' second advance was to interrogate their electrochemical aptamer probes using a method, known as square-wave voltammetry, operating at two discrete frequencies. Specifically, they identified matched frequency pairs at which the output signal drifts in concert while responding very differently to the presence of the target. Taking the difference between these two signals effectively eliminates drift. Combining the two approaches, the authors' device achieves multihour, continuous measurements on whole, undiluted blood with baseline stabilities in the submicromolar range of drug concentration.

The team demonstrated the ability of the MEDIC platform to monitor the chemotherapeutic drug doxorubicin and the antibiotic kanamycin in the blood of anaesthetized rats over the course of several hours. The pharmacokinetics derived correspond to long-established values[6] obtained by laboriously drawing blood samples and then, much later, measuring each by using chromatography. In the present paper, by contrast, the measurements were made in real time, which not only is convenient but also improves their precision.

There are some disadvantages to this platform, however. MEDIC requires continuous blood draws (of just a few hundred microlitres per hour) and a pump to maintain the flow of buffer through the device. It is therefore unsuitable for continuous, real-time monitoring of metabolites or drugs in the blood of a mobile patient going about their daily life. Nevertheless, by enabling convenient, high-precision measurements of pharmacokinetics in the clinic, the technology could fuel further advances in personalized medicine by supporting truly individualized dosing regimens. Indeed, the ability to monitor blood drug concentrations in real time could pave the way to proactive, high-precision dosing in which drug delivery is modulated on the go in response to hour-to-hour changes in a patient's metabolism or health status. Such feedback-controlled drug delivery could, in turn, open the door to therapies in which drugs with previously unduly complex dosing regimens or unacceptably narrow therapeutic indices are administered safely and effectively. ■

**Richard M. Crooks** *is in the Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712-1224, USA.*
*e-mail: crooks@cm.utexas.edu*

1. Ferguson, B. S. *et al. Sci. Transl. Med.* **5,** 213ra165 (2013).
2. Xiao, Y., Lubin, A. A., Heeger, A. J. & Plaxco, K. W. *Angew. Chem. Int. Edn* **44,** 5456–5459 (2005).
3. Lubin, A. A. & Plaxco, K. W. *Acc. Chem. Res.* **43,** 496–505 (2010).
4. Osborne, S. E., Matsumura, I. & Ellington, A. D. *Curr. Opin. Chem. Biol.* **1,** 5–9 (1997).
5. Swenson, J. S. *et al. J. Am. Chem. Soc.* **131,** 4262–4266 (2009).
6. Ueda, Y. *et al. Chem. Pharm. Bull.* **37,** 1639–1641 (1989).

QUANTUM PHYSICS

# An atomic SQUID

**Superconducting quantum circuits are the core technology behind the most sensitive magnetometers. An analogous device has now been implemented using a gas of ultracold atoms, with possible applications for rotation sensing.**

CHARLES A. SACKETT

When a magnetic field needs to be measured with the utmost precision, a superconducting quantum interference device (SQUID) is the instrument of choice[1]. Its exquisite sensitivity derives directly from a macroscopic manifestation of quantum mechanics, making it an archetype of quantum engineering. Reporting in *Physical Review Letters*, Ryu and colleagues[2] demonstrate an analogue of a SQUID using an ultracold gas of neutral atoms known as a Bose–Einstein condensate. Here, the analogue to the magnetic field is a physical rotation, so the atomic device could prove useful for rotation sensing and vehicle navigation. More broadly, it strengthens the correspondence between atomic and solid-state systems. Because atomic systems are better understood and more easily controlled than their solid-state counterparts, atoms might eventually serve as a design platform for complex solid-state quantum devices.

A conventional SQUID is a small ring of superconducting material cut in half by two non-superconducting barriers. Wire leads connected to each side of the device allow a current to pass through it (Fig. 1a). Within each of the superconducting regions, electrons act like a coherent quantum wave. Because the current passing through the SQUID can take either path around the ring, the two corresponding waves can interfere: they can add constructively with the peaks of the waves lined up, or cancel destructively with the peaks of one wave aligned to the troughs of the other. The total current through the ring depends sensitively on the type of interference. For charged particles such as electrons, the way that the waves align is set largely by the magnetic field threading the ring, which makes the SQUID a good magnetometer.

In the atomic analogue demonstrated by Ryu and colleagues, the superconducting electrons are replaced by a Bose–Einstein condensate consisting of a few thousand rubidium atoms at nanokelvin temperature, isolated in an ultra-high-vacuum chamber[3]. Like the electrons, the atoms in a Bose–Einstein condensate act as a wave, allowing similar physics to be probed. Here, the atoms are held in a ring-shaped trap that has two small potential-energy barriers through which the atoms can tunnel (Fig. 1b). The authors created the ring trap using a technique known as a painted potential. For

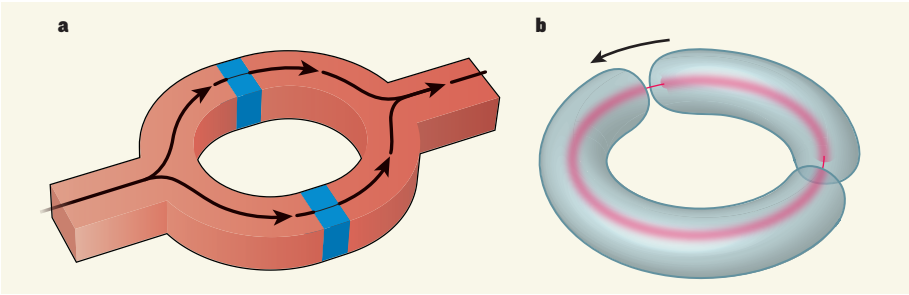



**Figure 1 | Quantum interference devices.** **a**, A conventional superconducting device consists of a ring of superconducting wire split by two non-superconducting barriers (blue). The current (thick black lines) through the loop must tunnel through the barriers (thinner lines). **b**, The atomic version demonstrated by Ryu *et al.*[2] is a Bose–Einstein condensate (red) held in a ring-shaped trap. The ring is broken by two potential-energy barriers. Instead of passing the atoms through the barriers, here one of the barriers is moved around the ring so as to pass through the atoms, as indicated by the arrow. The distribution of the atoms between the two regions reveals the dynamics of atom motion, which corresponds well to the electron currents in the superconductor.

this, a laser beam is rapidly scanned across the trapping region and selectively turned on so as to illuminate only a ring-shaped area. The atoms are attracted to the laser light and confined within the ring. The tunnelling areas are produced by reducing the light intensity at two spots on the ring.

Because the atoms are neutral, the condensate version is not especially sensitive to magnetic fields. However, if the whole apparatus rotates, then the atoms will experience the Coriolis force, which twists the path of any object moving on a rotating platform. For example, on the Earth, the Coriolis force causes the circulating air flow of hurricanes and cyclones. It affects the atomic waves much like a magnetic field affects electrons. By measuring how the atoms move through the ring, even a tiny Coriolis force can be detected, making the system useful for sensing rotation[4].

Ryu and colleagues' work builds on previous demonstrations (see, for example, ref. 4) of atomic systems similar to a SQUID, but for the first time uses the complete geometry of a ring with two barriers. The observed behaviour of the authors' atoms is in good accord with the phenomenological model used to describe superconducting devices. Indeed, for the atomic case, the expected behaviour can be derived nearly from first principles, so the system is on firm theoretical ground.

Rotation sensors are useful for vehicle navigation and other geophysical applications, and the atomic SQUID shows promise for advancing these technologies. A greater impact, however, may derive from the demonstration of how atomic systems can replicate solid-state devices. Although solid-state circuits have the practical benefit of not requiring lasers and vacuum chambers, developing a new device involves painstaking fabrication and characterization work. By contrast, the size and shape of an atom trap can be modified simply by reprogramming the behaviour of the laser beam. The atom system could thus serve as a design tool for complicated circuits, in which the geometry could be developed and optimized before being applied to superconductors. Although ordinary computer simulation can serve a similar purpose, a physical device can easily become too complex for simulation to be practical. Such complexity is common in quantum systems in which particle interactions are important and non-trivial, including high-temperature superconductors and, perhaps one day, quantum computers.

The idea that ultracold atoms could be used to simulate and explain solid-state systems has been a driving force in the atomic-physics community since the first observations of Bose–Einstein condensation[5]. Ryu and colleagues' demonstration that a useful device such as a SQUID can be implemented with atoms is a milestone in this effort. Nonetheless, substantial challenges remain. An immediate issue is the difficulty of using atomic systems to model macroscopic currents: the number of atoms in a condensate is relatively small, so there is no simple way to create a large current. The authors sidestep this problem by measuring a small current flowing through a barrier, rather than a large current passing through the ring as a whole (compare Fig. 1a and b). Although this set-up can be used for rotation measurements, it does not reflect the actual operation of a superconducting SQUID. A larger question is how well the correspondence between atoms and solid-state systems will hold up as the system's complexity grows. Until the systems become too complicated for computer simulation, the utility of the atomic experiments as a design platform for solid-state systems will probably be limited. Meeting the challenges involved will not be easy, but the steady progress in this field exemplified by Ryu and colleagues' achievement is encouraging. ■

**Charles A. Sackett** *is in the Department of Physics, University of Virginia, Charlottesville, Virginia 22904, USA.*
*e-mail: sackett@virginia.edu*

1. Clarke, J. & Braginski, A. I. (eds) *The SQUID Handbook* Vol. 1 (Wiley-VCH, 2004).
2. Ryu, C., Blackburn, P. W., Blinova, A. A. & Boshier, M. G. *Phys. Rev. Lett.* **111,** 205301 (2013).
3. Dalfovo, F., Giorgini, S., Pitaevskii, L. P. & Stringari, S. *Rev. Mod. Phys.* **71,** 463–512 (1999).
4. Wright, K. C., Blakestad, R. B., Lobb, C. J., Phillips, W. D. & Campbell, G. K. *Phys. Rev. Lett.* **110,** 025302 (2013).
5. Bloch, I., Dalibard, J. & Zwerger, W. *Rev. Mod. Phys.* **80,** 885–964 (2008).

**This article was published online on 18 December 2013.**

HIV

# Ringside views

**Two crystal structures reveal that the Vif and Vpx proteins of human and simian immunodeficiency viruses mediate evasion of host defences by reprogramming the cellular protein-degradation machinery.** **See Letters p.229 & p.234**

MICHAEL H. MALIM

The human immune system uses myriad adaptive and innate mechanisms to fight HIV infection and AIDS. Prominent among these is a collection of widely expressed cellular proteins called restriction factors, which can potently suppress viral replication[1]. But human and simian immunodeficiency viruses, including HIV-1, encode several dedicated regulatory proteins that enable them to evade restriction factors, thus ensuring their survival and propagation. Two papers[2,3] in this issue show how the viral Vif and Vpx regulatory proteins bind to key host-cell partners and targets, culminating in the removal of restriction factors from infected cells.

The substrates for Vif are members of the APOBEC3 (A3) protein family, namely A3D, A3F, A3G and A3H; the substrate for Vpx is the SAMHD1 protein. These are all enzymes that interfere with reverse transcription, an essential phase of HIV replication in which the viral RNA genome is copied into DNA. The APOBEC3 proteins are cytidine deaminases that are captured by virus particles as they assemble. The proteins induce destructive hypermutation of nascent viral DNA and suppress its synthesis[1]. SAMHD1 is a deoxynucleoside triphosphate

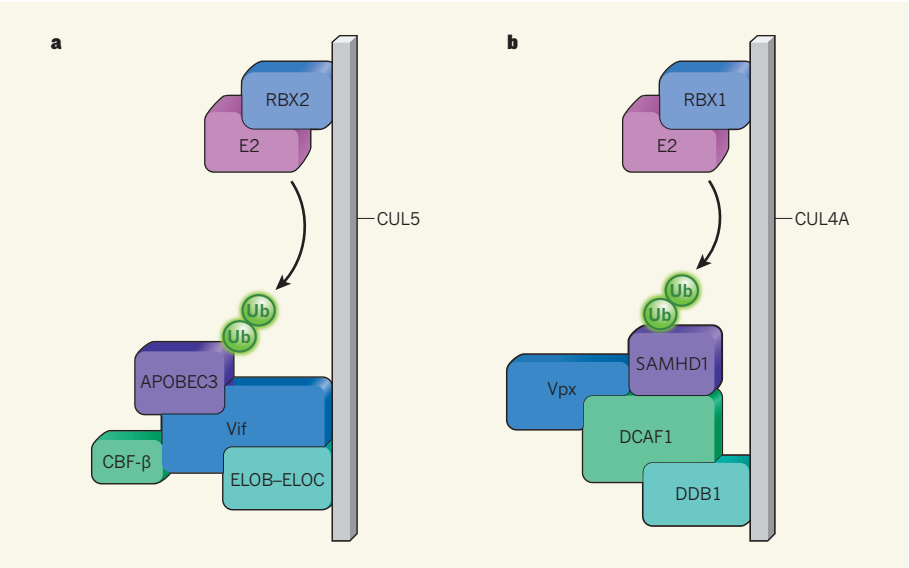



**Figure 1 | Destruction of host-cell antiviral proteins.** **a**, Guo and colleagues' crystal structure[2] shows the HIV-1 Vif protein occupying a central position as a substrate receptor in the CRL5 complex formed between host-cell proteins CBF-β, ELOB–ELOC, CUL5, an E2 enzyme and RBX2. The substrate, an APOBEC3 protein, is recruited by Vif and marked by ubiquitin molecules (Ub), which are transferred from E2 by RBX2. This tags the substrate for degradation by the cell. **b**, Schwefel *et al.*[3] show that the Vpx protein from sooty mangabey simian immunodeficiency virus occupies a comparatively peripheral position in the complex it forms with host-cell proteins. Vpx binds DCAF1 and recruits the substrate SAMHD1; DDB1 has an analogous role to ELOB–ELOC. The proteins interact with CUL4A, RBX1 and an E2 enzyme to form a CRL4A complex, and SAMHD1 is tagged for destruction by ubiquitination.

triphosphohydrolase that reduces the cellular levels of dNTPs (the substrates for DNA synthesis), thereby suppressing reverse transcription, especially in non-proliferating cells such as myeloid cells and resting T cells[4].

Vif and Vpx both work by directly recruiting their target restriction factors to host-cell cullin-RING ubiquitin ligases (CRLs), a diverse family of multicomponent enzymes that add ubiquitin chains to substrates[5], thereby marking them for destruction by the proteasome (a cellular protein-degrading machine). The CRLs are assembled with a cullin (CUL) protein (of which there are six in humans) as the central scaffold. Their catalytic core is built around the carboxy terminus of the CUL and also contains an RBX (or ROC) RING-finger protein and an associated E2 ubiquitin-conjugating enzyme (Fig. 1). The amino-terminal region is devoted to substrate recruitment: typically, the CUL binds a substrate adapter molecule, which in turn connects to a substrate receptor and its bound substrate. Now, Guo *et al.*[2] and Schwefel *et al.*[3] present the contrasting mechanisms used by Vif and Vpx to engage CRLs.

Previous attempts to resolve the structure of the Vif protein, either alone or in association with its CUL5-based CRL (CRL5) or APOBEC3 substrates, had been unsuccessful. A game-changing advance came with the discovery[6,7] that the transcription factor CBF-β is also required for the function of HIV-1 Vif. It rapidly emerged that Vif and CBF-β form a stable heterodimer, and this advance enabled the purification of enough soluble full-length Vif for structural studies.

Guo *et al.* (page 229) describe the first crystal structure of a complex comprising Vif–CBF-β, the ELOB–ELOC heterodimeric substrate adapter and an amino-terminal fragment of CUL5 (Fig. 1a). The structure shows that Vif occupies a crucial nucleating position within this pentameric complex, simultaneously interacting with CBF-β, CUL5 and ELOC, and promoting CRL assembly. By contrast, CBF-β contacts only Vif and seems to serve a chaperone-like function by helping Vif to fold into an active conformation. Interestingly, the contacts that Vif makes with ELOC (through an evolutionarily conserved peptide sequence called the BC box) and CUL5 imitate those made by the cellular protein SOCS2, a CRL substrate receptor involved in the downregulation of growth-hormone signalling. This suggests that the two proteins have adopted a similar mechanism for CRL5 recruitment. A fascinating future step would be to add APOBEC3 proteins to the complex; as discussed by the authors, the amino-acid residues in Vif that are required for engaging A3F or A3G are solvent accessible, and are therefore predicted to be available for direct interactions with these substrates.

Interest in Vpx intensified following the discovery that it provokes the degradation of SAMHD1 during the early stages of virus infection. HIV-2 and diverse simian immunodeficiency viruses (SIVs) encode Vpx or Vpr proteins with this function[8]. Interestingly, HIV-1 does not, raising important questions of whether and how it evades SAMHD1-mediated restriction. Building on earlier structural studies from their group[9], Schwefel *et al.* (page 234) now present the crystal structure of a complex between the Vpx-binding element of SAMHD1, the carboxy-terminal WD40 domain of the CRL4A substrate receptor DCAF1, and the Vpx protein of the SIV that infects sooty mangabeys (Fig. 1b). The structure shows that all three components contact each other, with extensive interactions between Vpx and DCAF1 creating a shared surface to which SAMHD1 binds. Although structures of Vpx-containing complexes with additional CRL components are eagerly anticipated, the authors proceeded to model a CRL–Vpx–SAMHD1 complex using previously determined structures. Satisfyingly, the model shows SAMHD1 positioned close to the RING domain of RBX1, a location that would be expected to render it receptive to ubiquitination.

Thus, these new reports show that although Vif and Vpx both manipulate CRLs to recognize host antiviral proteins and trigger their ubiquitination, they achieve this through contrasting mechanisms: Vif occupies a central organizing position and acts as a substrate receptor, whereas Vpx operates more peripherally to remodel the substrate receptor and facilitate substrate binding. These papers highlight not only the remarkable structural flexibility of assembled CRLs, but also the unrelenting capacity of viral proteins to commandeer cellular pathways for the benefit of the virus. Visualizing such structures and their underpinning protein interactions at atomic-level detail should inspire rational drug-design efforts aimed at interfering with fundamental aspects of CRL function. Pharmacological interventions that spare restriction factors from virus-induced elimination may offer a further therapeutic approach for treating HIV infections. ■

**Michael H. Malim** *is in the Department of Infectious Diseases, King's College London, London SE1 9RT, UK.*
*e-mail: michael.malim@kcl.ac.uk*

1. Malim, M. H. & Bieniasz, P. D. *Cold Spring Harb. Perspect. Med.* **2,** a006940 (2012).
2. Guo, Y. *et al. Nature* **505,** 229–233 (2014).
3. Schwefel, D. *et al. Nature* **505,** 234–238 (2014).
4. Baldauf, H.-M. *et al. Nature Med.* **18,** 1682–1689 (2012).
5. Zimmerman, E. S., Schulman, B. A. & Zheng, N. *Curr. Opin. Struct. Biol.* **20,** 714–721 (2010).
6. Jäger, S. *et al. Nature* **481,** 371–375 (2012).
7. Zhang, W., Du, J., Evans, S. L., Yu, Y. & Yu, X.-F. *Nature* **481,** 376–379 (2012).
8. Lim, E. S. *et al. Cell Host Microbe* **11,** 194–204 (2012).
9. Goldstone, D. C. *et al. Nature* **480,** 379–382 (2011).

**CORRECTION**

The News & Views article 'Cell biology: The beginning of the end' by Judith Campisi (*Nature* **505,** 35–36; 2014) omitted the name of the first author of ref. 5, Muñoz-Espín, in the final sentence of the first paragraph. The online versions of the article are correct.

ARTICLE

doi:10.1038/nature12789

# Diversity of ageing across the tree of life

Owen R. Jones[1,2]*, Alexander Scheuerlein[3]*, Roberto Salguero-Gómez[3,4], Carlo Giovanni Camarda[5], Ralf Schaible[3], Brenda B. Casper[6], Johan P. Dahlgren[1,2], Johan Ehrlén[7], María B. García[8], Eric S. Menges[9], Pedro F. Quintana-Ascencio[10], Hal Caswell[2,3,11,12], Annette Baudisch[3] & James W. Vaupel[1,3,13]

**Evolution drives, and is driven by, demography. A genotype moulds its phenotype's age patterns of mortality and fertility in an environment; these two patterns in turn determine the genotype's fitness in that environment. Hence, to understand the evolution of ageing, age patterns of mortality and reproduction need to be compared for species across the tree of life. However, few studies have done so and only for a limited range of taxa. Here we contrast standardized patterns over age for 11 mammals, 12 other vertebrates, 10 invertebrates, 12 vascular plants and a green alga. Although it has been predicted that evolution should inevitably lead to increasing mortality and declining fertility with age after maturity, there is great variation among these species, including increasing, constant, decreasing, humped and bowed trajectories for both long- and short-lived species. This diversity challenges theoreticians to develop broader perspectives on the evolution of ageing and empiricists to study the demography of more species.**

To examine demographic age trajectories across the tree of life, we studied life tables[1] (that is, patterns of mortality and fertility over age) and population projection matrices[2] for multicellular species from a wide range of taxonomic groups (Fig. 1; see Supplementary Methods for data sources and further rationale). We strived to find species with reliable data and from diverse taxa. From the data for each species we estimated smoothed trajectories of fertility, mortality and survivorship over age. Further research will undoubtedly refine the curves shown for many of the species in Fig. 1 and reveal variation in different environments and for different genotypes, but the general patterns are, we believe, serviceably accurate.

We standardized the demographic trajectories to facilitate comparison. Specifically we standardized the age axis so that it starts at the mean age of reproductive maturity and ends at a terminal age when only 5% of adults are still alive. After this terminal age, sample sizes were usually small and determination of age was often problematic. Fertility and mortality were mean-standardized by dividing age-specific fertility and mortality by the respective weighted average levels of fertility and mortality for all adults alive from maturity to the terminal age (see Methods). We refer to these standardized values as relative fertility and relative mortality. From the highest level of relative mortality at the terminal age (Fig. 1, top left) to the lowest level (Fig. 1, bottom right), species are ordered sequentially, row-by-row and from left-to-right. For the 46 diverse species depicted here, the range of variation in trajectories of fertility and mortality is unexpected. As an indication of variability across species, in modern Japanese women (Fig. 1, top left), mortality at the terminal age (102 years) is more than 20 times higher than the average level of adult mortality, whereas for white mangrove (*Avicennia marina*; Fig. 1, bottom right) the level of mortality at 123 years is less than half the average adult value.

Such variability is not predicted by the standard evolutionary theories of ageing[1,3–6]. Such theories provide explanations solely for age patterns of increasing mortality and decreasing fertility from maturity; the disposable soma theory[6] does so for species that segregate the germ line from the soma. Furthermore, for those species that show a lifetime increase in mortality, the canonical theory cannot account for the different magnitudes of that increase, although the disposable soma theory points to the crucial importance of trade-offs between the allocation of limited resources to repair and maintenance versus fertility and other imperatives.

## Mortality

The most notable pattern is the mortality trajectory for post-industrial humans, exemplified by Japanese women in 2009. The steep rise in relative mortality for the Japanese women is extreme compared with the rise for other species and sharper than that for historical populations such as the Swedish cohort born in 1881 and for hunter-gatherers such as the Aché of Paraguay whose mortality experience may be typical of humans over most of human existence[1,7]. The increased steepness of the rise of human mortality has largely occurred over the past century, indicating that it was behavioural and environmental change (including advances in health care) and not genetic change that moulded the current pattern[7–9]. Our close relatives, chimpanzees (*Pan troglodytes*) and baboons (*Papio cynocephalus*) also show a rise in mortality with age but far less than that for hunter-gatherers.

In several species mortality declines with age (Fig. 1, bottom row) and, in some cases, notably for the desert tortoise (*Gopherus agassizii*), the decline persists up to the terminal age. In other cases, an initial decline is followed by more or less constant mortality (for example, netleaf oak, *Quercus rugosa*). For species for which the underlying data are based on stages, such as dwarf gorse (*Ulex minor*) or the red-legged frog (*Rana aurora*), an asymptote is inevitable at older ages[8,10]. To alert readers to this, the mortality (and fertility and survival) curves derived from stage-classified models are represented by dashed curves in Fig. 1

[1]Max-Planck Odense Center on the Biodemography of Aging, Campusvej 55, 5230 Odense M, Denmark. [2]Department of Biology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark. [3]Max Planck Institute for Demographic Research, Konrad-Zuse-Strasse 1, 18057 Rostock, Germany. [4]School of Biological Sciences, Centre for Biodiversity and Conservation Science, University of Queensland, Brisbane QLD 4072, Australia. [5]Institut National d'Etudes Démographiques, 133 Boulevard Davout, 75980 Paris Cédex 20, France. [6]Department of Biology, University of Pennsylvania, 433 South University Avenue, Philadelphia, Pennsylvania 19104-6018, USA. [7]Department of Ecology, Environment and Plant Sciences, Stockholm University, Lilla Frescativägen 5, 10691 Stockholm, Sweden. [8]Pyrenean Institute of Ecology (CSIC), Avenida Montañana 1005, 50059 Zaragoza, Spain. [9]Archbold Biological Station, 123 Main Drive, Venus, Florida 33960, USA. [10]Department of Biology, University of Central Florida, 4110 Libra Drive, Orlando, Florida 32816-2368, USA. [11]Woods Hole Oceanographic Institution, Biology Department MS-34, Woods Hole, Massachusetts 02543 USA. [12]Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, PO Box 94248, 1090GE Amsterdam, The Netherlands. [13]Duke Population Research Institute, Duke University, Durham, North Carolina 27705, USA.

*These authors contributed equally to this manuscript.

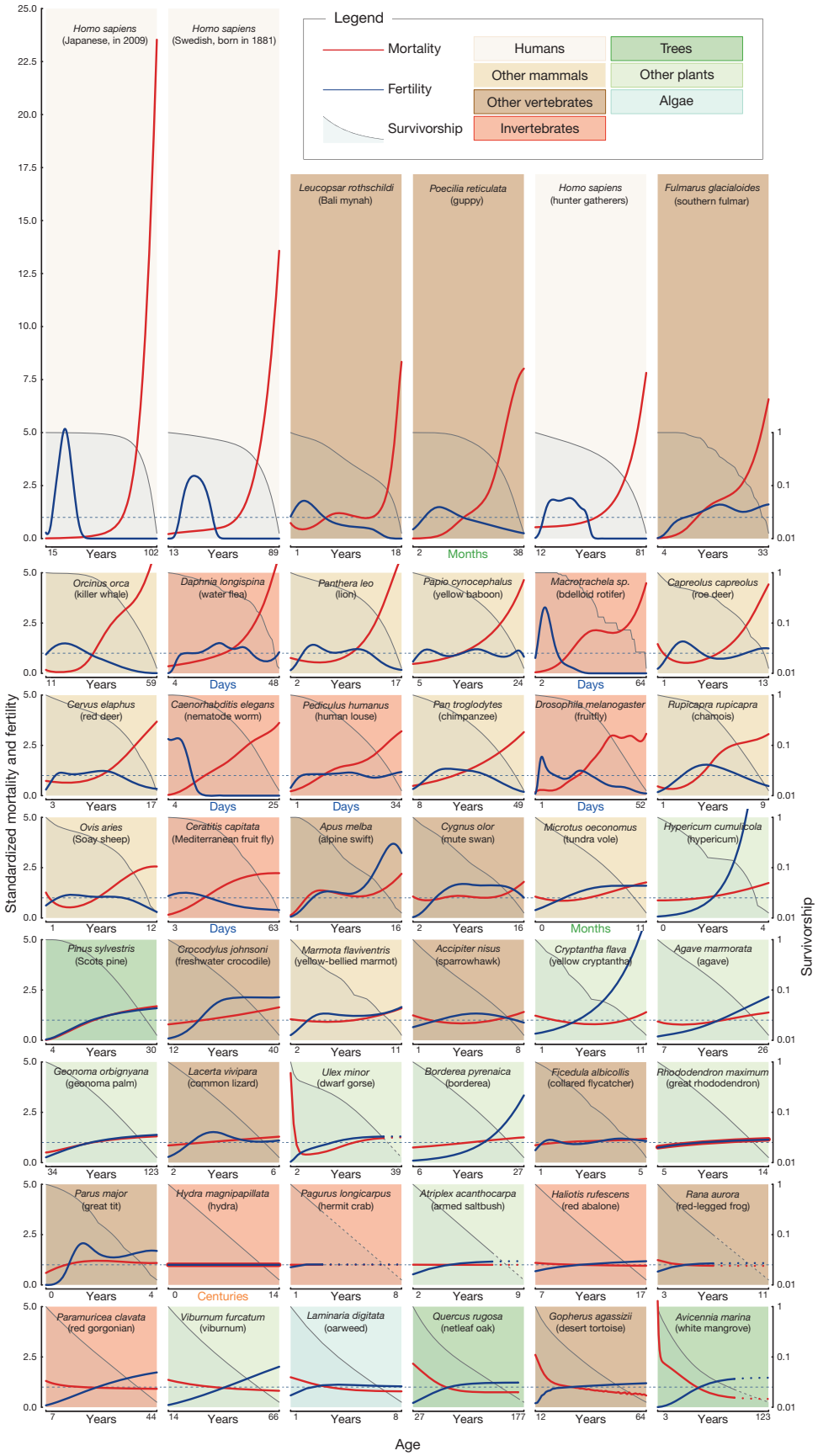



**Figure 1 | Demographic trajectories.** Relative mortality (red) and fertility (blue) as functions of age, from maturity to the age when only 5% of the adult population is still alive; mortality and fertility are scaled relative to their means. Subplots are arranged in order of decreasing relative mortality at the terminal age. Survivorship (on a log scale) from maturity is depicted by the shaded areas. Broken lines, for trajectories derived from projection matrices, start at the age when cohorts have converged to within 5% of their quasi-stationary distribution (see also Supplementary Methods).

at ages beyond which a cohort will have converged to within 5% of the quasi-stationary distribution (see Methods).

For most species in Fig. 1 the age pattern of mortality is derived from data on ages rather than stages. For some of these species, mortality levels off at advanced ages (for example, for the collared flycatcher, *Ficedula albicollis,* the great tit, *Parus major,* the fruitfly, *Drosophila melanogaster*) and in others remains constant at all adult ages (for example, for *Hydra magnipapillata*). For hydra in the laboratory, this risk is so small that we estimate that 5% of adults would still be alive after 1,400 years under those controlled conditions.

## Fertility

The fertility trajectories show considerable variation. For humans the trajectories are bell-shaped and concentrated at younger adult ages, but other shapes are apparent in Fig. 1. The patterns for killer whales (*Orcinus orca*), chimpanzees, chamois (*Rupicapra rupicapra*) and sparrowhawks (*Accipiter nisus*) are also approximately bell-shaped but spread over more of the course of life. Other species show trajectories of gradually increasing fertility (for example, southern fulmars, and the agave, *Agave marmorata*), asymptotic fertility (for example, tundra voles, *Microtus oeconomus*), or constant fertility (for example, hydra). In addition to humans and killer whales, bdelloid rotifers (*Macrotrachela* sp.), nematode worms (*Caenorhabditis elegans*) and Bali mynah birds (*Leucopsar rothschildi*) have post-reproductive life spans, which lends further support to the idea that this phenomenon may be widespread[3–6,11].

## Axes of senescence

Although the demographic trajectories in Fig. 1 vary widely, most of the 46 species can be roughly classified along a continuum of senescence; running from strong deterioration with age, to negligible deterioration, to negative senescence[12] and improvement with age. However, there are some deviations, for example, for Soay sheep (*Ovis aries*) and dwarf gorse, which show mortality reductions with adult age followed by deterioration. Fertility patterns show similar diversity.

A fast–slow continuum has been proposed to order species from those with short lives and intense early reproduction to those with long lives and an extended reproductive period[13–16]. Figure 1 displays mortality and fertility over the adult lifespan; pre-reproductive mortality trajectories are also of interest but beyond the scope of this article. If distinguished by the length of life, then fast and slow life histories are scattered irregularly across Fig. 1. Lifespans range from 1,400 years for the hydra to just 25 days for nematode worms. Species with fast life histories, such as water fleas (*Daphnia longispina*), are followed in Fig. 1 by species with slow life histories, such as the lion, and those with slow life histories, such as the chimpanzee, occur adjacent to those with fast life histories, such as the human louse (*Pediculus humanus*) and the fruitfly (*D. melanogaster*). Furthermore, species with very different life spans can display similar patterns of mortality, fertility and survivorship. For example, the water flea's trajectories are similar to the fulmar's, although water fleas reach advanced old age at 48 days, whereas the fulmars do so at 33 years.

If senescence is measured by how long it takes for death rates to rise from some level to a higher level, then long-lived species senesce slowly. It is more interesting to define senescence by the sharpness or abruptness rather than the speed of the increase in mortality. Baudisch[8] distinguishes the pace of life; that is, whether reproduction is fast and life spans are short or reproduction is slow and life spans are long, from the shape of mortality and fertility trajectories (whether mortality rises sharply with age and fertility falls sharply or whether mortality and fertility levels are more constant). One measure of pace, the measure that we have used, is the terminal age to which only 5% of adults survive; this measure is in days or years or some other unit of time. One measure of shape, the measure that we have used, is the ratio of mortality at the terminal age to the average level of adult mortality; this time-invariant measure does not change if time is measured in days versus years. More senescent species, with sharper increases in mortality with age, have higher values of this measure of shape.

The measure can be used to explore further the unexpected lack of association between the length of life and the degree of senescence. Among the first 24 graphs, those with the sharpest senescence, 11 species have relatively long life spans and 13 have relatively short life spans. Among the final 24 graphs, those with less senescence, 13 species have relatively long life spans and 11 have relatively short life spans. This weak negative association between the length of life and the degree of senescence is reflected in a weak Spearman rank correlation of $-0.13$, which is not significantly different from zero ($P = 0.362$). The Spearman correlations are also non-significant when assessed for animals ($P = 0.414$) and for plants ($P = 0.07$) examined separately. If the 12 plants in Fig. 1 are cross-tabulated as longer or shorter lived, and as more or less senescent, then three species fall into each of the four categories. Hence the data support Baudisch's[8] conjecture that pace and shape may be two orthogonal axes of life histories.

A survivorship curve indicates the proportion of individuals that are still alive at a given age. In Fig. 1, we plot survivorship from reproductive maturity on a logarithmic scale. If mortality increases with age, the log-survivorship curve is concave. If mortality is independent of age, log-survivorship is linear (for example, roughly from the hydra to the red abalone (*Haliotis rufesens*) in Fig. 1). For species with death rates that decline with age, the curve is convex (for example, from the red-legged frog to the white mangrove at the bottom of Fig. 1). The classification of survivorship curves into concave, linear and convex curves is known among biologists as type I, II and III, respectively[17,18], but normally the curves are plotted for lifespans starting at birth rather than at maturity. When the evolutionary theory of ageing[3–6] was being developed, there was very little empirical evidence for type III survivorship for adults and little evidence for type II survivorship. The widespread recognition that traditional theories of ageing predict adult senescence to be a universal trait led researchers to strive to find evidence for senescence in, for example, the mute swan (*Cygnus olor*)[19]. For this species, fertility does decline and mortality does increase at the oldest ages. However, the overall life course is characterized by fertility that increases and then slowly declines and by roughly constant mortality: the log-survivorship curve is nearly straight. It is clear from our analyses that the full spectrum of type I, II and III survivorship curves are found for adults in nature.

## Phylogenetic patterns

Phylogenetic relatedness seems to have some role in the order of species in Fig. 1, as shown by taxonomic clustering of mortality, fertility and survivorship patterns. All mammals are clustered in the top part of Fig. 1, whereas birds are somewhat more scattered, from the Bali mynah in the first row to the great tit in the seventh row. Amphibians and reptiles are found in the lower half of the panel, with flat mortality shapes and almost no overlap with mammals. In contrast, invertebrates are scattered across the continuum of senescence, with bdelloid rotifers and water fleas sharing the mammalian mortality pattern. The plants in our sample tend to occur lower in our ordering, with the first being *Hypericum cumulicola*. Although some angiosperm species seem to senesce[20–22], many angiosperm species seem not to[23], perhaps as an artefact of the use of stage-based data[10]. The only alga in our data set, oarweed (*Laminaria digitata*), falls in the last row.

Such clustering within broad taxonomic levels of kingdom (plants, animals), or class (mammals, birds), suggests that primitive traits related to the *bauplan* of species may have a pivotal role in determining patterns of ageing. In fact, the evolutionary conservatism of mechanistic determinants of ageing has been highlighted by genetic studies[24] and it has been suggested that asexual reproduction[25], modularity[26], lack of germ-line sequestration from the soma[27,28], the importance of protected niches[29], regenerative capacity, and the paucity of diverse cell types[30], may facilitate the escape from senescence in some clades. Many of the species in the lower half of Fig. 1—the reptiles, vascular plants, alga,

and coral—continue to grow after reproductive maturity to sizes much larger than those at maturity. For these indeterminate growers, mortality is approximately constant or decreases somewhat with age, whereas fertility is more or less constant or increases to some extent. Species with indeterminate growth may exhibit patterns of senescence that are fundamentally different from those of species with determinant growth[12,31–33].

Approximately constant mortality and fertility are experienced by vertebrates such as collared flycatchers and red-legged frogs, invertebrates such as hermit crabs (*Pagurus longicarpus*) and red abalone, and vascular plants such as great rhododendron (*Rhododendron maximum*) and armed saltbush (*Atriplex acanthocarpa*), with the age at 5% survivorship ranging from 5 years for the collared flycatcher to 14 centuries for hydra. It remains to be seen whether the similarity of patterns of mortality, fertility and survivorship among disparate groups of species is a coincidence or represents convergent solutions to similar evolutionary challenges.

## Continuing the exploration of ageing

Although hundreds of theories have been proposed to explain the proximate mechanisms of ageing[34,35], theories to explain the ultimate evolutionary causes of the varieties of ageing, illustrated by the diverse range of trajectories in Fig. 1, are in their infancy. However, scattered studies suggest profitable directions for research. It is only recently that researchers have extended their analyses beyond the traditional age-structured framework[36]; more complex demographic models show that selection gradients in clonal or stage-structured organisms can be non-monotonic[37–40]. As recognized in the disposable soma theory[6], differences in life-history constraints among species, and the resulting differences in optimal resource allocation among vital processes provide a promising direction for explaining empirical observations about diverse fertility[32,37–39,41,43] and mortality[32,41–43] trajectories. However, current theoretical approaches do not yet explain in detail why senescence has evolved in some species and not in others. Data sets that are currently available for research on ageing are taxonomically biased: high-quality data on hundreds of mammalian and bird species exist but data on other vertebrate taxa and on invertebrates are sparse. There is very limited knowledge of the age patterns of mortality and fertility in species of algae, fungi and bacteria[32,43,44].

The mortality and fertility trajectories of any species depend on the environment in which they are measured. Most human experience is bounded by the trajectories of modern Japanese and the hunter-gatherers in Fig. 1. Although population ecologists have long studied the responses of mortality and fertility to environmental factors, few studies have focused on the shape of the age trajectories. Environmental and genotypic variation has been documented in laboratory studies of nematode worms, medflies, *Drosophila* and other model species[45], and in a field study of *Plantago*[20]. Available evidence suggests that variation can be considerable for a species but that the qualitative shapes of mortality and fertility trajectories are similar, as illustrated by humans in Fig. 1 (see the Supplementary Note and Extended Data Fig. 1, which highlights intraspecific variation in the mortality trajectories of laboratory rats (*Rattus norvegicus*) and mice (*Mus musculus*)). In addition to the lack of data for most species, and for variation within a species, little information is available on mortality at advanced ages beyond the age cut-off in Fig. 1. In the species for which such data are available, mortality approaches a plateau at the oldest ages (for example, for humans, fruitflies (*D. melanogaster*) and nematode worms) or declines (for Mediterranean fruitflies, *Ceratitis capitata*)[45–47]. The deceleration of mortality at high ages is more apparent if death rates are plotted on a log scale rather than the linear scale used in Fig. 1 (ref. 45).

Deeper understanding of the evolutionary demography of ageing depends on the compilation of demographic data on diverse species investigated in the wild as well as in laboratories and zoos[8], and on the development of more inclusive theories that can account for negligible and negative senescence[42,48] as well as for the steepness of deterioration with age in senescent species. In such empirical and theoretical studies, researchers should guard against anthropocentric intuition about ageing: humans, especially modern humans, are extreme outliers in Fig. 1.

## METHODS SUMMARY

**Selection of examples.** We aimed to examine demographic trajectories for organisms across the tree of life. We therefore chose representative data sets compiled from the published literature for the major groups of organisms including vertebrate and invertebrate animals, plants and algae. Within the vertebrates we included exemplars of every major clade, including primates and other mammals, birds, reptiles, amphibians and fish. Representatives for the invertebrates included insects, molluscs, cnidarians and a crustacean. In the plant group we included both gymnosperms and angiosperms and, finally, we included a green alga. We favoured data sets that covered longer time periods, with larger sample sizes and, when possible, we preferred data sets that included information on realized reproduction and recruitment to those that simply recorded reproductive output. In addition, for dioecious species, we favoured data sets for females. See Supplementary Methods 1 and 2 for details.

**Calculation of standardised trajectories.** We classified the studies as: first, cohort studies; second, period studies with number at risk and numbers dying within a period; third, period studies depicting an age structure at a single point in time; or fourth, stage-structured population projection matrices (see Supplementary Methods 2 for details). We considered mortality and fertility trajectories from the age at maturity to the age at which 5% survivorship from maturity occurs. The trajectories of all data types, except the projection matrix data, were smoothed using P-splines[49]. We then calculated the force of mortality ($\mu_x$) and fertility rate ($m_x$) before standardizing them by dividing them by the respective averages, weighted by survivorship from maturity ($l_x$).

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 8 July; accepted 18 October 2013.**
**Published online 8 December 2013.**

1. Chiang, C. L. *The life table and its applications* (Krieger Publishing, 1984).
2. Caswell, H. *Matrix population models* (Sinauer Associates, 2001).
3. Medawar, P. B. *An unsolved problem of biology* (H. K. Lewis, 1952).
4. Williams, G. Pleiotropy, natural selection, and the evolution of senescence. *Evolution* **11,** 398–411 (1957).
5. Hamilton, W. D. The moulding of senescence by natural selection. *J. Theor. Biol.* **12,** 12–45 (1966).
6. Kirkwood, T. B. L. Evolution of ageing. *Nature* **270,** 301–304 (1977).
7. Burger, O., Baudisch, A. & Vaupel, J. W. Human mortality improvement in evolutionary context. *Proc. Natl Acad. Sci. USA* http:// dx.doi.org/10.1073/pnas.1215627109 (15 October 2012).
8. Baudisch, A. The pace and shape of ageing. *Methods Ecol. Evol.* **2,** 375–382 (2011).
9. Oeppen, J. & Vaupel, J. W. Broken limits to life expectancy. *Science* **296,** 1029–1031 (2002).
10. Horvitz, C. C. & Tuljapurkar, S. Stage dynamics, period survival, and mortality plateaus. *Am. Nat.* **172,** 203–215 (2008).
11. Cohen, A. A. Female post-reproductive lifespan: a general mammalian trait. *Biol. Rev. Camb. Philos. Soc.* **79,** 733–750 (2004).
12. Vaupel, J. W., Baudisch, A., Dölling, M., Roach, D. A. & Gampe, J. The case for negative senescence. *Theor. Popul. Biol.* **65,** 339–351 (2004).
13. Gaillard, J.-M. *et al.* An analysis of demographic tactics in birds and mammals. *Oikos* **56,** 56–76 (1989).
14. Promislow, D. E. L. & Harvey, P. H. Living fast and dying young: a comparative analysis of life-history variation among mammals. *J. Zool. (Lond.)* **220,** 417–437 (1990).
15. Stearns, S. C. *The Evolution of Life Histories* (Oxford Univ. Press, USA, 1992).
16. Jones, O. R. *et al.* Senescence rates are determined by ranking on the fast-slow life-history continuum. *Ecol. Lett.* **11,** 664–673 (2008).
17. Pearl, R. & Miner, J. R. Experimental studies on the duration of life. XIV. The comparative mortality of certain lower organisms. *Q. Rev. Biol.* **10,** 60–79 (1935).
18. Deevey, E. S. Life tables for natural populations of animals. *Q. Rev. Biol.* **22,** 283–314 (1947).
19. Charmantier, A., Perrins, C., McCleery, R. H. & Sheldon, B. C. Quantitative genetics of age at reproduction in wild swans: Support for antagonistic pleiotropy models of senescence. *Proc. Natl Acad. Sci. USA* **103,** 6587–6592 (2006).
20. Shefferson, R. P. & Roach, D. A. Longitudinal analysis in *Plantago*: strength of selection and reverse age analysis reveal age-indeterminate senescence. *J. Ecol.* **101,** 577–584 (2013).
21. Tuomi, J. *et al.* Prolonged dormancy interacts with senescence for two perennial herbs. *J. Ecol.* **101,** 566–576 (2013).
22. Salguero-Gómez, R., Shefferson, R. P. & Hutchings, M. J. Plants do not count… or do they? New perspectives on the universality of senescence. *J. Ecol.* **101,** 545–554 (2013).

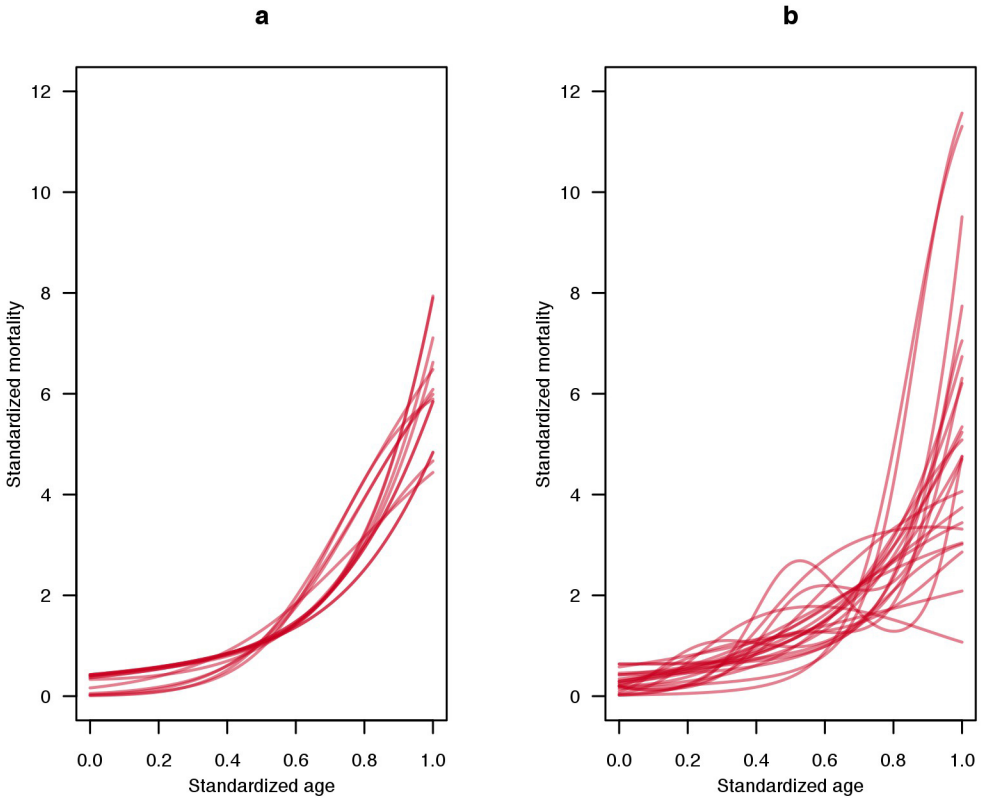
23. Baudisch, A. *et al.* The pace and shape of senescence in angiosperms. *J. Ecol.* **101,** 596–606 (2013).
24. McElwee, J. J. *et al.* Evolutionary conservation of regulated longevity assurance mechanisms. *Genome Biol.* **8,** R132 (2007).
25. Bell, G. Measuring the cost of reproduction. I. The correlation structure of the life table of a plank rotifer. *Evolution* **38,** 300–313 (1984).
26. Franco, M. & Silvertown, J. Life history variation in plants: an exploration of the fast-slow continuum hypothesis. *Phil. Trans. R. Soc. B* 1341–1348 (1996).
27. Buss, L. W. Diversification and germ-line determination. *Paleobiology* **14,** 313–321 (1988).
28. Martínez, D. E. & Levinton, J. S. Asexual metazoans undergo senescence. *Proc. Natl Acad. Sci. USA* **89,** 9920–9923 (1992).
29. Baudisch, A. & Vaupel, J. Senescence vs. sustenance: evolutionary-demographic models of aging. *Demogr. Res.* **23,** 655–668 (2010).
30. Martínez, D. E. Mortality patterns suggest lack of senescence in hydra. *Exp. Gerontol.* **33,** 217–225 (1998).
31. Finch, C. E. *Longevity, Senescence and the Genome* (Univ. Chicago Press, 1994).
32. Baudisch, A. *Inevitable Aging? Contributions to Evolutionary-Demographic Theory* (Springer, 2008).
33. Charnov, E. L. Reproductive constraints and the evolution of life histories with indeterminate growth. *Proc. Natl Acad. Sci. USA* **98,** 9460–9464 (2001).
34. Medvedev, Z. A. An attempt at a rational classification of theories of ageing. *Biol. Rev. Camb. Philos. Soc.* **65,** 375–398 (1990).
35. Kirkwood, T. B. L. Systems biology of ageing and longevity. *Phil. Trans. R. Soc. B* **366,** 64–70 (2010).
36. Charlesworth, B. *Evolution in Age-structured Populations* (Cambridge Univ. Press, 1994).
37. Caswell, H. Matrix models and sensitivity analysis of populations classified by age and stage: a vec-permutation matrix approach. *Theor. Ecol.* **5,** 403–417 (2012).
38. Pedersen, B. An evolutionary theory of clonal senescence. *Theor. Popul. Biol.* **47,** 292–320 (1995).
39. Caswell, H. in *Population Biology and Evolution of Clonal Organisms* (eds Jackson, J. B. C., Bus, L. W. & Cook, R. E.) 187–224 (Yale Univ. Press, 1985).
40. Caswell, H. & Salguero-Gomez, R. Age, stage and senescence in plants. *J. Ecol.* **101,** 585–595 (2013).
41. Orive, M. E. Senescence in organisms with clonal reproduction and complex life histories. *Am. Nat.* **145,** 90–108 (1995).
42. Baudisch, A. & Vaupel, J. W. Getting to the root of aging. *Science* **338,** 618–619 (2012).
43. Gadgil, M. & Bossert, W. H. Life historical consequences of natural selection. *Am. Nat.* **104,** 1–24 (1970).
44. Schaffer, W. M. Selection for optimal life histories — effects of age structure. *Ecology* **55,** 291–303 (1974).
45. Vaupel, J. W. *et al.* Biodemographic trajectories of longevity. *Science* **280,** 855–860 (1998).
46. Chen, J. *et al.* A demographic analysis of the fitness cost of extended longevity in *Caenorhabditis elegans*. *J. Gerontol. A Biol. Sci. Med. Sci.* **62,** 126–135 (2007).
47. Vaupel, J. W. Biodemography of human ageing. *Nature* **464,** 536–542 (2010).
48. Finch, C. E. Update on slow aging and negligible senescence - A mini-review. *Gerontology* **55,** 307–313 (2009).
49. Eilers, P. H. C. & Marx, B. D. Flexible smoothing with B-splines and penalties. *Stat. Sci.* **11,** 89–121 (1996).


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

**Acknowledgements** We thank S. Alberts for data on baboon demography, J. Curtsinger for data on *Drosophila* demography and O. Burger, D. Levitis, B. Pietrzak, F. Quade, F. Ringelhan and L. Vinicius for contributing published data about various species. J.W.V. and A.S. acknowledge support from NIH grant PO1 AG-031719. H.C. acknowledges a Research Award from the Alexander von Humboldt Foundation and Advanced Grant 322989 from the European Research Council. R.S.-G. acknowledges support from ARC DP110100727. A.B. acknowledges funding from the Max Planck Society to establish the Max Planck Research Group 'Modeling the Evolution of Aging'.

**Author Contributions** This research project was initiated by J.W.V. A.S. wrote the first draft; O.R.J., with help from A.S., R.S.-G., H.C., A.B. and J.W.V., wrote subsequent drafts; J.W.V. and O.R.J. completed the final draft. The Figure was produced by O.R.J. with suggestions from J.W.V., A.S., A.B. and H.C. A.B. suggested the method of standardization and the distinction between shape and pace. C.G.C. developed methods to smooth mortality and fertility trajectories. H.C. and R.S.-G. contributed to the analysis of stage-classified species. A.S., R.S.-G., O.R.J. and H.C. each provided data, derived from the literature, for several species. R.S. contributed unpublished data for hydra; J.E., J.D. and M.B.G. for *Borderea*; R.S.-G. and B.B.C. for *Cryptantha*; and E.M. and P.F.Q.-A. for *Hypericum*. O.R.J., A.S., R.S.-G. and H.C. screened the species for data quality.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.R.J. (jones@biology.sdu.dk).




**Extended Data Figure 1 | Standardized mortality trajectories.** **a**, Trajectories for laboratory rats. **b**, Trajectories for laboratory mice. Each line represents a different strain, sex or population (see Supplementary Methods for sources). We standardized the age axis to consider the trajectories from age at maturity to the age at which 5% survivorship from maturity occurs. The trajectories were smoothed using P-splines. We then calculated the force of mortality ($\mu_x$) and standardized it by dividing by the average value, weighted by survivorship from maturity ($l_x$). Note that the sample sizes in most cases were small (approximately 50 to 60 individuals) and thus random fluctuations may lead to erratic curves in some cases.

ARTICLE

OPEN

doi:10.1038/nature12826

# Elephant shark genome provides unique insights into gnathostome evolution

Byrappa Venkatesh[1,2], Alison P. Lee[1], Vydianathan Ravi[1], Ashish K. Maurya[3], Michelle M. Lian[1], Jeremy B. Swann[4], Yuko Ohta[5], Martin F. Flajnik[5], Yoichi Sutoh[6], Masanori Kasahara[6], Shawn Hoon[7], Vamshidhar Gangu[7], Scott W. Roy[8], Manuel Irimia[9], Vladimir Korzh[10], Igor Kondrychyn[10], Zhi Wei Lim[1], Boon-Hui Tay[1], Sumanty Tohari[1], Kiat Whye Kong[7], Shufen Ho[7], Belen Lorente-Galdos[11,12], Javier Quilez[11,12], Tomas Marques-Bonet[11,12], Brian J. Raney[13], Philip W. Ingham[3], Alice Tay[1], LaDeana W. Hillier[14], Patrick Minx[14], Thomas Boehm[4], Richard K. Wilson[14], Sydney Brenner[1] & Wesley C. Warren[14]

**The emergence of jawed vertebrates (gnathostomes) from jawless vertebrates was accompanied by major morphological and physiological innovations, such as hinged jaws, paired fins and immunoglobulin-based adaptive immunity. Gnathostomes subsequently diverged into two groups, the cartilaginous fishes and the bony vertebrates. Here we report the whole-genome analysis of a cartilaginous fish, the elephant shark (*Callorhinchus milii*). We find that the *C. milii* genome is the slowest evolving of all known vertebrates, including the 'living fossil' coelacanth, and features extensive synteny conservation with tetrapod genomes, making it a good model for comparative analyses of gnathostome genomes. Our functional studies suggest that the lack of genes encoding secreted calcium-binding phosphoproteins in cartilaginous fishes explains the absence of bone in their endoskeleton. Furthermore, the adaptive immune system of cartilaginous fishes is unusual: it lacks the canonical CD4 co-receptor and most transcription factors, cytokines and cytokine receptors related to the CD4 lineage, despite the presence of polymorphic major histocompatibility complex class II molecules. It thus presents a new model for understanding the origin of adaptive immunity.**

The emergence of gnathostomes from jawless vertebrates marks a major event in the evolution of vertebrates. This transition was accompanied by many morphological and phenotypic innovations, such as jaws, paired appendages and an adaptive immune system based on immunoglobulins, T-cell receptors and major histocompatibility complex (MHC) molecules[1–4] (Fig. 1a). How these novelties emerged and how they facilitated the divergence, adaptation and dominance of gnathostomes as the major group (99.9%) of living vertebrates are key unresolved questions. The living gnathostomes are divided into two groups, the cartilaginous fishes (Chondrichthyes) and bony vertebrates (Osteichthyes), which diverged about 450 Myr ago (Fig. 1a). A key feature distinguishing the two groups is that chondrichthyans have largely cartilaginous endoskeletons whereas osteichthyans have ossified endoskeletons. Although fossil jawless vertebrates (for example galeaspids) and jawed vertebrates (for example placoderms) possessed dermal and perichondral bone, endochondral bone is found only in osteichthyans[5]. Chondrichthyans include about 1,000 living species that are grouped into two lineages, the holocephalans (chimaeras) and elasmobranchs (sharks, rays and skates), which diverged about 420 Myr ago[6] (Fig. 1a). A detailed whole-genome evaluation of a chondrichthyan and comparative analysis with the available genome information on osteichthyans and a jawless vertebrate[7] might help us to understand features unique to chondrichthyans and provide insights into the ancestral state of gnathostome-specific morphological features and physiological systems.

We previously identified *C. milii*, a holocephalan, as a chondrichthyan genome model[8,9] because of its relatively small genome (~1.0 gigabase). Compared with elasmobranchs, the unique features of holocephalans include a single gill opening, a complete hyoid arch, fusion of the upper jaw to the cranium, and non-replaceable hypermineralized tooth plates[10] (Fig. 1b). *Callorhinchus milii* inhabits temperate waters of the continental shelves off southern Australia and New Zealand, typically at depths of 200 to 500 m (ref. 11). Here, we report the generation and analysis of a high-quality genome sequence of *C. milii*. Several key findings are presented here and further details on our in-depth characterization of this genome are presented in Supplementary Notes I to XI.

## Genome assembly and annotation

Genomic DNA of a single male *C. milii* was sequenced and assembled (Supplementary Note I) to a size of 0.937 gigabases, comprising 21,208 scaffolds (N50 contig, 46.6 kilobases; N50 scaffold, 4.5 megabases; Supplementary Table I.1). The average GC content of the *C. milii* genome is 42.3%, and approximately 46% of the genome is organized into isochores (Supplementary Note II). Using the Ensembl annotation pipeline[12] and RNA-seq transcript evidence, we predicted a total of 18,872 protein-coding genes. In addition, microRNA (miRNA) genes were identified by small-RNA sequencing and annotation of the genome assembly (Supplementary Note III). *Callorhinchus milii* have more miRNA gene loci (693 genes and 136 families) than do teleosts (for example, zebrafish have 344 genes and 94 families) but fewer than do humans (1,527 genes and 558 families) and other mammals (mirBase release 19). Several novel *C. milii*-specific miRNAs are expressed at high levels in a tissue-specific manner (Supplementary Figs III.1 and

[1]Comparative Genomics Laboratory, Institute of Molecular and Cell Biology, A*STAR, Biopolis, Singapore 138673. [2]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228. [3]Developmental and Biomedical Genetics Laboratory, Institute of Molecular and Cell Biology, A*STAR, Biopolis, Singapore 138673. [4]Department of Developmental Immunology, Max-Planck-Institute of Immunobiology and Epigenetics, Stuebeweg 51, 79108 Freiburg, Germany. [5]Department of Microbiology and Immunology, University of Maryland, Baltimore, Maryland 21201, USA. [6]Department of Pathology, Hokkaido University Graduate School of Medicine, Sapporo 060-8638, Japan. [7]Molecular Engineering Laboratory, Biomedical Sciences Institutes, A*STAR, Biopolis, Singapore 138673. [8]Department of Biology, San Francisco State University, San Francisco, California 94132, USA. [9]Banting and Best Department of Medical Research and Donnelly Centre, University of Toronto, Toronto, Ontario M5S 3E1, Canada. [10]Fish Developmental Biology Laboratory, Institute of Molecular and Cell Biology, A*STAR, Biopolis, Singapore 138673. [11]Institut de Biologia Evolutiva (UPF-CSIC), PRBB, 08003 Barcelona, Spain. [12]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain. [13]Center for Biomolecular Science and Engineering, School of Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA. [14]The Genome Institute at Washington University, St Louis, Missouri 63108, USA.

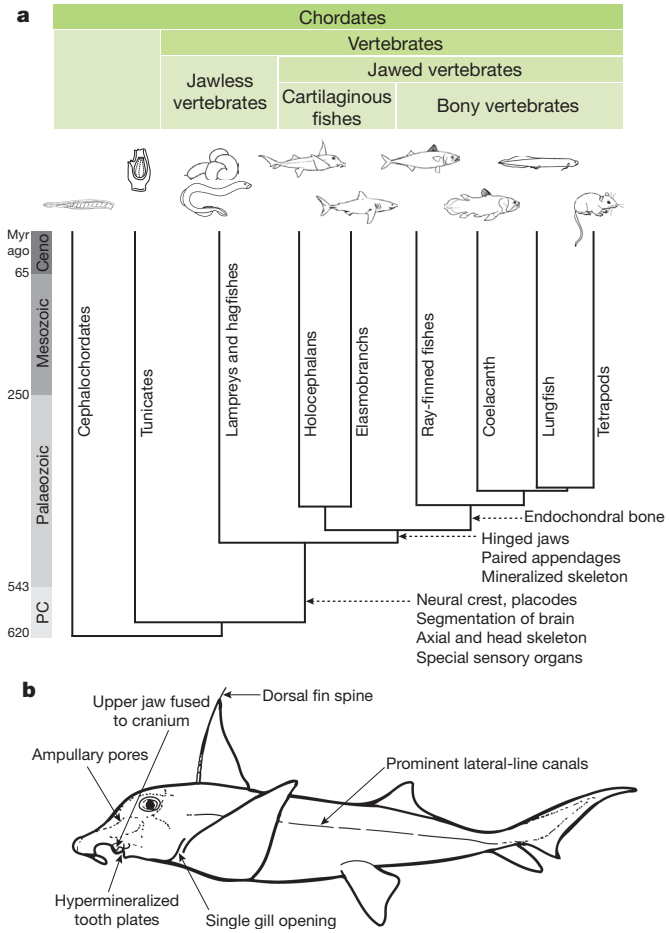



**Figure 1 | Phylogeny of chordates.** **a**, Major shared features of various vertebrate taxa; **b**, unique features of *C. milii*. Lampreys and hagfishes (cyclostomes) lack mineralized tissues. In contrast, cartilaginous fishes produce extensive dermal bone such as teeth, dermal denticle and fin spine. However, they lack the ability to make endochondral bone which is unique to bony vertebrates. Divergence times are from refs 6, 36. Ceno, Cenozoic era; PC, Precambrian era.

III.2). Notably, a considerable number (16%; 22/136) of *C. milii* miRNA families conserved in mammals have been secondarily lost in teleosts. Their explicit tissue-specific expression patterns in *C. milii* and mammals suggest that they have important roles in gene regulation. A total of 63,877 noncoding elements (average size, 271 base pairs) conserved between *C. milii* and bony vertebrates represent potential *cis*-regulatory elements (Supplementary Note IV). Surprisingly, only a tiny fraction (less than 1.0%) of these are found in the genomes of sea lamprey, sea squirt and amphioxus, suggesting that the emergence of gnathostomes was accompanied by major innovations in *cis*-regulatory elements and gene regulatory networks.

## Phylogenomics and evolutionary rate

Morphological and palaeontological studies have placed Chondrichthyes as a sister group to bony vertebrates[5]. However, subsequent molecular phylogenetic analyses based on mitochondrial or nuclear genes have produced conflicting topologies[13–15]. Using a genome-scale data set comprising 699 one-to-one orthologues from *C. milii* and 12 other chordates, we provide robust support for the traditional phylogenetic tree with an unambiguous split between Chondrichthyes and bony vertebrates (Supplementary Fig. V.1). Furthermore, analysis of gains and losses of introns provided independent support for Chondrichthyes as a sister group to bony vertebrates (Supplementary Note V).

Previous studies based on a few mitochondrial and nuclear protein-coding genes indicated that the nucleotide substitution rate in elasmobranchs is an order of magnitude lower than that in mammals[16,17]. Using the genome-wide set of 699 orthologues, we estimated the molecular evolutionary rate of *C. milii* and compared it with other gnathostomes, with sea lamprey as the outgroup. *Callorhinchus milii* protein-coding genes have evolved significantly slower than all other vertebrates examined ($P < 0.01$ for all comparisons; Supplementary Tables VI.1–VI.3), including the coelacanth, which has been considered to be the slowest evolving bony vertebrate[18]. A neutral tree based on fourfold-degenerate sites indicated that the low evolutionary rate is a reflection of the neutral nucleotide mutation rate, and confirmed that the neutral evolutionary rate of *C. milii* is the lowest (Fig. 2a).

The lower rates of molecular evolution of *C. milii* are also evident in the fewer changes in the intron–exon organization of its genes (Supplementary Note VII). *Callorhinchus milii* has experienced fewer intron gains or losses than any bony vertebrate since their divergence from the gnathostome ancestor (Fig. 2a). The highest rates of change were found in two teleost fishes, the stickleback and zebrafish, with the stickleback lineage experiencing the highest number of changes (603 gains and 126 losses after it split from the zebrafish lineage) recorded in any vertebrate lineage (Fig. 2a). In addition to a lower rate of intron changes, the *C. milii* genome also has experienced a relatively low rate of major interchromosomal rearrangements, comparable to that of chicken, which has one of the most stable karyotypes among tetrapods[19,20] (Supplementary Note VIII). An extensive conservation of synteny was observed in comparisons of *C. milii* scaffolds with chicken and human chromosomes, with a majority of *C. milii* scaffolds (93%) showing a one-to-one correspondence with chicken chromosomes (Supplementary Figs VIII.1 and VIII.2). A three-way comparison between *C. milii*, chicken and teleosts (medaka and zebrafish) revealed that teleosts have undergone a substantially higher number of interchromosomal rearrangements than previously demonstrated by simple tetrapod–teleost comparisons (Fig. 2b and Supplementary Tables VIII.7 and VIII.8).

## Evolution of protein-coding gene families

### Protein domains

Comparisons of Pfam domains (Supplementary Note IXa) identified 17 *C. milii* domains that are missing in bony vertebrates (Supplementary Table IX.2). Sixteen of these are present in amphioxus or other eukaryotes, and are thus ancient protein domains that are still retained in *C. milii* but have been lost from bony vertebrates. We note that 13 domains shared between *C. milii* and tetrapods are absent in teleosts (Supplementary Table IX.4), indicating that they have been secondarily lost from the teleost lineage.

### Lineage-specific gene losses

Orthologues of more *C. milii* genes were found to be lost from the teleost lineage (271 genes; Supplementary Note IXb and Supplementary Table IX.6) relative to the tetrapod lineage (34 genes; Supplementary Table IX.7). Human orthologues of many genes lost from teleosts are associated with genetic diseases (104 genes, 38%; Supplementary Table IX.6), indicating their importance for human physiology. The loss of these genes from teleosts supports the idea that teleosts represent a more derived group than do other gnathostomes. The functional annotation of zebrafish orthologues of the 34 genes lost from tetrapods highlighted several genes that are specific to the aquatic lifestyle, such as those regulating fin and lateral-line development and those encoding receptors for water-soluble odorants (Supplementary Table IX.7).

## Genetic basis of bone formation

Bone is the most widespread mineralized tissue in vertebrates, and its formation represents a major leap in vertebrate evolution. Although chondrichthyans produce dermal bone (for example teeth, dermal denticles and fin spines) and calcified cartilage[5,21], unlike bony vertebrates, their cartilage is not replaced with endochondral bone. Among vertebrates, the earliest mineralized tissue was found in the feeding

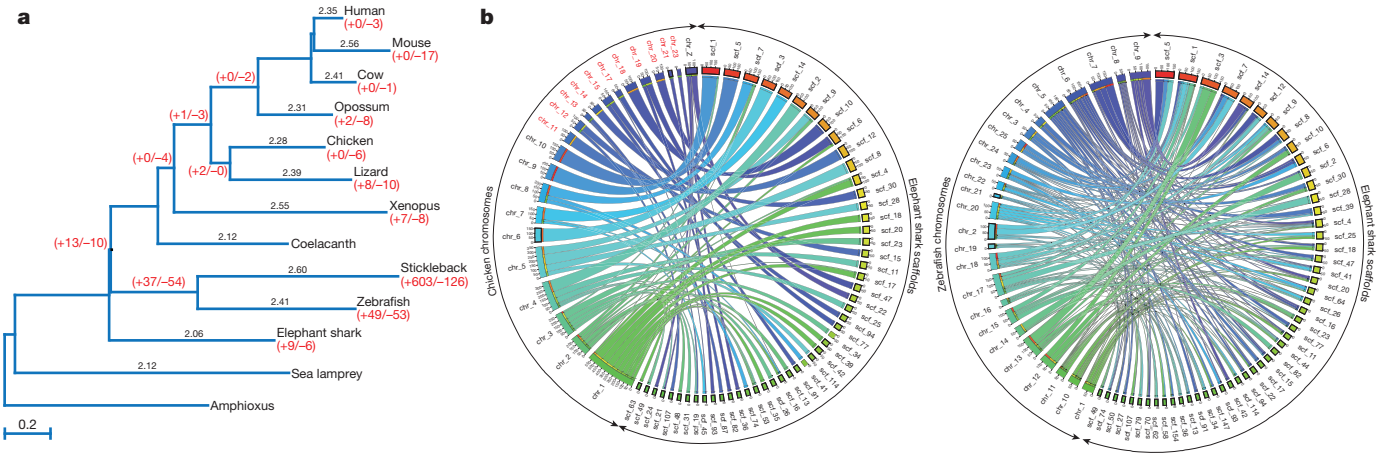



**Figure 2 | *Callorhinchus milii* possesses the slowest evolving vertebrate genome.** **a**, Neutral tree of 13 chordates based on fourfold degenerate sites. Pairwise distances to amphioxus are shown for each species above their respective branches. Intron gain/loss (+/−) events are shown in red below taxon labels and at ancestral nodes. **b**, Circos plots showing syntenic relationships between *C. milii* scaffolds and chicken and zebrafish chromosomes. Most *C. milii* scaffolds show a one-to-one relation with chicken chromosomes whereas they show a correspondence of one-to-two, or more, with zebrafish chromosomes. Chicken microchromosomes are labelled in red.

apparatus of extinct jawless fishes, the conodonts[21]. Early dermal bone was found in extinct jawless vertebrates such as heterostracans, whereas perichondral bone surrounding the cartilage was found in several fossil jawless vertebrates (osteostracans, galeaspids) and jawed vertebrates[5] (placoderms, acanthodians; Fig. 3). However, the highly complex process of endochondral ossification is unique to bony vertebrates. The *C. milii* genome sequence provided a unique opportunity to address the question of why the endoskeleton of chondrichthyans is not ossified.

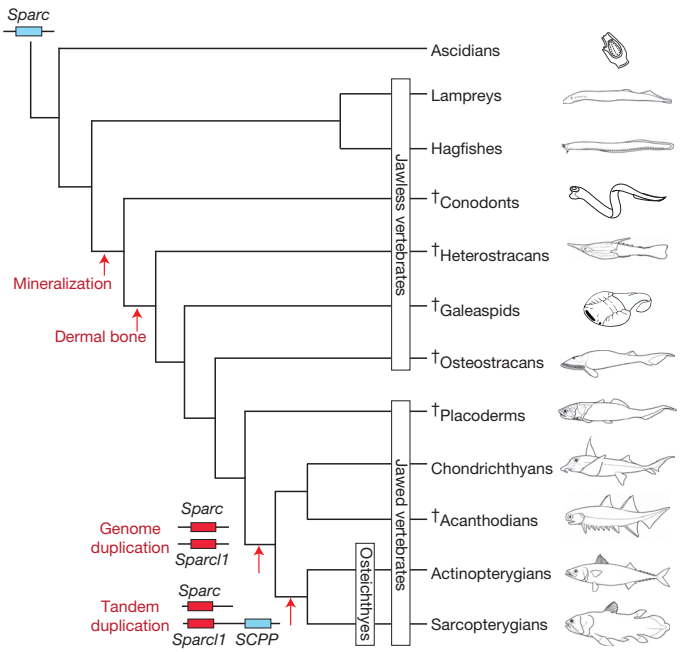



**Figure 3 | Genetic events underlying the emergence of bone formation in vertebrates.** Duplication of *Sparc* by whole-genome duplication initially gave rise to *Sparcl1*, and the subsequent tandem duplication of *Sparcl1* gave rise to the SCPP gene family responsible for endochondral ossification. Because the sea lamprey genome contains only *Sparc* but no *Sparcl1* (ref. 22), we have placed the genome duplication event that gave rise to *Sparcl1* after the divergence of jawless vertebrates from the jawed vertebrate ancestor. The sister relationship of chondrichthyans and acanthodians is based on ref. 37. *SCPP*, secretory calcium-binding phosphoprotein gene family member. †, extinct.

We searched the *C. milii* genome assembly and transcriptomes for genes known to be involved in bone formation in osteichthyans (Supplementary Note X). All gene family members involved in bone formation were present, except the secretory calcium-binding phosphoprotein (SCPP) gene family (Supplementary Table X.1). This gene family encodes a diverse array of secreted phosphoproteins that arose from the gene *Sparc-like 1* (*Sparcl1*) through tandem duplication, and *Sparcl1* itself arose from an ancient metazoan gene, *Sparc*, through whole-genome duplication[22]. There are two main categories of SCPP genes: one group encodes acidic proteins and the other encodes proline- and glutamine-rich (P/Q-rich) proteins. In the human genome, the two groups are found in two different clusters on chromosome 4; the acidic SCPP genes (*SPP1*, *MEPE*, *IBSP*, *DMP1* and *DSPP*, collectively known as SIBLING genes) occur between *PKD2* and *SPARCL1*, whereas the P/Q-rich SCPP genes are found in the enamel matrix protein-SCPP cluster ~17 megabases downstream of *SPARCL1* (Supplementary Fig. X.2). Acidic SCPP or SIBLING genes are involved in the ossification of collagenous matrix in bone and dentine, and P/Q-rich SCPP genes are involved in the production of enamel, milk, tears and saliva. Although there are variable numbers of P/Q-rich SCPP genes in teleosts[23], there is a single SIBLING gene, *spp1*, in zebrafish and medaka. Zebrafish *spp1* (also known as *osteopontin*) is expressed specifically in osteoblasts[24] and has therefore been proposed to have a primary function in bone formation similar to its mammalian orthologue[23].

The *C. milii* genome contains both *Sparc* and *Sparcl1* on different scaffolds that show extensive conserved synteny with orthologous loci in human and other bony vertebrates (Supplementary Figs X.1 and X.2). However, there is no SCPP gene cluster in the intergenic region between *Pkd2* and *Sparcl1* or elsewhere in the genome (Supplementary Fig. X.2). The genomic and transcriptomic resources available for other cartilaginous fishes such as the little skate (*Leucoraja erinacea*) and the small-spotted catshark (*Scyliorhinus canicula*) as well as the genome assembly of the sea lamprey, a jawless vertebrate, also do not contain any SCPP genes (Supplementary Note X). These findings suggest that the tandem duplication of *Sparcl1* that gave rise to SCPP genes occurred in the common ancestor of osteichthyans after this lineage split from the chondrichthyan lineage (Fig. 3). Because SCPP genes have a crucial role in the formation of bone, we propose that their absence from *C. milii* explains the absence of bone from the endoskeleton of cartilaginous fishes.

To test this hypothesis, we used two different methods to disrupt the function of the single bone-specific SIBLING gene *spp1* in zebrafish.

The knockdown of *spp1* using two gene-specific morpholinos (ATG MO and E2-I2 MO) resulted in a significant reduction in endochondral and dermal bone formation by comparison with embryos injected with 5-base-pair-mismatch control morpholinos (Supplementary Fig. X.4). Unlike the transient effects exerted by morpholinos, the genetic interference afforded by the CRISPR/Cas9 system[25] results in heritable genomic modifications; indeed, targeting exons 6 and 7 of the *spp1* gene using the CRISPR/Cas9 system resulted in the generation of specific insertion/deletion mutations at the target sites, including a ~2.6-kilobase deletion when exons 6 and 7 were simultaneously targeted (Supplementary Fig. X.6). Embryos 5 days post-fertilization (dpf) with deletions in exon 7 alone or in both exon 6 and exon 7 of *spp1* showed a significant reduction in the formation of endochondral and dermal bone (Fig. 4), with the defect in bone formation persisting in 15-dpf mutant embryos (Supplementary Fig. X.9). The similar phenotypes obtained using two different methods of manipulation indicate that the effects on bone formation are specific, and strongly suggest that *spp1* has an essential role in the modulation of bone formation in zebrafish.

The results of the zebrafish *spp1* knockdown experiments provide strong support for our hypothesis that the absence of SCPP genes from cartilaginous fishes is related to the unossified nature of their endoskeleton. In turn, the absence of SCPP genes from chondrichthyans raises questions about the genetic basis of dermal and perichondral bone formation in chondrichthyans, placoderms, acanthodians and extinct jawless vertebrates. We speculate that one or more SCPP-related genes, probably *Sparc*, *Sparcl1* or both, mediate the mineralization of skeleton in these vertebrates.

## Primordial adaptive immune system

The chondrichthyan immune system shares many features of the innate and adaptive immune systems of mammals[9,26]. However, several important differences, confirmed by transcriptome analysis of an elasmobranch cartilaginous fish, the nurse shark (*Ginglymostoma cirratum*), which diverged from the *C. milii* lineage ~420 Myr ago[6], highlight several unexpected features of the primordial state of gnathostome immune systems, especially for adaptive immunity (Supplementary Note XI, Supplementary Figs XI.1–XI.10 and Supplementary Tables XI.1–XI.13). First, the genome assembly suggests a close linkage of immunoglobulin and T-cell receptor genes, compatible with the notion that the somatically diversifying antigen receptor genes evolved from a common ancestor via *en bloc* duplications[27]. Indeed, the presence of the variant antigen receptor NAR–TCR, but lack of other specialized forms of antigen receptors, such as IgW and IgNAR in the *C. milii* genome suggest that the single-domain V region, subjected to somatic diversification by Rag proteins, was at first part of a T-cell receptor (TCR)-like structure before being co-opted by immunoglobulins. Second, the linkage of antigen receptor genes with certain MHC genes, whose products functionally interact in regulating the immune response, supports the co-evolutionary origin of antigen presentation and recognition[28,29]. Third, the immunogenome of *C. milii* is compatible with the presence of cytotoxic natural-killer and $CD8^+$ T cells; by contrast, the absence of the canonical CD4 co-receptor, the transcription factor RORC, several cytokines and cytokine receptors that are associated with helper and regulatory functions of $CD4^+$ T cells in mammals (Fig. 5) suggests the presence of a primordial type of helper function in cartilaginous fishes. Restricted helper functions (exemplified by the lack of the T-follicular helper lineage) in cartilaginous fishes may explain the long lag time required to generate humoral immunity (affinity maturation and memory) in sharks[30]. Thus, the emerging adaptive immune system seems to have been characterized by the presence of a full-blown cytotoxic system and a primordial helper system that was geared towards a $T_H1$-type response.

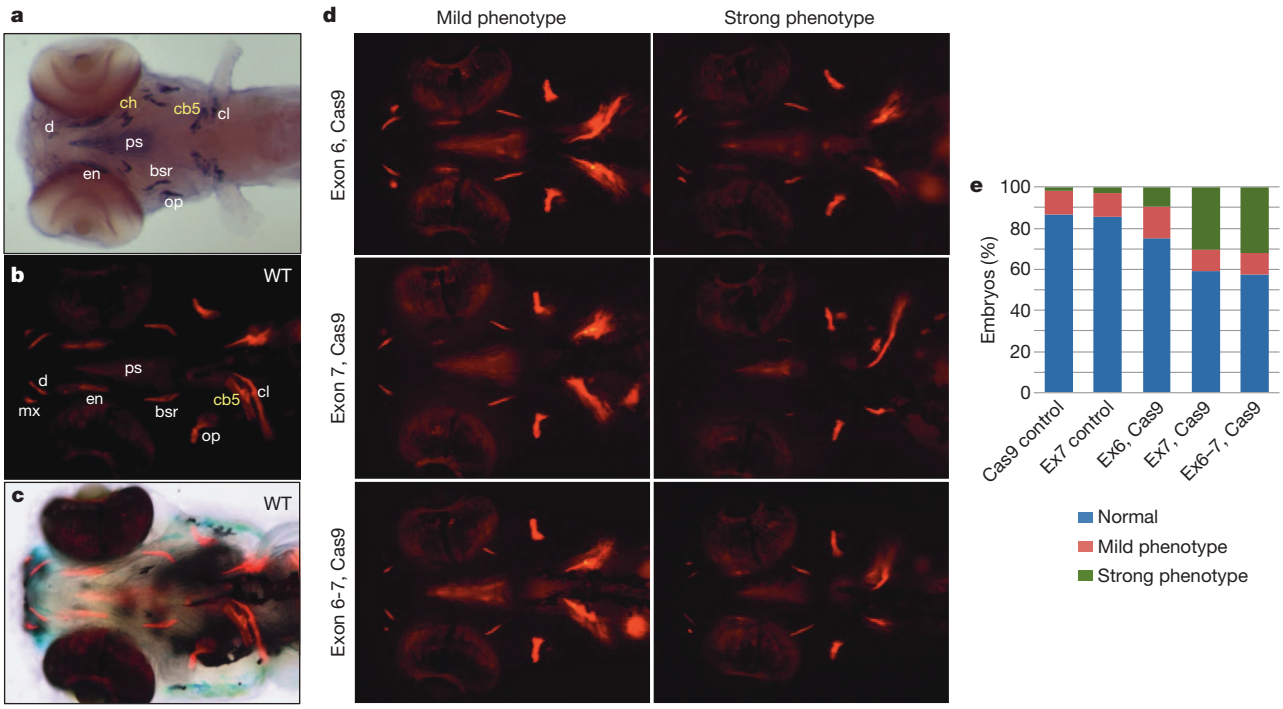



**Figure 4 | Targeted mutagenesis of zebrafish *spp1* by sgRNA:Cas9 results in reduced bone formation.** **a**, *spp1* is specifically expressed in cells surrounding the bone matrix. Ventral view of a 5-dpf embryo hybridized with a *spp1*-specific RNA probe. Yellow labels, endochondral bones (cb5, ceratobranchial 5; ch, ceratohyal); white labels, dermal bones (bsr, branchiostegal ray; cl, cleithrum; d, dentary; en, entopterygoid; op, operculum; ps, parasphenoid). **b**, Ventral view of a 5-dpf wild-type (WT) embryo stained with alizarin red to reveal sites of bone deposition (red fluorescence). mx, maxilla. **c**, Bright-field image merged with **b** to visualize anatomical structures and locations of bone deposition simultaneously. **d**, Ventral views of 5-dpf embryos injected with *Cas9* mRNA together with single guide RNA (sgRNA) targeting *spp1* exon 6, exon 7 or both (alizarin red staining). The embryos were scored as normal (resembling wild type), mild or strong bone phenotypes, with the latter showing the greatest reduction in bone formation. The variations in the extent of bone reduction are presumably due to somatic chimaerism with regard to *spp1* disruption. **e**, Proportions of mild and strong bone phenotypes induced by disruption of *spp1* by sgRNA/Cas9. Targeting of exon 7 ($n = 206$ embryos) or both exons 6 and 7 ($n = 143$) resulted in significantly higher proportions of strong bone phenotype ($P < 0.01$, Fisher's exact test) compared with control injections of *Cas9* mRNA ($n = 190$) and exon 7 sgRNA ($n = 143$) (Ex6, Cas9: $n = 72$).

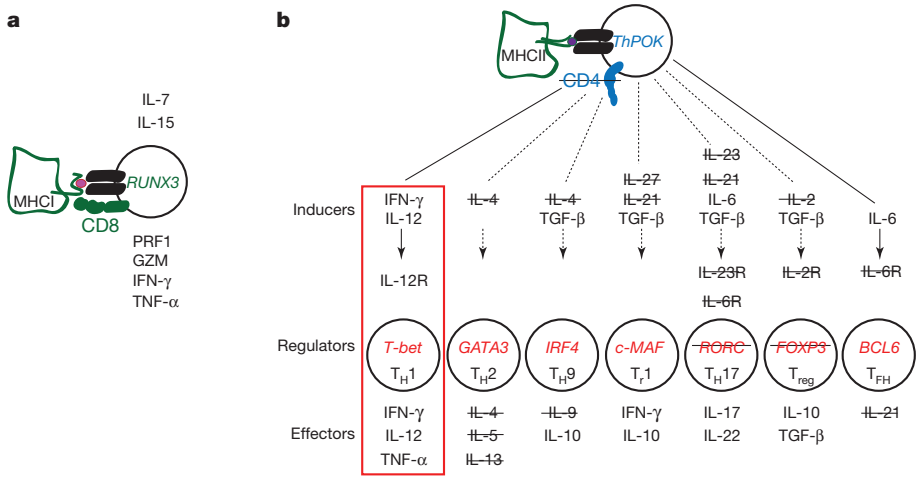



**Figure 5 | T-cell lineages in mammals and cartilaginous fishes.** **a**, Schematic diagram of mammalian $CD8^+$ cytotoxic cells producing various effector molecules, such as PRF1 (perforin), GZM (granzymes), INF-γ (interferon-γ) and TNF-α (tumour necrosis factor-α). Interleukins 7 (IL-7) and 15 (IL-15) are indispensable for induction of the transcription factor RUNX3, which controls expression of the genes encoding the CD8 co-receptor that interacts with the TCR and MHC class I molecules. All key elements of this T-cell lineage are present in the genomes of cartilaginous fishes. **b**, Diversification of CD4 lineages in mammals. The transcription factor ThPOK controls expression of the co-receptor CD4, which interacts with the TCR and MHC class II molecules and represses the CD8 lineage programme. The listed inducer cytokines activate key transcriptional regulators that define the multiple $CD4^+$ lineages specialized for producing various effector molecules; genes not found in cartilaginous fishes are struck through. Cartilaginous fishes probably possess only $T_H1$ cells (red rectangle) and lack other helper subsets. TGF-β, transforming growth factor-β.

Despite the apparent lack of $T_{reg}$ cells in cartilaginous fish, the elimination of self-reactive lymphocytes during the process of central tolerance seems to be a primordial vertebrate feature, as demonstrated by the presence of AIRE, which is responsible for the expression of peripheral self-antigens for presentation via MHC to T cells in the medulla of the thymus[31]. The absence of the gene encoding a bona fide CD4 co-receptor despite the presence of polymorphic MHC class II genes suggests the presence of unusual types of $CD8^-$ helper T cells, capable of exerting $T_H1$-like activities through IFN-γ, IL-12 and TNF-α (Fig. 5). We suggest that such cells are capable of recognizing exogenous antigens presented by MHC class II molecules but might also interact with other antigens in an MHC-independent, antibody-like recognition mode, as recently demonstrated in mice deficient in both CD8 and CD4 co-receptors[32]. It is to be noted that the secondary loss of CD4, MHC class II and invariant chain genes in a teleost, the Atlantic cod (*Gadus morhua*), is accompanied by compensatory changes, such as amplification of MHC class I and Toll-like receptor genes[33], most probably related to the disappearance of T-helper lineages only in this species. In stark contrast, the lack of such compensatory features in cartilaginous fishes, demonstrated in two highly divergent species, *C. milii* and *G. cirratum*, supports the hypothesis that the features described here are primordial. Our data also suggest that CD8 and CD4 were co-opted as co-receptors for the TCR at different times in evolution, and that T cells recognizing MHC class II molecules in cartilaginous fish recruit the downstream signalling components to the immunological synapse entirely through the TCR complex.

## Conclusion

Among the three living lineages of vertebrates (cyclostomes, cartilaginous fishes and bony vertebrates), bony vertebrates are the largest and most diverse group of vertebrates. Because cartilaginous fishes are the sister group of bony vertebrates, they constitute a critical outgroup for understanding the evolution and diversity of bony vertebrates. The whole-genome analysis of *C. milii*, a holocephalan cartilaginous fish, shows that the *C. milii* genome is evolving significantly slower than other vertebrates, including the coelacanth, which is considered a 'living fossil'. Although several physiological and environmental factors have been proposed to explain the interspecific variation in molecular evolutionary rates[34,35], the factors contributing to the lower evolutionary rate of *C. milii* are not known. Overall, the *C. milii* genome is the least derived among known vertebrates and is therefore a good model for inferring the state of the ancestral chondrichthyan and gnathostome genomes. Its value for comparative genomic studies is illustrated by our analysis of genetic events that led to the ossification of endoskeleton in bony vertebrates. Unexpected was our finding that the adaptive immune system of cartilaginous fishes possesses highly restricted subsets of T helper cells (perhaps only one) with unconventional antigen-binding properties; this suggests that helper and regulatory functions of T cells that recognize MHC class II molecules became more elaborate in the ancestor of bony vertebrates through the acquisition of transcription factors such as RORC, the CD4 co-receptor, conventional FOXP3 and a host of CD4-lineage-specific cytokines and cytokine receptors. Thus, the whole-genome analysis of *C. milii* provides fresh insight into the mechanism of bone formation and the origin of adaptive immunity of gnathostomes.

## METHODS SUMMARY

Genomic DNA was obtained from the testis of a single *C. milii* caught in Tasmania, Australia, and used to prepare libraries with inserts of different sizes. Sequencing was conducted on the Roche 454 GS FLX Titanium platform (for shotgun fragments and 3-kilobase and 8-kilobase inserts) and the ABI3730 instrument (for plasmid, fosmid and BAC ends). The *C. milii* genome was assembled using CABOG v6.1. PolyA-selected RNA from ten tissues of *C. milii* (brain, gills, heart, intestine, kidney, liver, muscle, ovary, spleen and testis) and the spleen and thymus of *G. cirratum* were sequenced on the Illumina GAIIx platform. Transcripts were assembled *de novo* using Trinity r2011-07-13. MicroRNA genes were identified by deep sequencing of small RNA from 16 tissues (brain, blood, eye, gills, heart, intestine, kidney, liver, muscle, ovary, pancreas, rectal gland, skin, spleen, testis and uterus) on the Illumina GAIIx platform. Annotation of the genome was carried out using the Ensembl gene annotation pipeline which integrated *ab initio* gene predictions and evidence-based gene models. The gene set can be viewed at http://esharkgenome.imcb.a-star.edu.sg/. Annotation of protein domains in the *C. milii* proteome was carried out by searching it against the Pfam v26 database using HMMER v3.0. Zebrafish *spp1* was knocked down using morpholinos or the CRISPR/Cas9 system. Methods are described in detail in individual sections of the Supplementary Information.

All animals were cared for in strict accordance with National Institutes of Health (USA) guidelines. The protocol was approved by the Institutional Animal Care and Use Committee of the Biological Resource Centre, A*STAR (protocol #100520).

**Received 22 May; accepted 1 November 2013.**

1. Shimeld, S. M. & Holland, P. W. Vertebrate innovations. *Proc. Natl Acad. Sci. USA* **97,** 4449–4452 (2000).
2. Gans, C. & Northcutt, R. G. Neural crest and the origin of vertebrates: a new head. *Science* **220,** 268–273 (1983).

3. Young, G. C. Placoderms (armored fish): dominant vertebrates of the Devonian period. *Annu. Rev. Earth Planet. Sci.* **38,** 523–550 (2010).
4. Boehm, T. Evolution of vertebrate immunity. *Curr. Biol.* **22,** R722–R732 (2012).
5. Janvier, P. *Early Vertebrates* (Oxford Univ. Press, 1996).
6. Inoue, J. G. *et al.* Evolutionary origin and phylogeny of the modern holocephalans (Chondrichthyes: Chimaeriformes): a mitogenomic perspective. *Mol. Biol. Evol.* **27,** 2576–2586 (2010).
7. Smith, J. J. *et al.* Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nature Genet.* **45,** 415–421 (2013).
8. Venkatesh, B., Tay, A., Dandona, N., Patil, J. G. & Brenner, S. A compact cartilaginous fish model genome. *Curr. Biol.* **15,** R82–R83 (2005).
9. Venkatesh, B. *et al.* Survey sequencing and comparative analysis of the elephant shark *(Callorhinchus milii)* genome. *PLoS Biol.* **5,** e101 (2007).
10. Nelson, J. S. *Fishes of the World* 4th edn (Wiley, 2006).
11. Last, P. R. & Stevens, J. D. in *Sharks and Rays of Australia* 465–483 (CSIRO Australia, 1994).
12. Flicek, P. *et al.* Ensembl 2012. *Nucleic Acids Res.* **40,** D84–D90 (2012).
13. Rasmussen, A. S. & Arnason, U. Molecular studies suggest that cartilaginous fishes have a terminal position in the piscine tree. *Proc. Natl Acad. Sci. USA* **96,** 2177–2182 (1999).
14. Blair, J. E. & Hedges, S. B. Molecular phylogeny and divergence times of deuterostome animals. *Mol. Biol. Evol.* **22,** 2275–2284 (2005).
15. Hallstrom, B. M. & Janke, A. Gnathostome phylogenomics utilizing lungfish EST sequences. *Mol. Biol. Evol.* **26,** 463–471 (2009).
16. Martin, A. P., Naylor, G. J. & Palumbi, S. R. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature* **357,** 153–155 (1992).
17. Martin, A. P. Substitution rates of organelle and nuclear genes in sharks: implicating metabolic rate (again). *Mol. Biol. Evol.* **16,** 996–1002 (1999).
18. Amemiya, C. T. *et al.* The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496,** 311–316 (2013).
19. Burt, D. W. Origin and evolution of avian microchromosomes. *Cytogenet. Genome Res.* **96,** 97–112 (2002).
20. Ellegren, H. Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol. Evol.* **25,** 283–291 (2010).
21. Donoghue, P. C., Sansom, I. J. & Downs, J. P. Early evolution of vertebrate skeletal tissues and cellular interactions, and the canalization of skeletal development. *J. Exp. Zoolog. B* **306B,** 278–294 (2006).
22. Kawasaki, K., Buchanan, A. V. & Weiss, K. M. Gene duplication and the evolution of vertebrate skeletal mineralization. *Cells Tissues Organs* **186,** 7–24 (2007).
23. Kawasaki, K. The SCPP gene repertoire in bony vertebrates and graded differences in mineralized tissues. *Dev. Genes Evol.* **219,** 147–157 (2009).
24. Laue, K., Jänicke, M., Plaster, N., Sonntag, C. & Hammerschmidt, M. Restriction of retinoic acid activity by Cyp26b1 is required for proper timing and patterning of osteogenesis during zebrafish development. *Development* **135,** 3775–3787 (2008).
25. Altekar, G., Dwarkadas, S., Huelsenbeck, J. P. & Ronquist, F. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20,** 407–415 (2004).
26. Dooley, H. & Flajnik, M. F. Antibody repertoire development in cartilaginous fish. *Dev. Comp. Immunol.* **30,** 43–56 (2006).
27. Glusman, G. *et al.* Comparative genomics of the human and mouse T cell receptor loci. *Immunity* **15,** 337–349 (2001).
28. Du Pasquier, L., Zucchetti, I. & De Santis, R. Immunoglobulin superfamily receptors in protochordates: before RAG time. *Immunol. Rev.* **198,** 233–248 (2004).
29. Flajnik, M. F., Tlapakova, T., Criscitiello, M. F., Krylov, V. & Ohta, Y. Evolution of the B7 family: co-evolution of B7H6 and NKp30, identification of a new B7 family member, B7H7, and of B7's historical relationship with the MHC. *Immunogenetics* **64,** 571–590 (2012).
30. Dooley, H. & Flajnik, M. F. Shark immunity bites back: affinity maturation and memory response in the nurse shark, *Ginglymostoma cirratum*. *Eur. J. Immunol.* **35,** 936–945 (2005).
31. Anderson, M. S. *et al.* Projection of an immunological self shadow within the thymus by the aire protein. *Science* **298,** 1395–1401 (2002).
32. Tikhonova, A. N. *et al.* αβ T cell receptors that do not undergo major histocompatibility complex-specific thymic selection possess antibody-like recognition specificities. *Immunity* **36,** 79–91 (2012).
33. Star, B. *et al.* The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477,** 207–210 (2011).
34. Martin, A. P. & Palumbi, S. R. Body size, metabolic rate, generation time, and the molecular clock. *Proc. Natl Acad. Sci. USA* **90,** 4087–4091 (1993).
35. Schlötterer, C., Amos, B. & Tautz, D. Conservation of polymorphic simple sequence loci in cetacean species. *Nature* **354,** 63–65 (1991).
36. Janvier, P. Palaeontology: modern look for ancient lamprey. *Nature* **443,** 921–924 (2006).
37. Zhu, M. *et al.* A Silurian placoderm with osteichthyan-like marginal jaw bones. *Nature* **502,** 188–193 (2013).


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

**Acknowledgements** Generation of *C. milii* genome sequence at the Genome Institute at Washington University was supported by grants from the National Human Genome Research Institute, USA. We thank members of the Production Sequencing group at the Genome Institute at Washington University. Sequencing of messenger RNA, small RNA and BAC ends was supported by grants from the Biomedical Research Council of A*STAR, Singapore. This work was supported by the A*STAR Computational Resource Centre through the use of its high-performance computing facilities. We would like to thank J. Danks, J. Bell and J. G. Patil for their help in collecting *C. milii* samples, and J. K. Joung for CRISPR and Cas9 plasmids. We also thank the following funding agencies: the Max Planck Society (T.B.); NIH grants RR006603 and AI27877 (M.F.F.); the Ministry of Education, Culture, Sports, Science and Technology, Japan (M.K.); the Human Frontiers Science Program Organization (M.I.); ERC Starting Grant (260372) and MICINN (Spain) BFU2011-28549 (T.M.-B.); and the Biomedical Research Council of A*STAR, Singapore (B.V., P.W.I., S. Hoon and V.K.).

**Author Contributions** B.V., R.K.W., S.B. and W.C.W. planned and oversaw the project. P.M., L.W.H. and W.C.W. sequenced and assembled the genome, and evaluated assembly coverage and quality. A.P.L., M.M.L. and B.V. annotated the genome. A.P.L. characterized repetitive sequences, GC content, isochores and CNEs. B.J.R. generated whole-genome alignments. B.L.-G., J.Q. and T.M.-B. analysed gene duplication and heterozygosity. S. Hoon, V.G., K.W.K. and S. Ho generated RNA-seq data and analysed miRNA families and protein domains. B.-H.T., S.T. and A.T. sequenced BAC ends and RNA-seq. M.M.L. and A.P.L. assembled and annotated RNA-seq data. V.R. performed phylogenomic and molecular evolutionary rate analyses. S.W.R. and M.I. investigated intron evolution. A.K.M., A.P.L., V.R., P.W.I. and B.V. analysed genes involved in bone formation. A.K.M. and Z.W.L. carried out functional analysis of zebrafish *spp1*. T.B., J.B.S., Y.O., M.F.F., Y.S., M.K., A.P.L., M.M.L. and B.V. analysed the immunogenome. V.K., V.R. and I.K. analysed evolution of protein-coding gene families. M.M.L., A.P.L. and B.V. analysed large-scale synteny. B.V., T.B. and W.C.W. wrote the manuscript with input from other authors.

**Author Information** The *C. milii* genome assembly has been deposited at DDBJ/EMBL/GenBank under the accession number AAVX00000000. The version described in this paper is the second version, AAVX02000000. The *C. milii* and *G. cirratum* RNA-seq data have been deposited at SRA under accession numbers SRA054255 and SRA062964, respectively. The transcripts have been deposited into GenBank under accession numbers JW861113–JW881738, KA353634–KA353668; and the miRNA sequences under accession numbers JX994303–JX994995. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.V. (mcbbv@imcb.a-star.edu.sg) or W.C.W. (wwarren@genome.wustl.edu).

ARTICLE

doi:10.1038/nature12879

# Patterning and growth control by membrane-tethered Wingless

Cyrille Alexandre[1]*, Alberto Baena-Lopez[1]* & Jean-Paul Vincent[1]

**Wnts are evolutionarily conserved secreted signalling proteins that, in various developmental contexts, spread from their site of synthesis to form a gradient and activate target-gene expression at a distance. However, the requirement for Wnts to spread has never been directly tested. Here we used genome engineering to replace the endogenous *wingless* gene, which encodes the main *Drosophila* Wnt, with one that expresses a membrane-tethered form of the protein. Surprisingly, the resulting flies were viable and produced normally patterned appendages of nearly the right size, albeit with a delay. We show that, in the prospective wing, prolonged *wingless* transcription followed by memory of earlier signalling allows persistent expression of relevant target genes. We suggest therefore that the spread of Wingless is dispensable for patterning and growth even though it probably contributes to increasing cell proliferation.**

Wnts are secreted signalling proteins[1,2] that have been suggested to act at a distance to control patterning and growth during development[3–6]. Long-range Wnt activity has been most extensively studied in wing imaginal discs of *Drosophila*[4–6]. These epithelial pockets, set aside in the embryo, grow from approximately 50 to 50,000 cells during larval stages to give rise to fully patterned wings during pupariation[7–9]. During early larval life, Wingless (Wg), the main *Drosophila* Wnt, is initially expressed throughout the prospective wing field to establish wing primordium[10–14]. At subsequent stages, Wg is produced in a narrow stripe of cells at the dorsoventral boundary. It is thought that from there it spreads throughout the prospective wing blade and activates target-gene expression in a concentration-dependent manner[4–6]. A common view is that, near the dorsoventral boundary, high signalling activity activates *senseless* expression, which contributes to wing margin fates[15–17],whereas further away, in the prospective wing blade (up to 50 cell diameters), low-level signalling stimulates the expression of more sensitive target genes like *vestigial, Distal-less* and *frizzled 3* (refs 4, 5, 18, 19) and promotes growth[20,21]. Although the importance of graded Wg signalling has been disputed[21], it is generally accepted that Wg needs to spread over the whole wing field for patterning and growth. To test this assumption directly, we sought to modify the *wingless* (*wg*) locus so that membrane-tethered Wg would be expressed at physiological level in the absence of the wild-type form.

## Patterning and growth by membrane-tethered Wingless

A fusion protein comprising the type 2 transmembrane protein Neurotactin and Wg (NRT–Wg) has been previously shown in clonal overexpression assays to activate target genes only within expressing cells and in surrounding adjoining cells[4,22,23]. As a first step towards introducing a complementary DNA encoding NRT–Wg into the *wg* locus, we used an improved protocol for homologous recombination[24] to delete the wild-type ATG-containing exon and replace it with a cassette that includes an *attP* recombination site, a mCherry-encoding cDNA, and other elements (Extended Data Fig. 1a). As expected, in heterozygous larvae, Cherry was expressed in a pattern that broadly recapitulated that of endogenous *wg* (Extended Data Fig. 1b). Most extraneous genetic elements were then removed, leaving only the *attP* site and a single *LoxP* site at the locus (Extended Data Fig. 1a). The resulting allele behaved as a null and is therefore referred to hereafter as *wg* knockout (*wg*{*KO*}) (Extended Data Fig. 1a). A wild-type *Wg* cDNA was inserted into the *attP* site, along with mini-*white* as a genetic marker, and the resulting flies (*wg*{*KO; Wg*}) were indistinguishable from wild-type flies, confirming faithful expression of the reintegrated cDNA (Extended Data Fig. 1d, f, g). Next, a cDNA encoding haemagglutinin-tagged NRT–Wg was introduced at the targeted locus, using either mini-*white* or *pax*-Cherry as a genetic marker, to generate *wg*{*KO; NRT–Wg*} (Extended Data Fig. 1e; see Methods for details about genetic marker usage throughout this manuscript). Notably, when grown without crowding, homozygous animals that have NRT–Wg as their sole source of Wg were viable and had normally patterned appendages and other cuticular structures (Fig. 1a, b and Extended Data Fig. 1f–o). The wings were slightly smaller than those of control siblings (10–12% surface area reduction; Extended Data Fig. 1f–i). However, the arrangement of margin sensory bristles appeared normal (as shown at high magnification in Extended Data Fig. 1j–l), an indication of appropriate *senseless* expression during imaginal development. Other target genes, *vestigial* (not shown), *Distal-less* (immunofluorescence) and *frizzled 3* (from a gene trap) were mildly affected, with their expression dropping off more sharply than in controls but without a noticeable consequence on adult wing patterning (Fig. 1b). Importantly, these reporters of Wg signalling continued to be expressed in a broad domain straddling the dorsoventral boundary (Fig. 1c–f).

## Signalling by Neurotactin–Wingless is juxtacrine

A key question is how NRT–Wg might activate target genes seemingly at a long range. Staining of homozygous *wg*{*KO; NRT–Wg*} imaginal discs with anti-Wg showed a restricted distribution by comparison to that in control discs although weak staining could be detected at the surface of cells located on either side of the main expression domain (Fig. 1g, h). To assess whether this could be due to protein perdurance or release (for example, by cleavage or on exosomes), we devised a system for timed removal of the NRT–Wg cDNA. A *wg*{*KO; FRT NRT–Wg FRT*} allele was created (FRT, Flp recombinase target) and combined with *hedgehog-gal4, UAS-flp* and *tubulin-gal80$^{ts}$*. Transferring the resulting larvae from 18 °C to 29 °C triggered excision of the *FRT NRT–Wg FRT* cassette in patches of cells within the P compartment. No anti-haemagglutinin immunoreactivity (that is,

[1]MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.
*These authors contributed equally to this work.

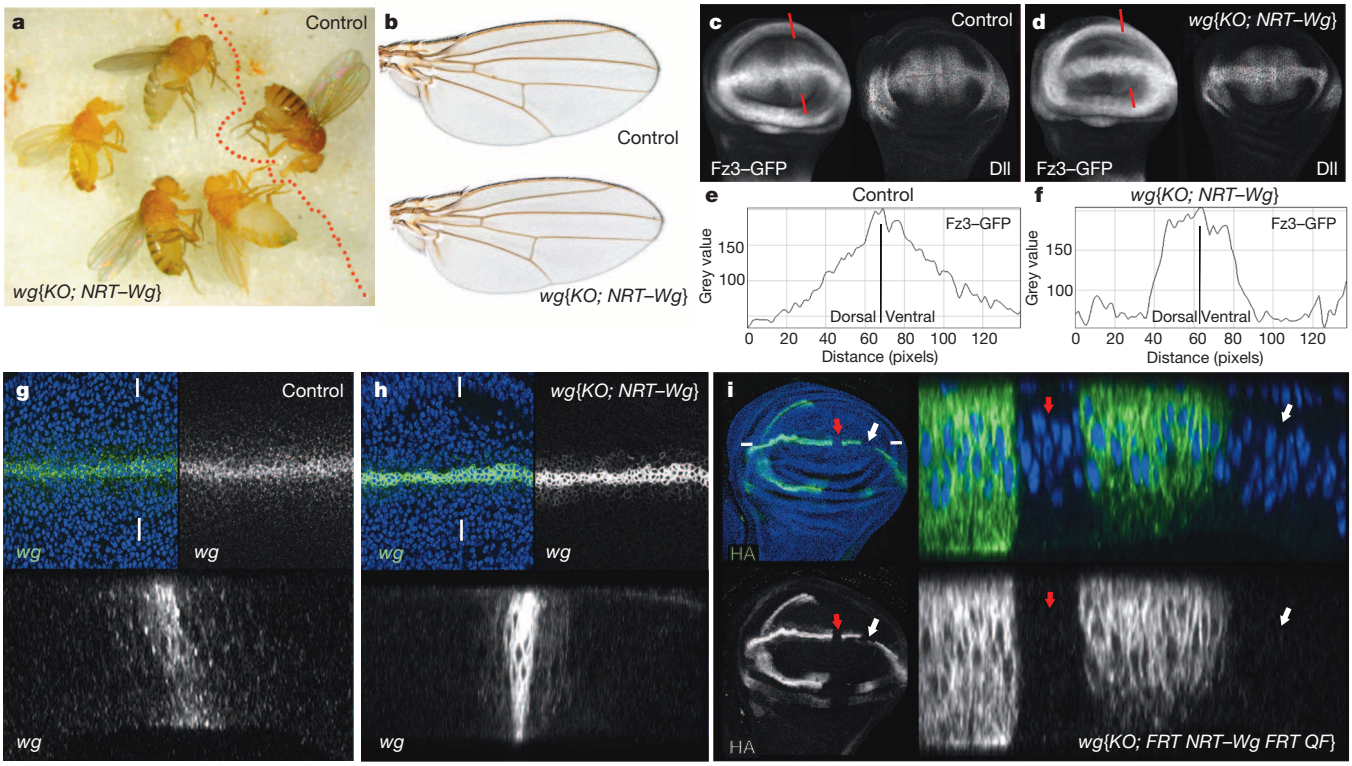



**Figure 1 | Characterization of membrane-tethered Wingless expressed from the *wingless* locus. a**, Homozygous *wg{KO; NRT–Wg}* flies and one control (*wg{KO; NRT–Wg}/GlaBC*) fly. **b**, Wings from a control (*wg{KO; NRT–Wg}/GlaBC*, and a homozygous *wg{KO; NRT–Wg}*fly (see size information and details of the wing margin in Extended Data Fig. 1i–l). **c**, **d**, Expression of Frizzled-3–GFP and Distal-less (Dll) in control and homozygous *wg{KO; NRT–Wg}* imaginal discs. **e**, **f**, Fluorescence intensity profile (grey value; arbitrary units) along lines connecting red arrowheads. **g**, **h**, Wg immunoreactivity in a control (wild type) and a homozygous *wg{KO; NRT–Wg}*disc, with corresponding transverse sections (along line connecting white marks). **i**, Excision of haemagglutinin (HA)-tagged NRT–Wg (red and white arrows) from *wg{KO; FRT NRT–Wg FRT QF}* with Flp produced from *hedgehog-gal4 tubulin-gal80$^{ts}$ UAS-flp* at 29 °C. Low-magnification frontal view and high-magnification transverse section along line defined by white marks are shown.

NRT–Wg) could be detected in the patches 12 h after the temperature shift (red and white arrows in Fig. 1i). As residual Gal80$^{ts}$ activity perdures as late as 5 h after shifting to 29 °C (ref. 25), we estimate that NRT–Wg persists for less than 7 h and is therefore not exceptionally stable. Importantly, lack of anti-haemagglutinin staining within the patches also shows that no detectable Wg is released from NRT–Wg-expressing cells. The range of NRT–Wg was functionally assessed by analysing the expression of *wg* target genes (*senseless* and *Distal-less*) in *wg*-null mutant clones generated in a *wg{KO; NRT–Wg}* background ('null-in-NRT' clones). For comparison, *wg* null mutant clones were also induced in a wild-type background ('null-in-WT' clones). In both cases, *senseless* was activated within one cell diameter from the clone edge (Extended Data Fig. 2a, b). Therefore, NRT–Wg achieves the level of signalling required to activate *senseless* expression. Next we assessed the range of NRT–Wg with *Distal-less*, a target gene thought to require lower signalling activity. Null-in-NRT clones located in region 1 (clones located elsewhere will be addressed later) failed to maintain *Distal-less* expression except in adjoining cells (Fig. 2b, red arrows). By contrast, *Distal-less* expression was maintained throughout similarly sized and positioned null-in-WT clones (Fig. 2a, turquoise arrowhead). These results confirm that wild-type Wg can spread and signal over a few cell diameters[4] and that, in region 1, Wg signalling is continuously required for high *Distal-less* expression. Importantly, they also demonstrate that no active Wg or other functional Wnts are released from the NRT–Wg-expressing cells.

## Prolonged transcription and cellular memory

So far we have shown that NRT–Wg does not act beyond adjoining cells and yet, extensive genetic analysis has shown that Wg signalling is required within the prospective wing blade, seemingly far from the source of Wg[4,5,21,26]. The combined effect of two processes provides a solution to this paradox. One such process was revealed by careful analysis of the pattern of *wg* transcription. This was facilitated by inserting a Gal4-encoding cDNA into the *attP* site of the *wg{KO}* allele, and crossing the resulting flies (*wg{KO; Gal4}*) to flies carrying UAS-GFP (Fig. 3a). GFP (green fluorescent protein) signal could be seen throughout the pouch until mid-third instar stage, suggesting that the 'Wg target cells' are indeed transcribing *wg* during this period of development. As the perdurance of GFP could lead to an overestimate of the time when transcription terminates, we devised a Flp-gated reporter of *wg* transcription, comprising *wg{KO; FRT Wg FRT QF}*, *hs-Flp* and *QUAS-Tomato* (Fig. 3b). In such larvae, excision of the *FRT Wg FRT* cassette by heat shock triggers expression of the heterologous transcription factor QF, and hence *QUAS-Tomato*, in cells that are transcribing *wg* at the time. As shown in Fig. 3b, heat shocks at early-third instar led to clones throughout the prospective wing, while clones induced later became spatially restricted. A similar behaviour was seen with another reporter designed to provide a permanent record of transcriptional activity at experimentally defined times (Extended Data Fig. 3b, c). To confirm directly that *wg* is transcriptionally active, imaginal discs were processed for fluorescence *in situ* hybridization (FISH) with a *wg* probe. Specific signal (using the prospective notum as a negative control) was seen in the prospective wing region as late as 84 ± 3 h after egg laying (Fig. 3c; see also ref. 13). Therefore, multiple experimental approaches showed that *wg* transcription is active throughout the wing pouch during a critical period of patterning and growth (see also ref. 14), but then recedes towards the dorsoventral boundary during the first half of third instar. Imaginal discs of larvae carrying *wg{KO; Gal4}*and

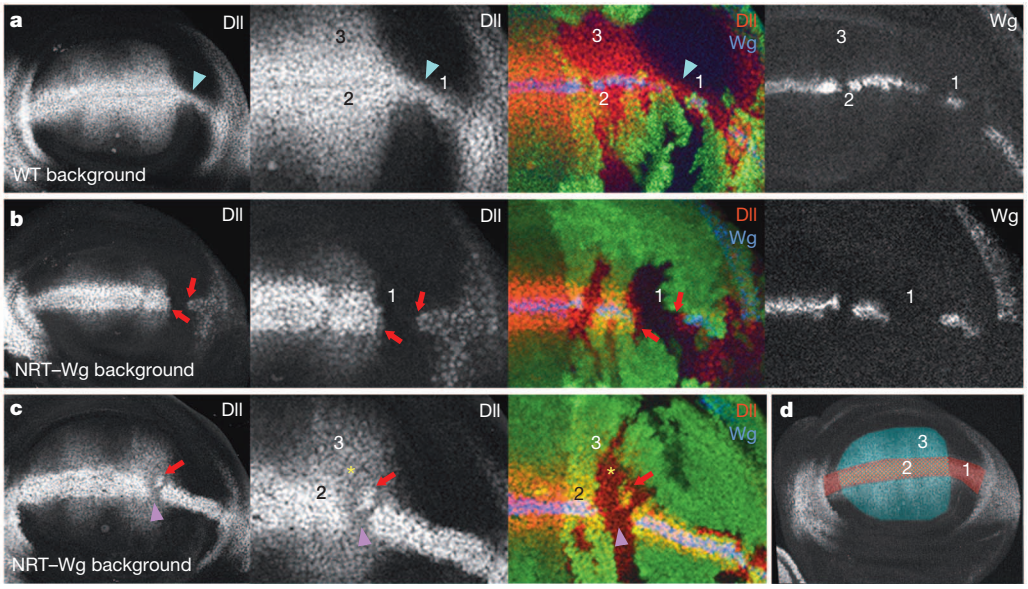



**Figure 2 | Gene expression in *wingless*-null mutant patches surrounded by wild-type or Neurotactin–Wingless-expressing cells. a–c**, Expression of Dll and Wg, as detected by immunofluorescence, in 'null-in-wt' (**a**) or 'null-in-NRT' (**b**, **c**) clones. Null *wg* mutant territory ($wg^{CX4}$) is marked by the absence of GFP. Red arrows indicate short-range activation of *Dll* expression by NRT–Wg. Turquoise arrowhead highlights longer range of wild type Wg. Purple arrowhead and yellow asterisk indicate persistent expression (memory) in zones 2 and 3, respectively. **d**, Diagram showing regions in which the behaviour of null mutant clones was analysed. Region 1 links the prospective wing to the prospective hinge. Region 2 represents the domain straddling the dorsoventral boundary that expresses high-level Dll. Region 3 represents the prospective wing blade, which expresses graded, low-level Dll. Age differences within the third instar account for disparities in disc size between **b** (younger) and **c** (older).

*UAS-HRP-CD8-GFP*, a particularly stable reporter[27], vividly illustrate that all the cells of the prospective wing blade derive from cells that express *wg* during early-third instar (Fig. 3d). The *wg* transcriptional program could provide a local source of low-level protein in the prospective wing blade during the first part of third instar.

Although *wg* transcription in the prospective blade extends later than previously thought, it does terminate around the mid-third instar. Therefore an additional mechanism is needed to sustain growth and patterning without releasable Wg beyond this stage. The nature of this mechanism is suggested from the behaviour of *vestigial* and *Distal-less*

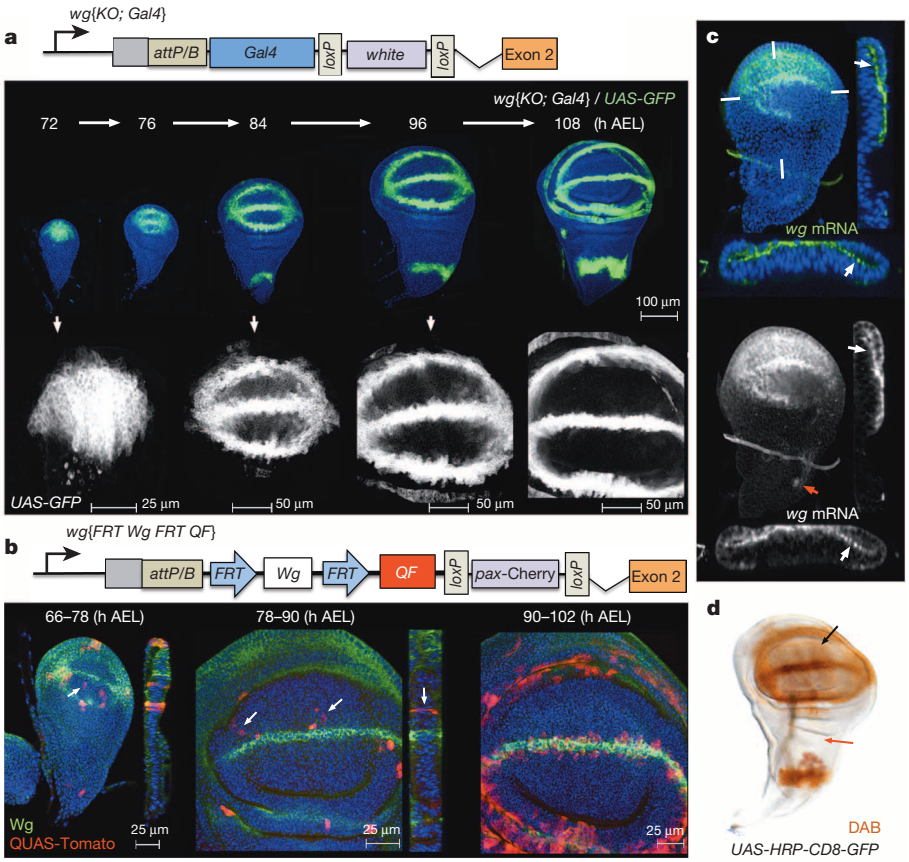



**Figure 3 | Activity of the *wingless* promoter in the prospective wing.** **a**, GFP expression driven by *wg{KO; Gal4}* at different stages. AEL, after egg laying. **b**, *QUAS-Tomato*-expressing clones in larvae of the genotype *hs-Flp QUAS-Tomato; wg{KO; FRT NRT–Wg FRT QF}/Cyo*. Larvae were heat shocked 18 h before fixation at the times indicated. White arrows indicate the presence of Tomato-expressing cells in the middle of the prospective wing blade, far from the prospective margin. **c**, FISH with a *wg* probe in wing imaginal discs obtained 84 ± 3 h AEL. Red arrow indicates incipient *wg* expression in prospective thorax. White marks indicate the plane of transverse section. White arrow shows FISH signal in the apical domain. **d**, HRP activity (detected by DAB staining) in a late-third instar *wg{KO; Gal4}/+, UAS-HRP-CD8-GFP* wing imaginal disc. Activity is detectable throughout the pouch (black arrow) but not in the prospective thorax (red arrow).

in response to changes in Wg signalling. As expected from classical target genes, these two genes are activated by ectopic signalling activity and their expression terminates in small patches of cells that are made unable to respond to Wg (for example, by genetic removal of the receptors)[4,5,22]. However, unlike classical target genes, their expression persists, albeit at a reduced level, when signalling is prematurely terminated throughout the disk or a compartment[21,26]. Such persistence can be seen for *Distal-less* in regions 2 and 3 of null-in-NRT mosaic discs. For example, in clones that transect the dorsoventral boundary in the central portion of the disc (region 2), *Distal-less* expression was reduced but not eliminated (for example, Fig. 2c, purple arrowhead). Note that high *Distal-less* continued to be expressed in null-in-NRT cells located at the edge of the clones, both in region 1 and in region 2 (Fig. 2b, c, red arrows). By contrast, high *Distal-less* expression was maintained throughout similarly sized null-in-WT clones (Fig. 2a). These observations further confirm that the NRT territory only triggers juxtacrine signalling. In region 3, further away from the dorsoventral boundary, *Distal-less* expression was essentially unchanged in 'null-in-NRT' clones (Fig. 2c, yellow asterisk), again confirming that *Distal-less* expression persists in this region even after removal of Wg signalling. Persistent target-gene expression within 'null-in-NRT' clones could explain their continued albeit slower (in comparison with null-in-WT) growth (Extended Data Fig. 2c–f). Memory of earlier signalling, which has been suggested previously for Dpp[28], could involve classical epigenetic control. Indeed, the expression of a *vestigial* reporter construct has been shown to be influenced by the presence or absence of Polycomb response elements[29]. Autoregulation could also contribute to the sustained expression of target genes[30]. Irrespective of the underlying mechanism, persistence of target-gene expression could explain, in part, why *wg{KO; NRT–Wg}* discs continue to grow beyond the time when *wg* transcription terminates in the prospective wing blade. It also suggests that during normal development, target-gene expression within the prospective blade (region 3) does not necessarily require Wg to spread from the dorsoventral boundary.

## Organ growth and developmental timing

Even though *wg{KO; NRT–Wg}* homozygous flies are morphologically normal, larvae of this genotype entered pupariation with a delay compared to controls (Fig. 4a). In addition, although they eclosed at a near normal frequency when cultured on their own, they largely failed to do so in co-culture with controls (Fig. 4b); a strong indication of reduced fitness. As imaginal disc damage can delay larval development[31–33], it is conceivable that preventing Wg release could affect developmental timing of the larva indirectly by slowing down disc development. To test this possibility, we developed a method for tissue-specific allele switching in second instar imaginal discs[24] (Extended Data Fig. 4a, b). Thus, we created larvae expressing NRT–Wg only in imaginal discs (*wg{KO; $WT^{Body}$; $NRT–Wg^{Discs}$}*; Fig. 4c, d). Larvae of this genotype developed at the same rate as controls (Extended Data Fig. 4c) and eclosed at near normal frequency when co-cultured with control siblings (Fig. 4e). By contrast, larvae expressing wild-type Wg in imaginal discs and NRT–Wg elsewhere *wg{KO; $NRT–Wg^{Body}$; $WT^{Discs}$}* (Fig. 4j, k) largely failed to eclose when co-cultured with control siblings even though they grew into normal-looking adults when grown in separate vials (Fig. 4l, n, p). These observations suggest that Wg release, or signalling activity not achieved by NRT–Wg, is required in tissues other than imaginal discs for timely development and general fitness. Nevertheless, further analysis of *wg{KO; $WT^{Body}$; $NRT–Wg^{Discs}$}* flies suggested that, in addition, the growth rate of imaginal discs expressing *NRT–Wg* is autonomously compromised. The wings of such mosaic flies were smaller than those of controls (−23%; Fig. 4f–h and Extended Data Fig. 4d). This constitutes a more extreme size reduction than that seen in wings obtained from animals expressing NRT–Wg throughout (*wg{KO; NRT–Wg}*; −12%). One possible interpretation is that in the absence of Wingless release, imaginal disc growth slows down in an organ-autonomous manner, preventing a normal size to be reached in larvae that develop at a normal rate, as in the case in the *wg{KO; $WT^{Body}$; $NRT–Wg^{Discs}$}* genotype. To confirm this possibility, we mea-

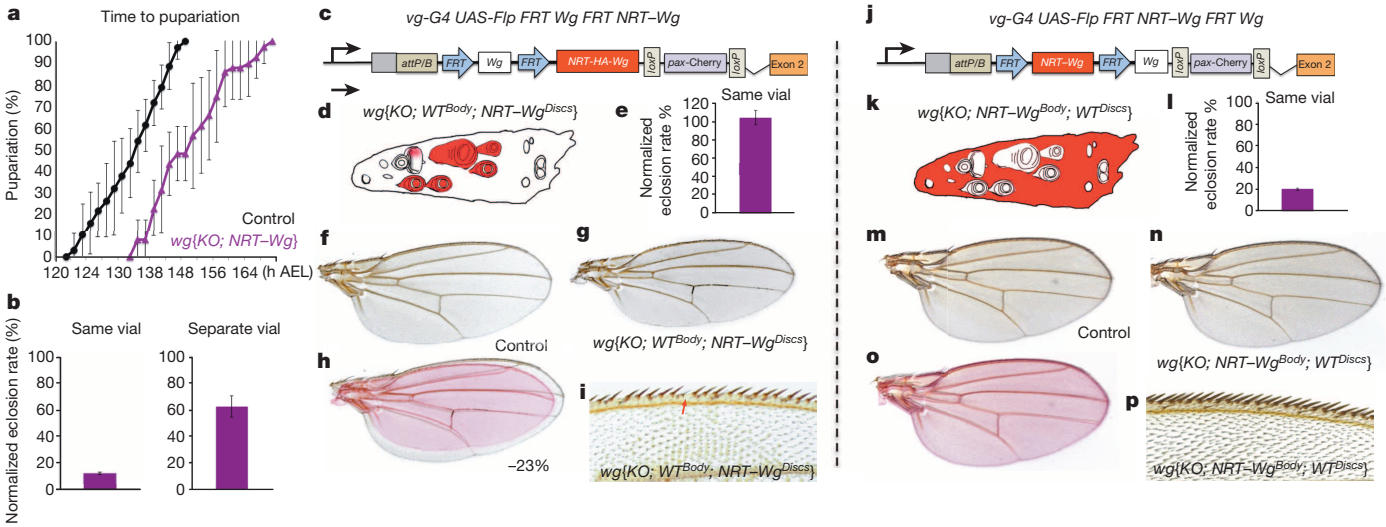



**Figure 4 | Tissue-specific allele switching to assess the contribution of Wingless release to organ-autonomous growth rate and organismal developmental timing.** **a**, Homozygous *wg{KO; NRT–Wg}* animals (70 animals, 8 experiments) were developmentally delayed ($P < 0.001$) relative to control siblings (*wg{KO; NRT–Wg}/Cyo-GFP*; 34 animals, 6 experiments). Larvae were grown in separate vials. **b**, Normalized eclosion rate of homozygous *wg{KO; NRT–Wg}* larvae cultured with (130 animals, 4 experiments) or without (90 animals, 6 experiments) control siblings. **c**, **d**, Wg to NRT–Wg conversion with *vestigial-gal4* (*vg-G4*) UAS-*Flp*. Expected NRT–Wg tissue is shown in red. **e**, Normalized eclosion rate of resulting *wg{KO; $WT^{Body}$; $NRT–Wg^{Discs}$}* larvae co-cultured with control siblings (80 animals, 4 experiments). **f–i**, The wings of *wg{KO; $WT^{Body}$; $NRT–Wg^{Discs}$}* flies were significantly smaller than those of control siblings (−23%, $P < 0.001$) and had an incomplete margin (missing bristles; red arrow). **j**, **k**, NRT–Wg to Wg conversion with *vestigial-gal4 UAS-Flp*. **l**, Normalized eclosion rate of the resulting *wg{KO; $WT^{Body}$; $NRT–Wg^{Discs}$}* larvae co-cultured with control siblings (>140 animals, 4 experiments) was significantly smaller than that of *wg{KO; $WT^{Body}$; $NRT–Wg^{Discs}$}* larvae ($P < 0.001$). **m**–**p**, The wing size and margin of *wg{KO; $WT^{Body}$; $NRT–Wg^{Discs}$}* flies were similar to those of sibling controls ($P > 0.05$). Error bars represent s.d. Statistical significance was assessed using Student's *t*-test.

sured disc size in *wg{KO; WT$^{Body}$; NRT–Wg$^{Discs}$}*, *wg{KO; NRT–Wg}*, and control animals at the onset of pupariation. The results showed that Wg tethering specifically in imaginal discs causes a ~15% size deficit (Extended Data Fig. 4e–g), a relatively mild reduction considering that growth occurs over a period of about 4 days. Another noteworthy feature of *wg{KO; WT$^{Body}$; NRT–Wg$^{Discs}$}* flies is that their wing margin lacked occasional sensory bristles (Fig. 4i), a phenotype not seen in homozygous *wg{KO; NRT–Wg}* flies grown in optimal conditions (Extended Data Fig. 1j–l). As margin specification occurs at the very end of larval life, we suggest that in *wg{KO; WT$^{Body}$; NRT–Wg$^{Discs}$}* flies, the margin cannot be completed for lack of time before pupariation (see Methods for another example of wing–body asynchrony). The above considerations suggest that a disc-autonomous growth delay does not hold up developmental progression of the organism and therefore that, at least in this instance, there is no organ-size control checkpoint. Moreover, our mosaic analysis by allele switching also shows that Wg release, or a level of signalling activity not achieved by NRT–Wg (see Methods for further discussion on their relative importance), contributes independently to ensuring timely growth in an organ-specific manner and to pacing organismal developmental progress.

## Conclusion

In this paper we have asked directly to what extent the spread of Wg, the main Wnt of *Drosophila*, is required for normal development of the wing. Transcription of *wg* takes place throughout wing precursors at early stages, and later becomes restricted to a narrow stripe at the dorsoventral boundary. From there, the protein product spreads to form a gradient[4,5,34], as expected from a morphogen. However, the requirement for graded expression has been questioned[21]. In addition, as we showed here, juxtacrine Wg signalling suffices for extensive growth and patterning. Early *wg* expression throughout the wing primordium, along with the persistent effect of signalling[21,26,35], ensure continued target-gene expression in the absence of Wg release. We infer that these processes contribute substantially to patterning and growth in the wild type. Nevertheless, target genes remain responsive until late stages[4,5]. Therefore, Wg spreading from the dorsoventral boundary is likely to boost proliferation, at least in cells within its reach. Our results, along with those of previous studies[14,21,35–37], suggest that the requirement for long-range spreading of other Wnts should be revisited. Furthermore, our genome-editing approach and associated tools for mosaic analysis provide a template for further investigation of other signalling proteins.

## METHODS SUMMARY

All the standard fly strains used in this study are described at http://www.flybase.org and all experiments were conducted at 25°C unless otherwise indicated. *UAS-HRP-CD8-GFP* encodes a transmembrane protein comprising extracellular HRP, the transmembrane domain of mouse CD8 and intracellular GFP (details upon request). Gene targeting was performed with a new vector and protocol described elsewhere[24]. The key elements of the targeting vector are shown in Extended Data Fig. 1a. The homology arms used for targeting of the *wg* locus were amplified by PCR from a BAC (http://www.pacmanfly.org) with primers listed in Methods. The mini-*white* marker of the targeting vector was used for initial identification of candidate targeted flies. True recombinants were confirmed by several criteria, including expression of Cherry, the phenotype of homozygous animals and PCR (primers and protocols described in Methods). Various constructs were integrated into the *attP* site of the targeted allele with a variety of 'reintegration vectors (RIVs)' using mini-*white* or *pax*-Cherry as a marker[24]. Both markers were used indistinctly for *wg{KO; NRT–Wg}*. Immunostaining of imaginal discs was performed according to standard protocols with the following primary antibodies: rabbit anti-HA (1:1,000, Cell Signalling Technology, C29F4), mouse anti-Wingless (1:100, Hybridoma Bank), rabbit anti-vestigial (1:50, gift from S. Carroll), mouse anti-Distalless (1:300, gift from S. Carroll), rabbit anti-Senseless (1:500, gift from H. Bellen), and chicken anti-β-galactosidase (1/200, Abcam). In all micrographs, blue staining shows DAPI, a nuclear marker. FISH was performed according to standard protocols[38]. Temperature shifts were performed by transferring culture vials between incubators maintained at the desired temperatures. Details on measuring developmental timing are provided in Methods. For all quantitative measurements, the error bars represent standard deviation. Samples were normally distributed. Sample size was chosen to ensure statistical significance, which was assessed by Student's *t*-test using Prism 6.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 30 May; accepted 12 November 2013.**
**Published online 25 December 2013.**

1. van Amerongen, R. & Nusse, R. Towards an integrated view of Wnt signaling in development. *Development* **136,** 3205–3214 (2009).
2. Clevers, H. & Nusse, R. Wnt/β-catenin signaling and disease. *Cell* **149,** 1192–1205 (2012).
3. Kiecker, C. & Niehrs, C. A morphogen gradient of Wnt/β -catenin signalling regulates anteroposterior neural patterning in *Xenopus*. *Development* **128,** 4189–4201 (2001).
4. Zecca, M., Basler, K. & Struhl, G. Direct and long-range action of a wingless morphogen gradient. *Cell* **87,** 833–844 (1996).
5. Neumann, C. J. & Cohen, S. M. Long-range action of Wingless organizes the dorsal-ventral axis of the *Drosophila* wing. *Development* **124,** 871–880 (1997).
6. Swarup, S. & Verheyen, E. M. Wnt/Wingless signaling in *Drosophila*. *Cold Spring Harb. Perspect. Biol.* **4** (2012).
7. Garcia-Bellido, A. & Merriam, J. R. Parameters of the wing imaginal disc development of *Drosophila melanogaster*. *Dev. Biol.* **24,** 61–87 (1971).
8. Johnston, L. A. & Gallant, P. Control of growth and organ size in Drosophila. *BioEssays* **24,** 54–64 (2002).
9. Martín, F. A., Herrera, S. C. & Morata, G. Cell competition, growth and size control in the *Drosophila* wing imaginal disc. *Development* **136,** 3747–3756 (2009).
10. Williams, J. A., Paddock, S. W. & Carroll, S. B. Pattern formation in a secondary field: a hierarchy of regulatory genes subdivides the developing *Drosophila* wing disc into discrete subregions. *Development* **117,** 571–584 (1993).
11. Couso, J. P., Knust, E. & Martinez Arias, A. Serrate and wingless cooperate to induce vestigial gene expression and wing formation in *Drosophila*. *Curr. Biol.* **5,** 1437–1448 (1995).
12. Ng, M., Diaz-Benjumea, F. J., Vincent, J. P., Wu, J. & Cohen, S. M. Specification of the wing by localized expression of wingless protein. *Nature* **381,** 316–318 (1996).
13. García-García, M. J., Ramain, P., Simpson, P. & Modolell, J. Different contributions of pannier and wingless to the patterning of the dorsal mesothorax of *Drosophila*. *Development* **126,** 3523–3532 (1999).
14. Martinez Arias, A. Wnts as morphogens? The view from the wing of *Drosophila*. *Nature Rev. Mol. Cell Biol.* **4,** 321–325 (2003).
15. Couso, J. P., Bishop, S. A. & Martinez Arias, A. The wingless signalling pathway and the patterning of the wing margin in *Drosophila*. *Development* **120,** 621–636 (1994).
16. Nolo, R., Abbott, L. A. & Bellen, H. J. Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in *Drosophila*. *Cell* **102,** 349–362 (2000).
17. Jafar-Nejad, H., Tien, A.-C., Acar, M. & Bellen, H. J. Senseless and Daughterless confer neuronal identity to epithelial cells in the *Drosophila* wing margin. *Development* **133,** 1683–1692 (2006).
18. Sato, A., Kojima, T., Ui-Tei, K., Miyata, Y. & Saigo, K. Dfrizzled-3, a new *Drosophila* Wnt receptor, acting as an attenuator of Wingless signaling in *wingless* hypomorphic mutants. *Development* **126,** 4421–4430 (1999).
19. Sivasankaran, R., Calleja, M., Morata, G. & Basler, K. The Wingless target gene *Dfz3* encodes a new member of the *Drosophila* Frizzled family. *Mech. Dev.* **91,** 427–431 (2000).
20. Giraldez, A. J. & Cohen, S. M. Wingless and Notch signaling provide cell survival cues and control cell proliferation during wing development. *Development* **130,** 6533–6543 (2003).
21. Baena-López, L. A., Franch-Marro, X. & Vincent, J.-P. Wingless promotes proliferative growth in a gradient-independent manner. *Sci. Signal.* **2,** ra60 (2009).
22. Zecca, M. & Struhl, G. A feed-forward circuit linking wingless, fat-dachsous signaling, and the warts-hippo pathway to *Drosophila* wing growth. *PLoS Biol.* **8,** e1000386 (2010).
23. Herr, P. & Basler, K. Porcupine-mediated lipidation is required for Wnt recognition by Wls. *Dev. Biol.* **361,** 392–402 (2012).
24. Baena-López, L. A., Alexandre, C., Mitchell, A., Pashakarnis, L. & Vincent, J.-P. Accelerated genome engineering in *Drosophila* without sequence constraints. *Development* **140,** 4818–4825 (2013).
25. Bergantiños, C., Corominas, M. & Serras, F. Cell death-induced regeneration in wing imaginal discs requires JNK signalling. *Development* **137,** 1169–1179 (2010).
26. Piddini, E. & Vincent, J.-P. Interpretation of the wingless gradient requires signaling-induced self-inhibition. *Cell* **136,** 296–307 (2009).
27. Dubois, L., Lecourtois, M., Alexandre, C., Hirst, E. & Vincent, J. P. Regulated endocytic routing modulates wingless signaling in *Drosophila* embryos. *Cell* **105,** 613–624 (2001).
28. Lecuit, T. *et al.* Two distinct mechanisms for long-range patterning by Decapentaplegic in the *Drosophila* wing. *Nature* **381,** 387–393 (1996).
29. Pérez, L. *et al.* Enhancer-PRE communication contributes to the expansion of gene expression domains in proliferating primordia. *Development* **138,** 3125–3134 (2011).
30. Halder, G. *et al.* The Vestigial and Scalloped proteins act together to directly regulate wing-specific gene expression in *Drosophila*. *Genes Dev.* **12,** 3900–3909 (1998).
31. Smith-Bolton, R. K., Worley, M. I., Kanda, H. & Hariharan, I. K. Regenerative growth in *Drosophila* imaginal discs is regulated by Wingless and Myc. *Dev. Cell* **16,** 797–809 (2009).

32. Colombani, J., Andersen, D. S. & Léopold, P. Secreted peptide Dilp8 coordinates *Drosophila* tissue growth with developmental timing. *Science* **336,** 582–585 (2012).
33. Garelli, A., Gontijo, A. M., Miguela, V., Caparros, E. & Dominguez, M. Imaginal discs secrete insulin-like peptide 8 to mediate plasticity of growth and maturation. *Science* **336,** 579–582 (2012).
34. Strigini, M. & Cohen, S. M. Wingless gradient formation in the *Drosophila* wing. *Curr. Biol.* **10,** 293–300 (2000).
35. Giorgianni, M. W. & Mann, R. S. Establishment of medial fates along the proximodistal axis of the *Drosophila* leg through direct activation of *dachshund* by Distalless. *Dev. Cell* **20,** 455–468 (2011).
36. Wilder, E. L. & Perrimon, N. Dual functions of wingless in the *Drosophila* leg imaginal disc. *Development* **121,** 477–488 (1995).
37. Couso, J. P., Bate, M. & Martínez-Arias, A. A wingless-dependent polar coordinate system in *Drosophila* imaginal discs. *Science* **259,** 484–489 (1993).
38. Artero, R., Furlong, E. E., Beckett, K., Scott, M. P. & Baylies, M. Notch and Ras signaling pathway effector genes expressed in fusion competent and founder cells during *Drosophila* myogenesis. *Development* **130,** 6257–6272 (2003).

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

**Acknowledgements** This work was supported by the UK Medical Research Council (U117584268), an ERC grant (WNTEXPORT) from the European Union to JPV and a Sir Henry Wellcome post-doctoral fellowship to L.A.B-L. (082694/Z/07/Z). We are grateful to U.-M. Fiuza for discussion and A. Mitchell for help in generating the first *wingless* knockout allele. Discussions with G. Struhl have led to significant improvement of the manuscript. We thank colleagues listed in the Methods Summary, as well as the Developmental Studies Hybridoma Bank and the Bloomington Stock Center for providing antibodies and fly strains.

**Author Contributions** All the experiments were performed jointly by L.A.B.-L. and C.A. C.A., L.A.B.-L., and J.-P.V. contributed equally to the conception of the work, the interpretation of results, and manuscript preparation.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-P.V. (jvincen@nimr.mrc.ac.uk).

## METHODS

**Editing the *wingless* locus.** Gene targeting was performed with a targeting vector and according to protocols described elsewhere[24]. The primers used to amplify homology arms and to confirm gene targeting are the following: forward 5′ arm, 5′-GATCAGTGCGGCCGCGCCGAGAAGAGATCGCCACCACCACTCTACTCTTTTGCACATGCC-3′; reverse 5′ arm, 5′-GATCGCTAGCCGGCACACACACTCTCACACTGACACACGGGGTATGATAGATACTTCC-3′; forward 3′ arm, 5′-GATCATGCATGGACACTGCCCGCCTCCAGCCCAGTCCCGTCCTCTGAAGCCGCCC-3′; reverse 3′ arm, 5′-GATCCCTAGGGCCGATCTGTTGCAATTTCCAAATCAAACAGCGCGCGAAACGTGTGGC-3′; for confirmation of 5′ recombination (genomic, forward), 5′-CAGCACTAAAATGGCTTCCTCCGC-3′; for confirmation of 5′ recombination (vector, reverse), 5′-CAACTGAGAGAACTCAAAGG-3′ (within *attP* site); for confirmation of 3′ recombination (vector, forward), 5′-TCGTATAATGTATGCTATACG-3′ (within Gal4 Poly A); for confirmation of 3′ recombination (genomic, reverse), 5′-GTTCCCGGAATAGTTTAGACCTC-3′.

After confirmation of targeting into *wingless*, much of the targeting vector was removed by crossing to a strain expressing Cre constitutively (Bloomington stock 851) (procedure outlined in ref. 24). The resulting strain, referred to as *wg{KO}* in the manuscript, was used as a host for reintegration of various constructs via the *attP* site. Reintegration was achieved by injecting a *wg{KO}* strain expressing the PhiC31 integrase.

The following reintegration vectors (RIV) were generated: RIV{*wg*, mini-*white*} (Extended Data Fig. 1d), the full-length *wg* cDNA, along with 135 bp of 5′ untranslated region (UTR) and 1,200 bp of 3′ UTR was inserted as a NotI-AscI fragment into RIV{*MCS*; mini-*white*} (ref. 24); RIV{*FRT Wg FRT QF*; *pax*-Cherry} (Fig. 3b), this was obtained by cloning the above NotI-AscI *wg* fragment into RIV{*FRT-MCS2-FRT QF*; *pax*-Cherry}, which is described in ref. 24; RIV{*NRT–Wg*; mini-*white*} (Fig. 1a–c and Extended Data Fig. 1e), pMT-NRT-HA-Wg was made by replacing the NotI-BglII fragment of pMT-Wg with a synthetic fragment (GENEWIZ) containing, from 5′ to 3′, the 5′ UTR of *wg*, the NRT open reading frame, DNA encoding two HA epitopes, and remaining *wg* coding sequences (up to the BglII), then DNA encoding NRT-HA-Wg was transferred to RIV{MCS; mini-*white*} as a NotI-AscI fragment; RIV{*FRT NRT–Wg FRT QF*; *pax*-Cherry} (Fig. 1i) was obtained by cloning the above NotI-AscI NRT-HA-Wingless fragment into RIV{FRT-MCS2-FRT QF; *pax*-Cherry}, which is described in[24]; RIV{*Gal4*; mini-*white*} (Fig. 3a and Extended Data Fig. 3c) is described elsewhere[24]; RIV{*FRT Wg FRT NRT–Wg*; *pax*-Cherry} (Fig. 4c and Extended Data Fig. 4b), first, the QF open reading frame was excised from RIV{*FRT Wg FRT QF*; *pax*-Cherry}with AvrII and AgeI and replaced by an XbaI-AgeI fragment containing NRT-HA-Wg obtained from pMT-NRT-HA-Wg; RIV{*FRT NRT–Wg FRT Wg*; *pax*-Cherry} (Fig. 4j), an XbaI-AgeI fragment encoding Wingless was obtained from pMT-Wg and cloned into RIV{*FRT NRT–Wg FRT QF*; *pax*-Cherry} pre-digested with AvrII and AgeI.

**Immunostaining, FISH and microscopy.** The following primary antibodies were used: rabbit anti-HA (1:1,000, Cell Signalling Technology, C29F4), mouse anti-Wingless (1:100, Hybridoma Bank), rabbit anti-Vestigial (1:50, gift from S. Carroll), mouse anti-Distalless (1:300, gift from S. Carroll), Rabbit anti-Senseless (1:500, gift from H. Bellen), and chicken anti-β-galactosidase (1/200, Abcam). Secondary antibodies labelled with Alexa 488 or Alexa 555 (used at 1:200) were obtained from Molecular Probes. Imaginal discs were mounted in Vectashield with DAPI (Vector Laboratories). In all micrographs, blue staining shows DAPI, a nuclear marker. Staining for HRP activity with DAB was performed as previously described[27]. FISH was performed according to standard protocols[38]. Fluorescence micrographs were acquired with a Leica SP5 confocal microscope. Embryo cuticles were prepared according to standard protocol. Bright field images from embryo cuticles were obtained with a Zeiss Axiophot2 microscope with an Axiocam HRC camera. Bright field and confocal images were processed with Photoshop CS4 (Adobe). Measurements of adult wing size, cell density and generation of fluorescence intensity profiles in wing imaginal discs were carried out as described previously[21].

**Temperature shifts to assess Neurotactin–Wingless perdurance and *wingless* promoter activity.** The perdurance of NRT–Wg was estimated in larvae of the genotype *wg{FRT NRT–Wg FRT QF}/wg{KO}; hedgehog-gal4 tubulin-gal80$^{ts}$/UAS-Flp* (Fig. 1i). Larvae of this genotype were kept at 18 °C until late-third instar. They were then transferred to 29 °C for 12 h to allow excision of NRT–Wg. Under these conditions, excision occurred in groups of cells without any apparent spatial reproducibility. To assess the activity of *wingless* promoter, *wg{KO; Gal4}, tubulin-gal80$^{ts}$/Cyo* flies were crossed to flies carrying *UAS-Flp* (Bloomington stock number 4539) and *Actin FRT stop FRT LacZ* (Bloomington stock number 6355). The progeny were allowed to lay eggs for 24 h at 18 °C. The resulting larvae were kept at 18 °C and then transferred to 29 °C at the indicated stages. Larvae were allowed to continue development at this temperature until late wandering stage, before puparation. The timeline of this protocol is illustrated in Extended Data Fig. 3b. In both experiments imaginal discs were then processed for conventional immunofluorescence.

**Assessing the size of *wingless* mutant territories.** *wg* mutant territories were generated with *hedgehog* (*hh*)*-gal4* and *UAS-flp*, which trigger mitotic recombination in virtually all cells of the P compartment at early stages of larval development. In control experiments[26], genetically marked wild-type tissue generated in a wild-type background occupied half of the posterior compartment. Unlike standard heat shock-induced clonal analysis, this approach overcomes the issues raised by differences in developmental timing between controls and experimentals. The genotypes used to generate the data shown in Extended Data Fig. 2c–f were: FRT40 $wg^{CX4}$/FRT40 ubiquitin-GFP; *hh-gal4, UAS-Flp* (Extended Data Fig. 2c) FRT40 $wg^{CX4}$/FRT40 ubiquitin-GFP *NRT–Wg*; *hh-gal4, UAS-Flp* (Extended Data Fig. 2d).

For both genotypes, larvae were dissected and fixed at the end of the third instar, before pupariation. At least 30 confocal planes were combined by maximal projection. The areas of GFP-positive and GFP-negative territories were measured in the prospective wing blade (area encircled by white dots in Extended Data Fig. 2e) with ImageJ. Over 10 discs were analysed for each genotype.

**Normalized eclosion rates.** For co-culture assays, eggs deposited by balanced flies were allowed to develop in a single vial and all the eclosed adults were genotyped (using the dominant marker on the balancer chromosome). The number of experimental (non-balancer) was normalized to the Mendelian ratio-adjusted number of sibling controls (balancer). For example, eggs deposited by *wg{KO; NRT–Wg}/Cyo* gave rise to homozygous *wg{KO; NRT–Wg}* (experimental) and *wg{KO; NRT–Wg}/Cyo* (control) flies. As these are expected in a 1:2 ratio, the normalized eclosion rate was calculated as (no. of experimental animals)/(0.5 × no. of controls). Note that the normalized eclosion rate can be larger then 100% as balancer chromosomes are likely to reduce fitness. For separate culture experiments, first instar larvae (24–28 h after egg laying) were genotyped using GFP-expressing balancer chromosomes. Fifteen larvae of each genotype were then transferred into separate vials and kept at 25 °C under normal laboratory conditions. The number of eclosed experimental flies was normalized to the number of eclosed control.

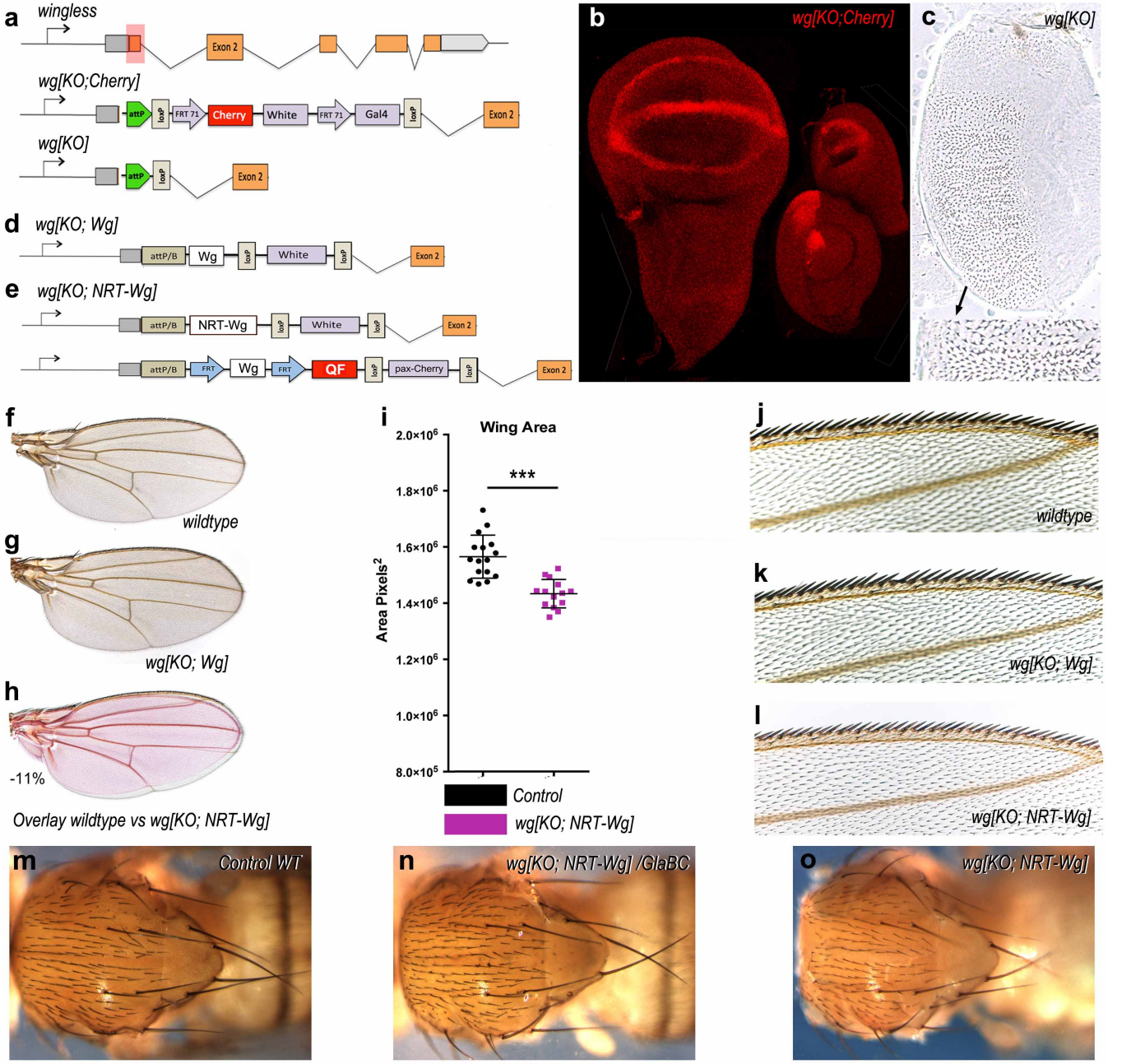
**Time of puparation.** First instar larvae (24–28 h after egg laying) were genotyped using GFP-expressing balancer chromosomes. Ten to fifteen larvae of each genotype were transferred into separate vials and kept at 25 °C under normal laboratory conditions until they started puparation. The number of larvae pupariating within a 2-h interval was then recorded. At least four independent experiments were performed for each genotype (Fig. 4a).

**Genetic markers of reintegration.** Depending on experimental requirements, two genetic markers, mini-*white* and *pax*-Cherry, were used alternately to track reintegrated genetic elements (as shown for *wg{KO; NRT–Wg}* in Extended Data Fig. 1e). We found that the presence of mini-*white* and *pax*-Cherry downstream of the reintegrated cDNA had an equally detrimental impact on quantifiable physiological functions such as developmental progression (see below), and organismal motility (not shown). Although the choice of marker is unlikely to be relevant, marker usage is listed in Supplementary Notes.

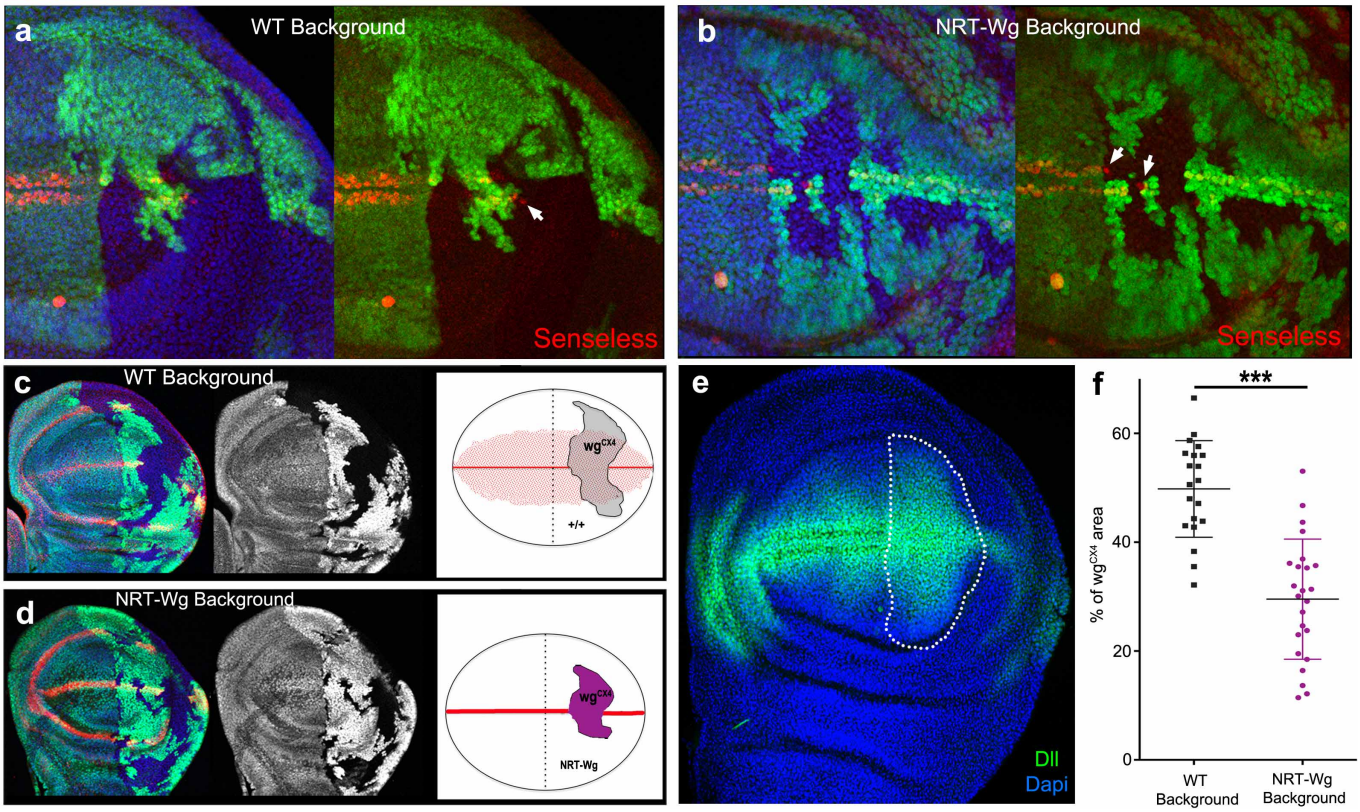
**Removal of the genetic marker from *wg{KO; NRT–Wg}* preferentially ameliorates developmental timing.** In the main text we have referred to *wg{KO; NRT–Wg; pax*-Cherry} and *wg{KO; NRT–Wg;* mini-*white*} indistinctly as *wg{KO; NRT–Wg}* as they behaved identically in all assays. In preliminary experiments, we assessed the effect of removing the genetic markers (*pax*-Cherry or mini-*white*) by Cre-mediated excision (see Extended Data Fig. 1e for position of the LoxP sites). We found that homozygous *wg{KO; NRT–Wg; floxed}* larvae grew faster and competed more effectively in co-culture than homozygous *wg{KO; NRT–Wg}* larvae (carrying mini-*white* or *pax*-Cherry), a phenotypic improvement that is likely to be due to enhanced NRT–Wg expression. However, the wing margin of homozygous *wg{KO; NRT–Wg; floxed}* flies were seen occasionally to miss sensory bristles. This confirms our suggestion that a minor extension of the growth period helps NRT–Wg flies to make a perfectly patterned wing margin. Our evidence so far is consistent with the notion that the developmental delay of *wg{KO; NRT–Wg}* animals could be due mostly to a mild reduction in expression and hence activity. It remains to be determined whether the wing autonomous growth defect is due to the lack of Wg release or reduced expression.

**Sample collection.** Animals for analysis were chosen randomly from a large collection at specific developmental age as required. When necessary, a given sex was studied as specified.

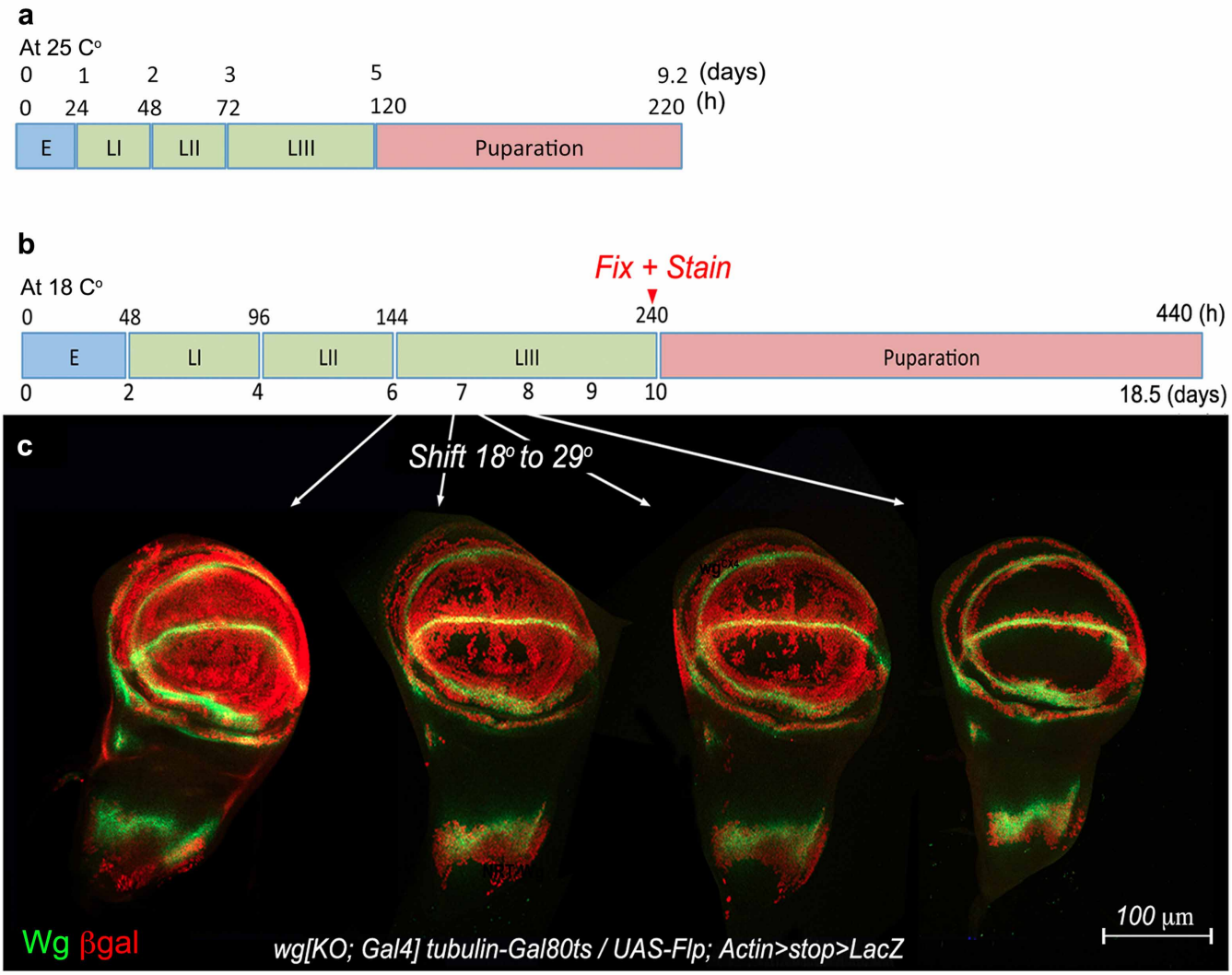
39. Alexandre, C., Lecourtois, M. & Vincent, J. Wingless and Hedgehog pattern *Drosophila* denticle belts by regulating the production of short-range signals. *Development* **126,** 5689–5698 (1999).



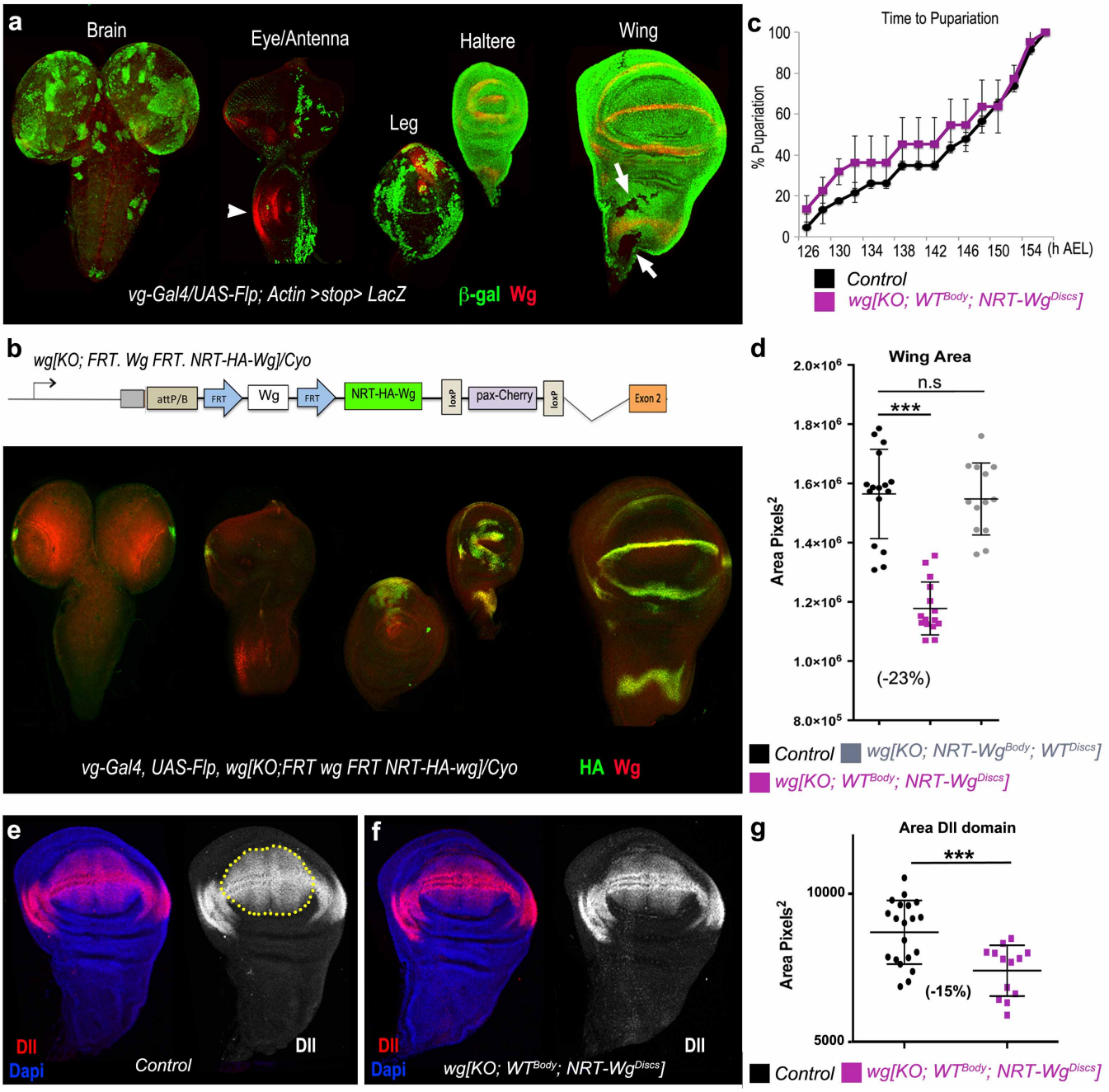

**Extended Data Figure 1 | Engineering the *wg* locus to express membrane-tethered Wg. a**, Structure of the *wingless* locus before targeting, after targeting, and after Cre-mediated excision. The *wg{KO}* allele was used as a founder line for subsequent reintegration. **b**, Cherry expression in wing, leg and haltere imaginal discs of larvae carrying one copy of *wg{KO; Cherry}*. **c**, Cuticle preparation of a homozygous *wg{KO}*larva at low and high magnification (black arrow). The phenotype is identical to that of $wg^{CX4}$ homozygous embryos[39]. **d**, Diagram showing the reintegration of a wild type *wingless* cDNA in the *wg{KO}* to generate *wg{KO; Wg}* (note presence of mini-*white*). **e**, Diagram showing the reintegration of the NRT–Wg cDNA in *wg{KO}*. This was achieved using either *pax*-Cherry or mini-*white* as a genetic marker, as indicated. **f**, Wing of a wild-type fly. **g**, Wing of a *wg{KO; Wg}* homozygous fly. **h**, Overlay of the wings shown in Fig. 1b to illustrate the mild wing size reduction in *NRT–Wg* flies. **i**, Wing size of *wg{KO; NRT–Wg}* homozygous ($n = 14$) and control (*wg{KO; NRT–Wg}/GlaBC*) flies ($n = 16$, $^{***}P < 0.001$). **j–l**, High-magnification view of the wing margin of wild-type, homozygous *wg{KO; Wg}*, and homozygous *wg{KO; NRT–Wg}*. They are barely distinguishable. **m–o**, Views of the dorsal thorax illustrate the normal arrangement of pattern elements such as microchaetes and macrochaetes in the genotypes indicated. Error bars represent s.d. Statistical significance was assessed using Student's *t*-test.




**Extended Data Figure 2 | Senseless expression and growth in *wingless*-null patches surrounded by wild-type or Neurotactin–Wingless-expressing cells.** **a**, **b**, Expression of Senseless (red) is lost in patches of *wingless* mutant cells (GFP-negative; $wg^{CX4}$ homozygotes) except in the cells located within one cell diameter of surrounding GFP-positive cells, which are wild-type cells (**a**) or homozygous *wg{KO; NRT–Wg}* (**b**) (white arrows). Mosaics were created by mitotic recombination in a way that generates approximately the same number progenitors for the two genotypes, as described in Methods. **c**, Example of mosaic imaginal discs generated as above to measure the growth of *wingless* mutant territory ($wg^{CX4}$ homozygous; GFP-negative) relative to that of wild type (**c**) or *wg{KO; NRT–Wg}* homozygous (**d**) tissue. Wg and NRT–Wg, detected with anti-Wg, are shown in red. **e**, Outline of the territory where the surface areas were assessed. **f**, Quantification of the areas colonised by *wingless* mutant cells (GFP-negative) in the two genetic backgrounds. On average, the *wingless*-null territory was smaller in the *wg{KO; NRT–Wg}* homozygous background ($n = 24$) than in the wild type ($n = 20$, $^{***}P < 0.001$). Error bars represent s.d. Statistical significance was assessed using Student's *t*-test.




**Extended Data Figure 3 | Activity of the *wingless* promoter during imaginal disc development.** **a**, Timing of key developmental stages at 25 °C. **b**, Developmental timing at 18 °C as it relates to the results illustrated in **c**. **c**, Permanent labelling of *wg*-expressing cells and their descendants at different stages of development. Genotype was *wg{KO; Gal4}, tubulin-gal80$^{ts}$/UAS-Flp; Actin FRT stop FRT LacZ* so that the stop cassette was only excised in cells that express *wingless* at the time of shifting to 29 °C to activate Flp expression and hence excision of the stop cassette. Discs were shifted from 18 °C to 29 °C at different stages (shown in **b**) but they were fixed and stained at the same stage, just before puparation.




**Extended Data Figure 4 | Tissue-specific allele switching to determine the anatomical origin of organismal developmental delay in Neurotactin–Wingless animals. a**, Cumulative pattern of *vestigial-gal4* activity in various organs precursors. Expression of *vestigial-gal4* at any stage or place leads to excision of the stop cassette in *Actin FRT stop FRT LacZ* thus marking permanently the corresponding cells. As expected, nearly the whole wing and haltere discs were labelled at the end of larval development. In wing imaginal discs, only a few cells were β-Galactosidase-negative that did not overlap with the domain of Wg expression (anti-Wg, red). In the eye antennal disc, the patterns of Wg (white arrowhead) and β-Galactosidase expression are also non-overlapping. Therefore, in combination with *UAS-Flp, vestigial-gal4* is expected to excise an FRT cassette throughout the domain of wingless expression. Examination of the brain and CNS shows that *vestigial-gal4* is unexpectedly active in these tissues. **b**, In larvae of genotype *vestigial-gal4, UAS-Flp, wg{FRT Wg FRT NRT–Wg}/Cyo wg,* most of the *wg*-expressing cells in leg, haltere and wing imaginal discs, but not in the brain and CNS, were converted to expressing NRT-HA-Wg (anti-HA; green). **c**, Developmental timing in *wg{KO; WT$^{Body}$; NRT–Wg$^{Disc}$}* (*vestigial-gal4, UAS-Flp, wg{FRT Wg FRT NRT–Wg}*) and control (*vestigial-gal4, UAS-Flp, wg{FRT Wg FRT NRT–Wg}/GlaBC*) larvae (80 animals, 4 experiments). The two data sets cannot be statistically distinguished ($P > 0.05$). **d**, Adult wing size for three genotypes: *wg{KO; WT$^{Body}$; NRT–Wg$^{Discs}$}/GlaBc* obtained from selfed *vestigial-gal4, UAS-Flp, wg{FRT Wg FRT NRT–Wg}/GlaBc* ($n = 16$, shown in black); *wg{KO; WT$^{Body}$; NRT–Wg$^{Discs}$}*, obtained from homozygous *vestigial-gal4, UAS-Flp, wg{FRT Wg FRT NRT–Wg}* ($n = 15$, shown in purple); and *wg{KO; NRT–Wg$^{Body}$;WT$^{Discs}$}*, obtained from homozygous *vestigial-gal4, UAS-Flp, wg{FRT NRT–Wg FRT Wg}* ($n = 13$, shown in grey). **e**–**g**, Extent of *Distal-less* expression in *wg{KO; WT$^{Body}$; NRT–Wg$^{Discs}$}* heterozygotes (over Cyo; **e**) and homozygotes (**f**). All the discs were obtained from immobile larvae at the time of anterior spiracle eversion, an event that marks the onset of pupariation. The extent of the *Distal-less* domain was estimated from the surface area of a polygon drawn around the zone of immunoreactivity, as shown. The results, plotted in panel **g**, show a mild reduction in *wg{KO; WT$^{Body}$; NRT–Wg$^{Discs}$}*discs ($n = 13$) compared to controls ($n = 20$; ***$P < 0.001$; n.s., not significantly different). Error bars represent s.d. Statistical significance was assessed using Student's *t*-test.

LETTER

doi:10.1038/nature12765

# The rarity of dust in metal-poor galaxies

David B. Fisher[1,2], Alberto D. Bolatto[1], Rodrigo Herrera-Camus[1], Bruce T. Draine[3], Jessica Donaldson[1], Fabian Walter[4], Karin M. Sandstrom[4], Adam K. Leroy[5], John Cannon[6] & Karl Gordon[7]

**Galaxies observed at redshift $z > 6$, when the Universe was less than a billion years old, thus far very rarely show evidence[1–3] of the cold dust that accompanies star formation in the local Universe, where the dust-to-gas mass ratio is around one per cent. A prototypical example is the galaxy Himiko ($z = 6.6$), which—a mere 840 million years after the Big Bang—is forming stars at a rate of 30–100 solar masses per year, yielding a mass assembly time of about $150 \times 10^6$ years. Himiko is thought to have a low fraction (2–3 per cent of the Sun's) of elements heavier than helium (low metallicity), and although its gas mass cannot yet be determined its dust-to-stellar mass ratio is constrained[3] to be less than 0.05 per cent. The local dwarf galaxy I Zwicky 18, which has a metallicity about 4 per cent that of the Sun's[4] and is forming stars less rapidly (assembly time about $1.6 \times 10^9$ years) than Himiko but still vigorously for its mass[5], is also very dust deficient and is perhaps one of the best analogues of primitive galaxies accessible to detailed study. Here we report observations of dust emission from I Zw 18, from which we determine its dust mass to be 450–1,800 solar masses, yielding a dust-to-stellar mass ratio of about $10^{-6}$ to $10^{-5}$ and a dust-to-gas mass ratio of $3.2–13 \times 10^{-6}$. If I Zw 18 is a reasonable analogue of Himiko, then Himiko's dust mass must be around 50,000 solar masses, a factor of 100 below the current upper limit. These numbers are quite uncertain, but if most high-$z$ galaxies are more like Himiko than like the very-high-dust-mass galaxy SDSS J114816.64 + 525150.3 at $z \approx 6$, which hosts a quasar[6], then our prospects for detecting the gas and dust inside such galaxies are much poorer than hitherto anticipated.**

The recent study[7] of HFLS3, a 'maximum starburst' (that is, a galaxy that converts gas into stars at a rates that are close to the theoretical limit) at $z = 6.3$, provides an example of a galaxy with a large amount of dust (about a billion solar masses, $M_\odot$), and a ratio of dust mass to gas mass of 0.01 and a ratio of dust mass to star mass of 0.04, which are more like those of nearby starbursting galaxies. This galaxy has an astonishing star-formation rate of around 3,000 $M_\odot$ per year, and converts its gas into stars at rates 2,000 times that of typical galaxies, which are properties rare even for gas-rich high-$z$ galaxies. Frequently, observations of dust and molecules in high-$z$ galaxies tend to target bright active galaxies[8], such as HFLS3 or J114816.64+525150.3. Such massive galaxies are well known to be rare at all redshifts. For those 'normal' galaxies that are not gravitationally lensed, the dust emission is typically detected at a signal-to-noise ratio of 3–5 in 1–2 h of integration time[9,10]. Here we report observations of dust emission from I Zw 18, which is a galaxy with a metallicity of 4 per cent that of the Sun's[4] and a star-formation rate of only 0.1 $M_\odot$ per year[5], and which is therefore a good analogue of primitive galaxies.

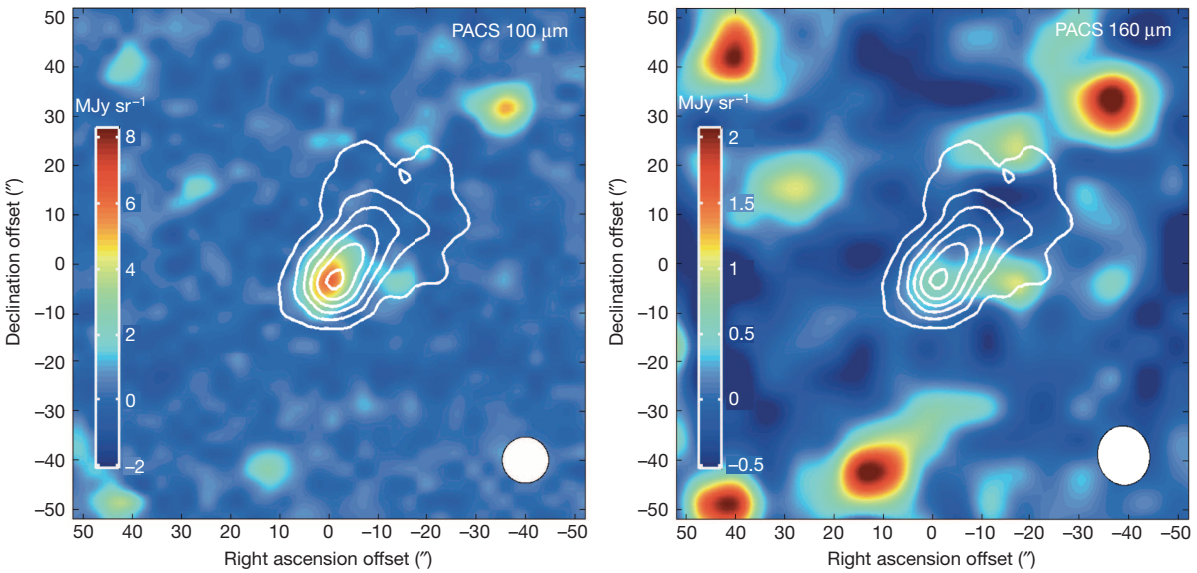



**Figure 1 | The 100-μm and 160-μm images of I Zw 18.** The colour scale is in units of megajanskys per steradian. Shown is the far-infrared detection, at $\lambda = 100$ μm, of dust emission in galaxy I Zw 18, and the marginal detection of I Zw 18 at 160 μm. These new observations (proposal identification, PID: OT_dbfisher_1) were obtained with the Herschel PACS. White contours represent the surface density of atomic hydrogen from the Very Large Array map[11]. The beam size of the H I map is $8.8 \times 8.3$ arcsec. For display purposes, the pixel size of the infrared maps is resampled to match the pixel size of the H I map. At 100 μm I Zw 18 is clearly detected, and well matched to the centre of the H I gas contours. The emission we detect in both far-infrared filters is contained within a small region (15″ or 1.3 kpc). We note that the off-target peaks in the 160-μm map are not noise: they are all coincident with peaks in the 100-μm map, and are therefore most probably background targets. At 160 μm we detect emission at the $3\sigma$ level that is consistent with the peak of both H I and the infrared emission at 24 μm, 70 μm and 100 μm, and which we attribute to I Zw 18. The peak and extent of the emission in our images is coincident with that of the Hα emission from Hubble Space Telescope images[5], and also the peak of H I emission[11]. The white ovals represent the shape and size of the beam of the infrared map.

[1]Department of Astronomy, Laboratory for Millimeter-wave Astronomy and Joint Space Institute, University of Maryland, College Park, Maryland 20742, USA. [2]Centre for Astrophysics and Supercomputing, Swinburne University, PO Box 218, Hawthorn, Victoria 3122, Australia. [3]Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08544, USA. [4]Max-Planck-Institut für Astronomie, Königstuhl 17, Heidelberg 69117, Germany. [5]National Radio Astronomy Observatory, Charlottesville, Virginia 22903, USA. [6]Department of Physics and Astronomy, Macalester College, 1600 Grand Avenue, Saint Paul, Minnesota 55105, USA. [7]Space Telescope Science Institute, Baltimore, Maryland 21218, USA.

galaxies where deep submillimetre observations have been performed, at present only upper limits for both [C II] and the submillimetre continuum exist[1,2]. These observational limits suggest that for the first 800 million years of galaxy evolution, galaxies with very little dust and low metallicity, like Himiko, are more typical. Results from stellar-population analyses do indeed show that high-$z$ galaxies have very little evidence of dust extinction[9]. An understanding of the physical conditions under which stars form in these primitive systems, however, can come only from the study of local analogues.

Located at a distance of 18 Mpc (ref. 10), I Zw 18 is the archetypal star-forming, very low-metallicity[4] galaxy ($12 + \log(O/H) = 7.17$, or 1/30 the solar metallicity). I Zw 18 is gas-rich[11–13] (atomic hydrogen mass of $2.3 \times 10^8\ M_\odot$ and molecular hydrogen mass of less than $5 \times 10^7\ M_\odot$) and actively star-forming[5] at a rate of $(0.05 \pm 0.02)\ M_\odot$ per year. Given its stellar mass (the combined mass of all the stars in the galaxy) of $9 \times 10^7\ M_\odot$, this galaxy has a very high gas fraction (its gas mass is two-thirds of the total mass of its gas and stars). I Zw 18 is currently undergoing a starburst phase[14], but despite its active star formation no CO emission has been detected; such emission would indicate the presence of molecular gas[12] in I Zw 18. The lowest-metallicity detection of CO has recently been reported in the dwarf galaxy Wolf–Lundmark–Melotte (WLM)[15] ($12 + \log(O/H) \approx 7.8$, or 1/8 the solar value). Unlike WLM, I Zw 18 has a very active star-forming environment, which probably photodissociates CO. These properties mean that it is among the closest analogues to primitive high-$z$ galaxies, although I Zw 18 contains a larger population of evolved stars than may be found in the early Universe.

We note that the difference in stellar mass between I Zw 18 (or any local low-metallicity galaxy) and observed high-$z$ galaxies is large. Because of the smaller potential well, identical starbursts can in principle more easily drive dust and metal-enriched gas out of I Zw 18 than out of Himiko, for example. Nonetheless, I Zw 18 and galaxies like it remain our best candidates for the study of metal-poor, starbursting environments.

Using the Photodetector Array Camera and Spectrometer (PACS) for the Herschel Space Observatory, we measured the flux of I Zw 18 to be 21.1 mJy at wavelength $\lambda = 100\ \mu$m, with an uncertainty in flux (calculated by placing apertures randomly in the map) of $\pm 2.9$ mJy (signal-to-noise ratio, $S/N \approx 7$) and a calibration uncertainty of 10% ($\pm 2.1$ mJy). Recent results find a 100-μm flux that is consistent with our result[16]. 160 μm we measured a flux of 5.6 mJy, with a flux uncertainty in the map of $\pm 1.3$ mJy ($S/N \approx 4$) and a calibration uncertainty of $\pm 0.6$ mJy. (Maps are shown in Fig. 1 and our procedure is discussed in the Supplementary Information.) Together with these detections, we used a number of ancillary data sources to construct a full infrared spectral energy distribution for modelling the dust and star formation. We use data from the Spitzer Space Telescope[12,17] covering 3.6 μm, 4.5 μm, 5.8 μm, 8.0 μm, 24 μm and 70 μm, as well as a spectrum from the Spitzer Infrared Spectrograph[14] (see Supplementary Information).

The dust mass we determined from models (shown in Fig. 2) with mixed dust grain temperatures is $912^{+912}_{-456}\ M_\odot$ (see Supplementary Information for a complete discussion of the uncertainties). Modified blackbody models are also usually used for fitting dust spectral energy distributions, although they yield unrealistically low dust masses because of the assumption of a single temperature. For comparison, a modified blackbody model with $F_\nu \propto \nu^{1.5} B_\nu(T_{\rm dust})$—where $F_\nu$ is the flux, $\nu$ is the frequency, $B$ is the brightness of a blackbody function

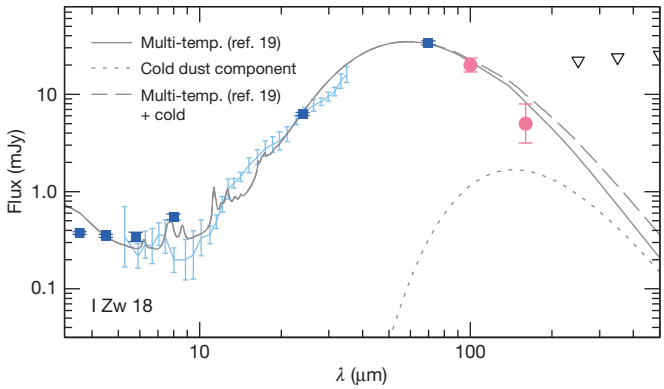



**Figure 2 | The far infrared spectral energy distribution of I Zw 18.** Blue squares represent fluxes measured by the Infrared Array Camera (IRAC) and the Multiband Imaging Photometer (MIPS) onboard the Spitzer Space Telescope[12,17]. The blue line is a smoothed Spitzer Infrared Spectrograph spectrum[14]. The open triangles are upper limits from the Herschel Spectral and Photometric Imaging Receiver (SPIRE) instrument. The red circles show our new Herschel PACS data at 100 μm and 160 μm. Error bars represent $1\sigma$ uncertainties. We fitted the spectral energy distribution with models based on a mixture of carbonaceous and silicate dust grains[19], which assume a distribution of grain sizes matching that of the Milky Way. This is a commonly adopted dust model, allowing direct comparison to dust masses in the literature, and is not expected to introduce large errors in the dust mass estimate[19]. Most of the modelled dust is heated by a single starlight intensity $U_{\rm min}$, but a fraction is heated by a power-law distribution of intensities, with an adjustable slope $\alpha$ and upper cut-off of $U_{\rm max}$. The best-fitting dust model to the spectral energy distribution of I Zw 18, shown here, returns the following values: $\alpha = 2.4$, $U_{\rm min} = 100$, and $<U> \approx 200$, for $U_{\rm max} = 10^7$ and no polycyclic aromatic hydrocarbons, with $U$ in units of the radiation field in the vicinity of the Sun. The solid grey line represents the best-fitting dust model[19]. The dotted line represents a cold ($T_{\rm dust} = 20$ K) dust model with a mass of 1,000 $M_\odot$ and a $\kappa$ value that is proportional to $\nu^{1.5}$. The dashed line represents the linear combination of the cold component and the dust model. We find that including the cold component of dust increases the discrepancy between the model and the 160-μm flux to $2.3\sigma$. Our assumption of a factor-of-two uncertainty in the dust mass, therefore accounts for the possibility of the cold component.

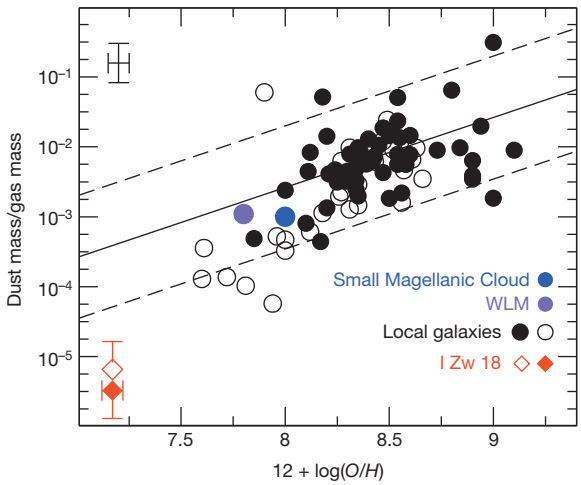



**Figure 3 | The dust-to-gas ratio of galaxies compared to metallicity for local galaxies and I Zw 18.** The local galaxy sample consists of dust-to-gas ratios from two recent papers[20,21]; the sample is representative of typical disk and dwarf galaxies in the local Universe. For some galaxies in the local sample, flux measurements that allow for the gas-mass determination of either H I or $H_2$ are not available in the literature. In this case we estimate the total gas mass with an empirical correlation relating molecular to atomic gas: $M(H_2) = 0.008 M(H\,I)^{1.2}$. Those galaxies are plotted as open circles. We also include two nearby, well known, low-metallicity galaxies: WLM[15] (purple circle) and the Small Magellanic Cloud[22] (blue circle). The error bars in the upper left corner represent the median $2\sigma$ for the local galaxy sample. A common assumption is that dust-to-gas ratio in galaxies scales linearly with metallicity[18]. We therefore show a linear bisecting line set to match the local galaxy sample; the dashed lines represent the $\pm 2\sigma$ root mean squared scatter around this bisector. I Zw 18 is a clear outlier from this correlation. (The open red diamond symbol represents the local dust-to-gas ratio.) Both the Small Magellanic Cloud and WLM have a linear relationship between dust-to-gas ratio and metallicity; however, we note that, unlike those two galaxies, I Zw 18 is a starbursting galaxy. If I Zw 18 is representative of starbursting low-metallicity environments, this implies that the dust mass is lower than we may expect in primitive galaxies of the early Universe.

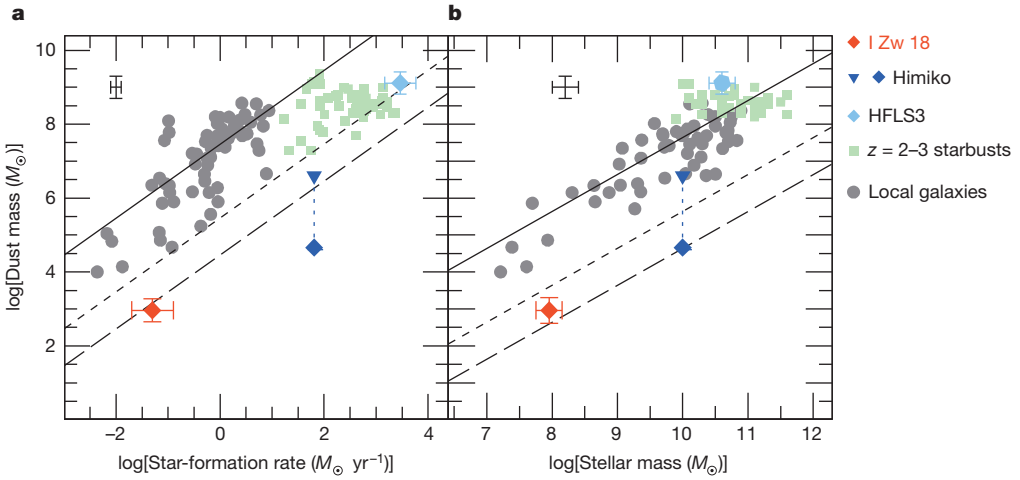



**Figure 4 | Dust mass versus star-formation rate and stellar mass for local disks, high-$z$ starbursts and I Zw 18.** Here we compare dust masses to star-formation rates (**a**) and stellar masses (**b**) of a sample of normal and starbursting galaxies[20–25], including I Zw 18. The definitions of symbols are the same in both panels. The red diamond represents I Zw 18. Squares represent $1 < z < 3$ starbursts[26]. The blue triangle represents the present upper limit on the dust mass of Himiko; the blue diamond shows the dust mass of Himiko if it has a similar dust-to-stellar mass ratio to I Zw 18. Solid lines represent linear relationships matching the local galaxies; the dashed lines represent 1/100 (short dashes) and 1/1,000 (long dashes) of the local galaxy sample. The error bars in both panels represent the median $2\sigma$ error bar for the sample. We note that very significant differences exist between I Zw 18 and the other starbursting systems shown here; the distant starbursts have star-formation rates five orders of magnitude higher than I Zw 18, and many of them have solar metallicity[26]. I Zw 18 is clearly an extreme outlier towards lower ratios of dust to star-formation rate and lower ratios of dust to stellar mass when compared with typical nearby galaxies, and the dust mass is even lower per unit of star-formation rate than for starbursting galaxies. As indicated by the extreme difference in the current upper limit of Himiko, and its predicted dust mass using the ratio of dust mass to star mass of I Zw 18, observations of the highest-$z$ galaxies may be significantly overestimating the dust mass.

representing the dust of temperature $T_{\rm dust}$—and mass emissivity $\kappa_{200} = 6.37\ {\rm cm^2\ g^{-1}}$ reproduces the flux at 70 μm and 160 μm for $T_{\rm dust} = 70$ K and a dust mass of about 250 $M_\odot$. We find that a significant mass of cold dust with $T_{\rm dust} > 15$ K would be detected by our 160-μm flux (see Supplementary Information). Independently of the assumed model, the dust mass necessary to explain the spectral energy distribution observed in I Zw 18 is extremely low.

Under simplistic model assumptions, decreasing the amount of heavy elements, of which the dust grains are formed, results in a proportional decrease in the dust-to-gas ratio. This scenario is frequently assumed in cosmological models of star formation[18]. Observational constraints for this relationship in the early Universe are scarce. I Zw 18 provides a probe of such environments, and our measurements directly constrain the relationship between dust-to-gas ratio and metallicity at very low metallicities[12,19,20] ($12 + \log(O/H) \leq 8$). We find that I Zw 18 falls roughly two orders of magnitude below the linear correlation between metallicity and dust-to-gas ratio (Fig. 3). The distance to the linear relation is very significant, approximately four times larger than the spread in the data, and much larger than the error bars on the measurement. Using dust-to-gas ratios measured only in the infrared-emitting region (the central 15 arcsec) typically results in a linear relationship even at low metallicity[19], $12 + \log(O/H) \approx 8$, but not in the case of I Zw 18, for which the local (that is, measured only in the infrared-emitting region) dust-to-gas mass ratio is still a factor of 38 below the linear relationship.

I Zw 18 stands out in the local galaxy population because it has an environment that is both starbursting and also lacking heavy elements. Our results suggest that in starbursting galaxies with very low metallicities the dust-to-gas ratio is determined by more than just the availability of heavy elements. The essentially dust-free character of nascent galaxies[1–3] like Himiko therefore probably reflects a combination of low metallicity and the balance of the dust production-and-destruction mechanisms in a starburst environment, which act together to keep the dust-to-gas ratio very low in these galaxies.

These are properties that are common at very high redshift ($z$>6), and consequently we expect those primitive galaxies to exhibit very low dust masses compared to their star-formation rates and stellar masses. The ratio of dust mass to star-formation rate (Fig. 4) in I Zw 18 is more than two orders of magnitude lower than typical in local galaxies, and an order of magnitude lower than observed in the $z = 2$–3 starbursts. Even when normalized by its stellar mass or star-formation rate, the dust mass of I Zw 18 is extremely small compared with both local and $z = 2$–3 galaxies. In Fig. 4 we show that if one assumes that Himiko has the same dust-to-gas ratio as I Zw 18, rather than using the upper limit on the 1.2-mm flux[3], this would significantly affect the amount of dust relative to the gas mass and star mass in high-z galaxies. With considerable uncertainty, we can scale the stellar mass of Himiko with the dust-to-stellar-mass ratio of I Zw 18, and place it on Fig. 4 with a dust mass of about 50,000 $M_\odot$.

If the dust temperature is 40 K (ref. 3), we calculate that Himiko would have a flux density of 0.5 μJy at Earth at 260 GHz, which would require several tens of days of integration with the complete Atacama Large Millimeter Array (ALMA) to detect. Maximum starbursts like HFLS3 are very rare, even in the early Universe[7] (about one per cubic gigaparsec), whereas blue dust-poor 'drop-out' (so highly redshifted that they can only be seen in the red filters) galaxies are much more common at high redshifts[9] (one in a thousand per cubic megaparsec). This implies that the interstellar medium of I Zw 18 may indeed be representative of the primitive galaxy population in the early Universe. If this is the case, the prospects for detecting dust emission at $z > 6$ will probably be limited to unusually evolved sources like HFLS3 and SDSS J114816.64 + 525150.3.

**Received 18 June; accepted 11 October 2013.**
**Published online 8 December 2013.**

1. Walter, F. *et al.* Evidence for low extinction in actively star-forming galaxies at z > 6.5. *Astrophys. J.* **752,** 93–98 (2012).
2. Kanekar, N., Wagg, J., Ram, C. R. & Carilli, C. L. A search for C II 158 μm line emission in HCM 6A, a Lyα emitter at z = 6.56. *Astrophys. J.* **771,** L20–L25 (2013).
3. Ouchi, M. *et al.* An intensely star-forming galaxy at z~7 with low dust and metal content revealed by deep ALMA and HST observations. Preprint at http://arxiv.org/abs/1306.3572 (2013).
4. Skillman, E. D. & Kennicutt, R. C. Jr. Spatially resolved optical and near-infrared spectroscopy of I ZW 18. *Astrophys. J.* **411,** 655–666 (1993).
5. Cannon, J. M., Skillman, E. D., Garnett, D. R. & Dufour, R. J. Dust in I Zw 18 from Hubble Space Telescope narrowband Imaging. *Astrophys. J.* **565,** 931–940 (2002).
6. Walter, F. *et al.* Molecular gas in the host galaxy of a quasar at redshift z = 6.42. *Nature* **424,** 406–408 (2003).

7. Riechers, D. A. *et al.* A dust-obscured massive maximum-starburst galaxy at a redshift of 6.34. *Nature* **496,** 329–333 (2013).
8. Vieira, J. D. *et al.* Dusty starburst galaxies in the early Universe as revealed by gravitational lensing. *Nature* **495,** 344–347 (2013).
9. Bouwens, R. J. *et al.* UV-continuum slopes of >4000 z~4–8 galaxies from the HUDF/XDF, HUDF09, ERS, CANDELS-South, and CANDELS-North fields. Preprint at http://arxiv.org/abs/1306.2950 (2013).
10. Aloisi, A. *et al.* I Zw 18 revisited with HST ACS and cepheids: new distance and age. *Astrophys. J.* **667,** L151–L154 (2007).
11. van Zee, L., Westpfahl, D., Haynes, M. P. & Salzer, J. J. The complex kinematics of the neutral hydrogen associated with I ZW 18. *Astron. J.* **115,** 1000–1015 (1998).
12. Herrera-Camus, R. *et al.* Dust-to-gas ratio in the extremely metal-poor galaxy I Zw 18. *Astrophys. J.* **752,** 112–119 (2012).
13. Lelli, F., Verheijen, M., Fraternali, F. & Sancisi, R. Dynamics of starbursting dwarf galaxies: I Zw 18. *Astron. Astrophys.* **537,** A72 (2012).
14. Wu, Y. *et al.* Dust in the extremely metal-poor blue compact dwarf galaxy I Zw 18: the Spitzer mid-infrared view. *Astrophys. J.* **662,** 952–958 (2007).
15. Elmegreen, B. G. *et al.* Carbon monoxide in clouds at low metallicity in the dwarf irregular galaxy WLM. *Nature* **495,** 487–489 (2013).
16. Remy-Ruyer, A. *et al.* Revealing the cold dust in low-metallicity environments. I. Photometry analysis of the Dwarf Galaxy Survey with Herschel. *Astron. Astrophys.* **557,** A95 (2013).
17. Engelbracht, C. W. *et al.* Metallicity effects on dust properties in starbursting galaxies. *Astrophys. J.* **678,** 804–827 (2008).
18. Gnedin, N. Y. & Kravtsov, A. V. Environmental dependence of the Kennicutt-Schmidt relation in galaxies. *Astrophys. J.* **728,** 88–108 (2011).
19. Draine, B. T. & Li, A. Infrared emission from interstellar dust. IV. The silicate-graphite-PAH model in the post-Spitzer era. *Astrophys. J.* **657,** 810–837 (2007).
20. Galametz, M. *et al.* Probing the dust properties of galaxies up to submillimetre wavelengths. II. Dust-to-gas mass ratio trends with metallicity and the submm excess in dwarf galaxies. *Astron. Astrophys.* **532,** 56–74 (2011).
21. Draine, B. T. *et al.* Dust masses, PAH abundances, and starlight intensities in the Sings galaxy sample. *Astrophys. J.* **663,** 866–894 (2007).
22. Leroy, A. *et al.* The Spitzer Survey of the Small Magellanic Cloud: far-infrared emission and cold gas in the Small Magellanic Cloud. *Astrophys. J.* **658,** 1027–1046 (2007).
23. Kennicutt, R. C. Jr *et al.* Dust-corrected star formation rates of galaxies. I. Combinations of Hα and infrared tracers. *Astrophys. J.* **703,** 1672–1695 (2009).
24. Skibba, R. *et al.* The emission by dust and stars of nearby galaxies in the Herschel KINGFISH Survey. *Astrophys. J.* **738,** 89–107 (2011).
25. Fisher, D. B. *et al.* The molecular gas density in galaxy centers and how it connects to bulges. *Astrophys. J.* **764,** 174–198 (2013).
26. Magnelli, B. *et al.* A Herschel view of the far-infrared properties of submillimetre galaxies. *Astron. Astrophys.* **539,** 155–190 (2012).


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

**Acknowledgements** D.B.F., A.D.B. and R.H.-C. acknowledge support from the University of Maryland and the Laboratory for Millimeter Astronomy and NSF grant number AST0838178. A.D.B. acknowledges partial support from CAREER NSF grant numbers AST0955836 and AST1139998 and from a Research Corporation for Science Advancement Cottrell Scholar award. B.T.D. acknowledges partial support from NSF grant number AST1008570. J.M.C. is supported by NSF grant number AST1211683. K.M.S. acknowledges support from a Marie Curie International Incoming fellowship. PACS has been developed by a consortium of institutes led by MPE (Germany) and including UVIE (Austria); KU Leuven, CSL, IMEC (Belgium); CEA, LAM (France); MPIA (Germany); INAF-IFSI/OAA/OAP/OAT, LENS, SISSA (Italy) and IAC (Spain). This development has been supported by the funding agencies BMVIT (Austria), ESA-PRODEX (Belgium), CEA/CNES (France), DLR (Germany), ASI/INAF (Italy) and CICYT/MCYT (Spain).

**Author Contributions** D.B.F. and A.D.B. wrote the text of both the proposal and this manuscript. D.B.F. and R.H.-C. performed detailed calculations. J.D. reduced the Herschel data. B.T.D. modelled the spectral energy distribution. A.K.L. and F.W. obtained and reduced the CO observations. J.C. obtained the Hα flux for I Zw 18. All authors participated in discussion of results and helped with revision of the manuscript.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.B.F. (dfisher@swin.edu.au) or A.D.B. (bolatto@astro.umd.edu).

LETTER

doi:10.1038/nature12763

# Face-to-face transfer of wafer-scale graphene films

Libo Gao[1,2], Guang-Xin Ni[1,3], Yanpeng Liu[2], Bo Liu[1,2], Antonio H. Castro Neto[1,3] & Kian Ping Loh[1,2]

**Graphene has attracted worldwide interest since its experimental discovery[1,2], but the preparation of large-area, continuous graphene film on $SiO_2$/Si wafers, free from growth-related morphological defects or transfer-induced cracks and folds, remains a formidable challenge[3]. Growth of graphene by chemical vapour deposition on Cu foils[4–7] has emerged as a powerful technique owing to its compatibility with industrial-scale roll-to-roll technology[6]. However, the polycrystalline nature and microscopic roughness of Cu foils means that such roll-to-roll transferred films are not devoid of cracks and folds[6,7]. High-fidelity transfer or direct growth of high-quality graphene films on arbitrary substrates is needed to enable wide-ranging applications in photonics or electronics, which include devices such as optoelectronic modulators, transistors, on-chip biosensors and tunnelling barriers[3,8,9]. The direct growth of graphene film on an insulating substrate, such as a $SiO_2$/Si wafer, would be useful for this purpose, but current research efforts remain grounded at the proof-of-concept stage, where only discontinuous, nanometre-sized islands can be obtained[10]. Here we develop a face-to-face transfer method for wafer-scale graphene films that is so far the only known way to accomplish both the growth and transfer steps on one wafer. This spontaneous transfer method relies on nascent gas bubbles and capillary bridges between the graphene film and the underlying substrate during etching of the metal catalyst, which is analogous to the method used by tree frogs to remain attached to submerged leaves[11,12]. In contrast to the previous wet[4,5,13–15] or dry[6,7] transfer results, the face-to-face transfer does not have to be done by hand and is compatible with any size and shape of substrate; this approach also enjoys the benefit of a much reduced density of transfer defects compared with the conventional transfer method. Most importantly, the direct growth and spontaneous attachment of graphene on the underlying substrate is amenable to batch processing in a semiconductor production line, and thus will speed up the technological application of graphene.**

Much effort has been directed to controlling the grain size[15,16], doping[17] and heterostructure[18] of graphene by fine-tuning growth conditions during chemical vapour deposition (CVD), but there has been a lack of breakthroughs in the after-growth transfer process of the graphene film[3]. Conventional transfer methods for CVD graphene can be classified as either dry[6,7] or wet[4,5,13–15] transfer, depending on the environment in which graphene touches the target substrate. The dry transfer method seems to be more applicable to industrial applications, because it can realize a 30-inch graphene film on a flexible substrate[6], but plenty of transfer defects occur—for example, cracks, folds and wrinkles[6,7]. The wet transfer method is generally difficult to scale up, and the surface tension experienced by the floating graphene at the air–water interface causes warping, rippling and rolling of the films during transfer.

The inspiration for our face-to-face method is found in the study of how the feet of a terrestrial beetle or tree frog remain attached to a fully submerged leaf[11,12]. Microscopic observations reveal that air bubbles that are trapped around the feet of the beetle form capillary bridges and keep the beetle's feet attached to the submerged leaf. In a similar fashion, the formation of capillary bridges, as indicated schematically in Fig. 1a, ensures that the graphene film remains attached to the substrate and does not undergo delamination during the etching process (Fig. 1b). During the etching of the copper film between the graphene and the $SiO_2$/Si substrate, dissolution of the copper generates voids and channels, and these create capillary forces that allow the liquid etchant to infiltrate between the graphene film and the substrate. In the case of hydrophobic graphene surfaces[19], the unfavourable interactions between the water molecules and the soft graphene film induce instability of the planar interface, which leads to fluctuations of the water interface and spontaneous cavitation. Strong negative pressure will operate in short capillary bridges, resulting in long-ranged attraction between the contacting surfaces. In fact, such capillary adhesion forces can be larger than the van der Waals interaction between the two solids. We suggest that the evolution of bubbles during the etching process contributes to the formation of capillary bridges between the graphene–substrate interfaces, thus allowing the graphene film to remain attached to the substrate even with the infiltration of liquid. However, gas bubbles can also generate buoyancy forces that will separate the graphene film from the underlying substrate, and this is in fact the situation exploited in the conventional float transfer process. Therefore, the fate of the graphene film—with regard to either delamination from, or adhesion to, the substrate—is decided by whether a sufficient number of capillary bridges can be formed between the graphene and the substrate to counteract the pull-off forces due to buoyancy.

The presence of gas bubbles under the graphene can be seen by atomic force microscopy (AFM) imaging of the surface after CVD growth. Inductively coupled plasma CVD (ICP-CVD), required for the growth of graphene (see Methods), produces energetic ions that can be absorbed by the metal catalyst at high temperature. As the temperature is reduced at the end of the growth process, some gas precipitates at the surface and gets trapped between the impermeable graphene[20] and the substrate, forming bubbles. These bubbles alter the adhesion and strain properties of graphene, which are manifested as bright spots in the AFM phase contrast images (Supplementary Fig. 2). To facilitate the formation of capillary bridges, a pre-treatment step involving plasma nitridation of the $SiO_2$/Si wafer is helpful. Our experiments showed that without this treatment step, the graphene film delaminates at the end of the etching process. The nitrogen plasma treatment converts the top several nanometres of the surface to silicon oxynitride phases[21], and these decompose readily during the CVD process at greater temperature and act as an additional source of bubbles under the graphene. After the plasma nitridation of the substrate, Cu catalyst is sputtered onto the surface and growth of graphene is performed by ICP-CVD. The graphene/Cu/$SiO_2$/Si wafer is then coated with polymethyl methacrylate (PMMA) for protection and immersed in an aqueous etchant bath. The Cu film is then etched (Supplementary Figs 3 and 4). Throughout, the PMMA/graphene film adheres firmly to the underlying wafer. It is worth pointing out that our face-to-face transfer process has a shorter Cu etching time than does conventional float transfer (Fig. 2a). The Cu film can usually be completely dissolved within 2 h, even in dilute etchant.

To estimate the thickness of the infiltrated water layer, we produced an 8-inch $SiO_2$/Si wafer onto which a graphene layer had been transferred by our face-to-face process; this wafer, with the layer structure

[1]Graphene Research Centre, National University of Singapore, 6 Science Drive 2, 117546 Singapore. [2]Department of Chemistry, National University of Singapore, 3 Science Drive 3, 117543 Singapore. [3]Department of Physics, National University of Singapore, 2 Science Drive 3, 117542 Singapore.

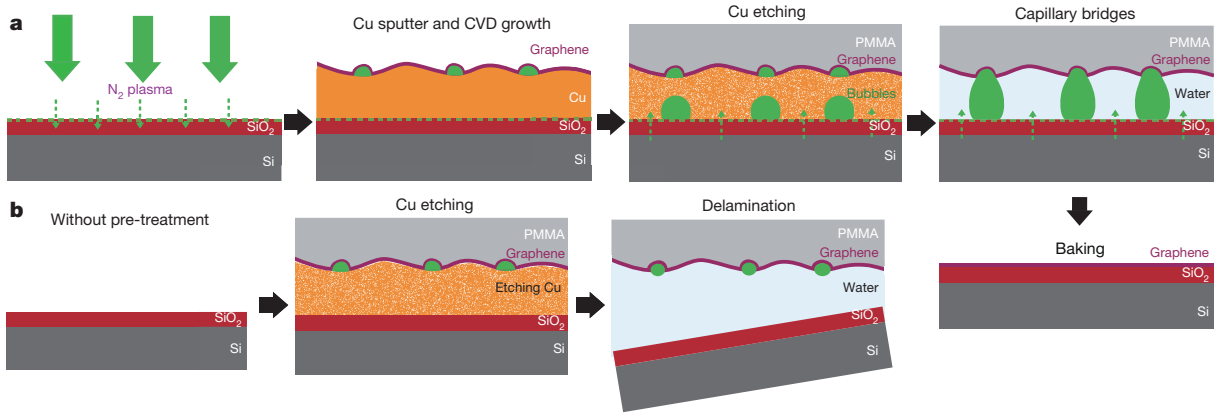



**Figure 1 | Illustration of our face-to-face method for transferring graphene mediated by capillary bridges.** **a**, Schematic illustration showing 'bubble seeding' by plasma treatment, CVD growth, Cu film etching, formation of capillary bridges and removal of water and PMMA. See text for details. **b**, Schematic illustration showing that in the absence of plasma treatment, delamination of the film results.

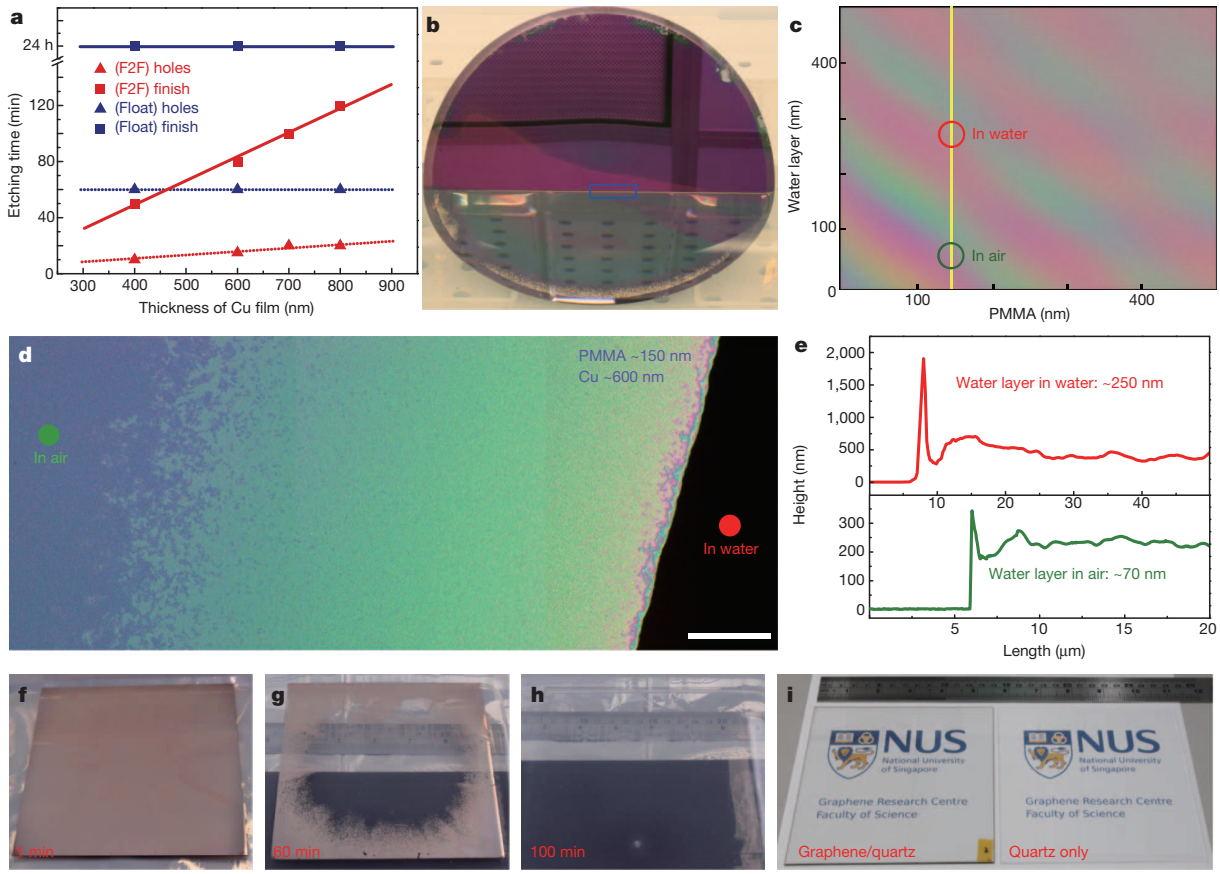



**Figure 2 | Characterization of intercalated water layer with capillary bridges during face-to-face transfer.** **a**, Comparison of etching time between face-to-face (F2F) and float transfer. The triangles indicate the time when visible holes in the copper film appear, and the squares indicate the time of complete etching of the Cu film. **b**, Photograph of a partially submerged 8-inch PMMA/graphene/water layer/$SiO_2$/Si wafer in water. The blue box shows the border between water/PMMA/water and air/PMMA/water interfaces. **c**, Simulated colour change of the PMMA/graphene/water layer/$SiO_2$/Si wafer placed in air. Red and green circles indicate sampling positions for data in **e**. The vertical yellow line indicates the relevant data for different thicknesses of the water layer and a fixed thickness of 150 nm for the PMMA layer, which is used here. **d**, Optical image of PMMA/graphene at the water–air border. The red dot identifies the bulk water. Scale bar, 500 μm. **e**, AFM profiles of thicknesses of PMMA/graphene with infiltrated water layer, measured in water (top) and in air (bottom). **f**–**h**, Photographs of PMMA/graphene/Cu (700 nm)/quartz (14 cm) immersed in Cu etchant for different times (given in red at bottom of panels), showing that the face-to-face transfer also works on quartz. **i**, Left, photograph of face-to-face transferred graphene on quartz after removal of PMMA and baking. Right, quartz only.

PMMA/graphene/water layer/$SiO_2$/Si, was pulled out partially from water for optical characterization (Fig. 2b). The combined effects of reflection and refraction of light at the multiple interfaces can be simulated using the CIE (International Commission on Illumination) colour-matching equation to reproduce the colour change[22,23] (Fig. 2c; the method is illustrated in Supplementary Fig. 5). The observed colour change of the PMMA/graphene/water layer/$SiO_2$/Si system is due to the change in thickness of the water layer under the graphene as it escapes, which generates coloured fringes (Fig. 2d). The PMMA/graphene floating on infiltrated water at the air–water border is pale red in colour (right) at first, and changes into olive, green and violet as the thickness of water layer decreases. On the basis of the simulated colour changes, the thickness of the intercalated water layer is estimated to be 60–80 nm in air (violet) and 250–300 nm in water (pale red). To confirm this measurement, the thickness of intercalated water layer when the wafer is immersed in water is determined by an AFM that can perform measurements under water, and its thickness in air after some water escapes is measured by a fast-scan AFM (Supplementary Fig. 6). Figure 2e shows that the mean thickness of the water layer when the wafer is under water has a mean value of ~250 nm, and that the thickness in air is ~70 nm. These values agree with the simulation in Fig. 2c, marked by circles. It is worth noting that the change in height profile is reversible when the wafer is immersed into water or withdrawn from it again; this implies that water can infiltrate freely in the intercalated layer between graphene and the substrate. The infiltrated water can be completely evaporated by baking the wafer at 150 °C for more than 10 min, producing a dry graphene film on $SiO_2$/Si. The face-to-face transfer also works on quartz substrates, as shown in Fig. 2f–i for a 700-nm sputtered Cu film on a 14 cm × 14 cm quartz plate.

The Greenwood–Williamson contact mechanics theory adapted to randomly rough surfaces in close contact can be used to estimate the maximal height of the capillary bridges formed at the interface[24,25]. PMMA/graphene can be considered an elastic soft surface in contact with a hard and rough substrate (Cu/$SiO_2$/Si). In the contact region between the two surfaces, a liquid capillary bridge will form. The meniscus radius $r_K$ is given by the Kelvin equation, $r_K = -\gamma v_0/(k_B T)\ln(P_v/P_{sat})$, where $\gamma$ is the surface tension, $P_v$ is the actual vapour pressure, $P_{sat}$ is the saturated vapour pressure, $v_0$ is the molar volume of water, $k_B$ is Boltzmann's constant and $T$ is temperature, and the thickness of the water is given by $d_K = r_K(\cos\theta_1 + \cos\theta_2)$, where $\theta_1$ and $\theta_2$ are the contact angles of the two surfaces[25].

When the PMMA/graphene is almost delaminating from the wafer surface because of a pull-off force, $d_K$ becomes largest. At this point, the work of adhesion ($w$; equation (1)) will be decreased to zero, and the maximal height of the capillary bridge ($d_c$) can be estimated from equation (2)[25]:

$$w \approx 2\gamma\left[1 - \frac{h}{2d_K}\ln\left(\frac{q_0 h E d_K}{2\gamma}\right)\right] \quad (1)$$

$$d_c = \frac{h}{2}\ln\left(\frac{q_0 h E d_c}{2\gamma}\right) \quad (2)$$

Here $q_0$ is the roll-off wavevector of the surface roughness power spectrum, $h$ is the roughness and $E$ is the Young's modulus of the film. Accordingly, $d_c$ is calculated to be in the range 300–450 nm by assuming a roughness of $h$ = 100–150 nm for the PMMA/graphene (root mean squared roughness as determined by AFM). The significance of $d_c$ is that it sets an upper limit on the thickness of the Cu films that can be sputtered onto the $SiO_2$/Si surface, because a thick Cu film will result in the infiltration of a thick water layer that exceeds the threshold thickness $d_c$. As observed in our experiments, if the thickness of the Cu film is three times larger than $d_c$, then delamination of the PMMA/graphene film occurs as a result of the infiltration of an excessively thick water layer. Under typical conditions, the thickness of the infiltrated water is about one-third the thickness of the Cu film because hydrostatic pressure compresses the film after the removal of the Cu. Owing to the fact that PMMA/graphene is a soft membrane, it can deform elastically in response to the capillary forces caused by the gas bridges, and this allows the film to be pulled closer to the hard substrate.

The high surface tension of water exerts a pulling force on the graphene, and undesirable ripples and folds are created. The surface tension can be reduced by adding isopropyl alcohol or by increasing the temperature of the water (to 80 °C). Figure 3a is a typical AFM image of a face-to-face graphene film on a $SiO_2$/Si wafer, where a high density of graphene nanobubbles with heights up to ~20 nm can be seen. After the addition of isopropyl alcohol (Fig. 3b) or increasing water temperature (Fig. 3c), there is a significant reduction in the density of graphene wrinkles, and the height of the nanobubbles is reduced by a factor of at least ten, resulting in a visible flattening of the film. Therefore, the addition of a surfactant into the water can help to 'iron out' the creases on the graphene during the face-to-face transfer process.

Figure 4a shows a photograph of face-to-face transferred graphene on $SiO_2$/Si wafers after removing PMMA (both a 4-inch and an 8-inch wafer are shown). The transferred graphene film appears to be highly uniform, with no visible transfer defects such as cracks and folds, and its torn edges can be observed by the naked eye (Fig. 4a inset). Raman spectroscopy is a commonly applied technique for characterizing the properties of graphene in terms of its layer number, defects and doping[26,27]. The Raman spectra of graphene transferred by face-to-face and, respectively, conventional float transfer are shown in Fig. 4b, displayed together with the spectrum of graphene on a Cu film as a reference. Although as-grown graphene before transfer shows nearly no defect-related D band (at ~1,350 $cm^{-1}$), both face-to-face and float-transferred graphene show small D peaks due to the Raman signal enhancement on a dielectric substrate[28]. The 2D band (at ~2,690 $cm^{-1}$), which is due to two phonons with opposite momenta in the highest optical branch, is

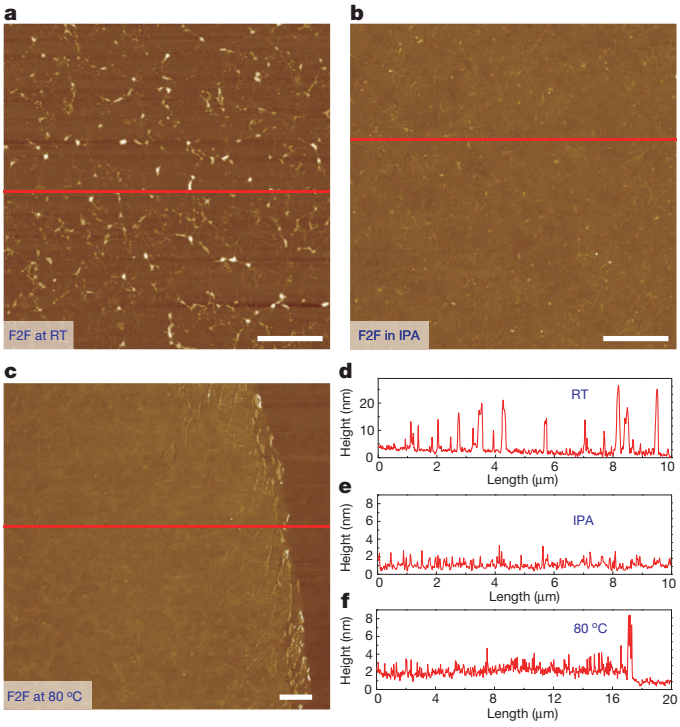



**Figure 3 | AFM images and height profiles of graphene on a $SiO_2$/Si wafer transferred by our face-to-face technique.** **a–c**, AFM images of films transferred in room-temperature (RT) water (**a**); in room-temperature isopropyl alcohol (IPA; **b**); and in water at 80 °C (**c**). **d–f**, Corresponding height profiles for **a–c**, respectively. The red horizontal line in **a–c** shows the related profile positions for **d–f.** The density of the nanobubbles decreases after reducing the liquid surface tension in **b** and **c**. Scale bars, 2 μm.

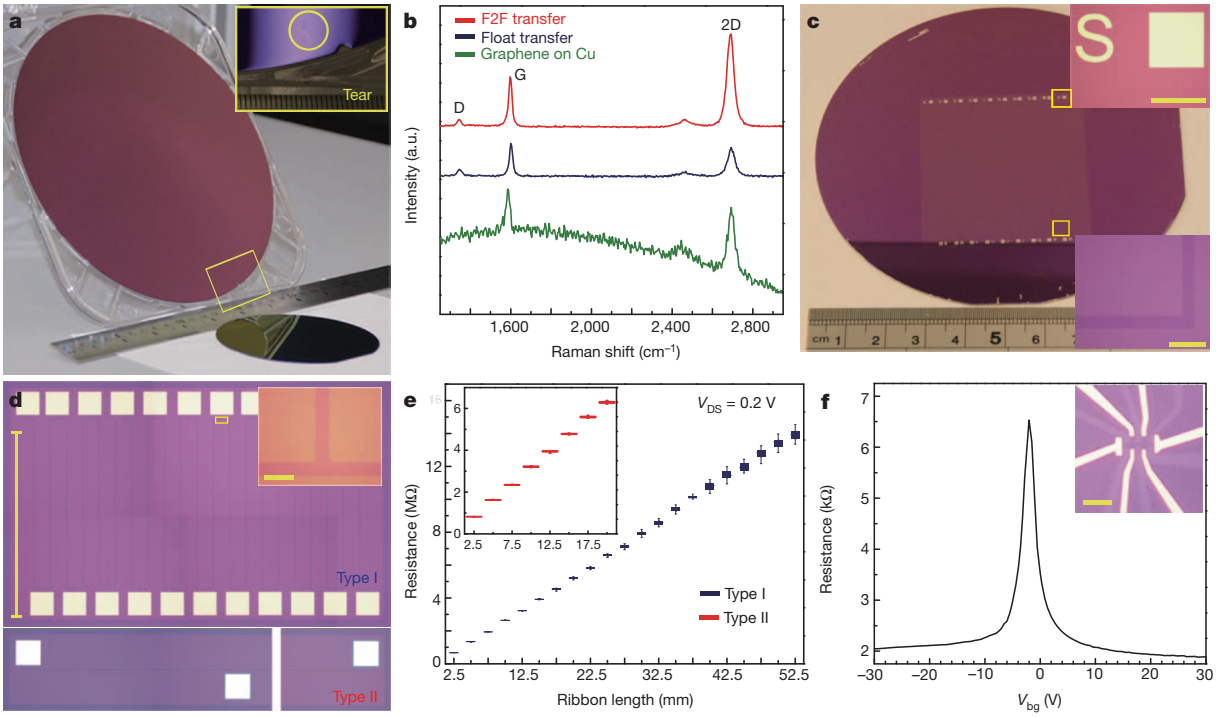



**Figure 4 | Characterization of face-to-face transferred graphene on a $SiO_2$/Si wafer.** **a**, Photograph of face-to-face transferred graphene on an 8-inch wafer and a 4-inch wafer; the boxed area is shown enlarged in the inset, where a torn edge shows the presence of a graphene sheet. **b**, Raman spectra of graphene prepared by face-to-face transfer (red) and by float transfer onto $SiO_2$/Si substrates (blue), and of graphene on Cu film before transfer (green). a.u., arbitrary units. **c**, Photograph of unbroken 1-m-long and 15-μm-wide zigzag graphene ribbon on face-to-face transferred graphene wafer. Insets at top and bottom right are magnified optical images from the positions marked by yellow boxes. 'S' is a location marker for the metal electrode on the ribbons. **d**, Optical image of a graphene ribbon 5 cm long and 10 μm wide. The graphene ribbon is fabricated in a zigzag configuration (top panel, type I) and as a straight ribbon (bottom panel, type II); inset at top right is the zoomed optical image for type I, from the position marked by the yellow box. **e**, Box plot showing resistance versus length for graphene ribbons of type I and type II. The linear relationship between resistance and length shows the long-distance continuity of the ribbons. All the measurements were performed at a voltage ($V_{DS}$) of 0.2 V between two electrodes. **f**, Carrier mobility for a monolayer graphene Hall bar device under ambient conditions; the Hall effect mobility of this device is ~3,800 $cm^2 V^{-1} s^{-1}$. Inset, optical image of the related graphene Hall bar device. $V_{bg}$ is the gate voltage during the measurement. Scale bars: **d** (main panel), 2.5 mm; in insets of **c**, **d** and **f**, 1 mm (top right), 20 μm (bottom right), 20 μm and 5 μm, respectively.

more sensitive to long-range order. Here it can be seen that the intensity of the 2D band is much higher in face-to-face than float-transferred graphene, which attests to the higher crystalline quality of the former. More characterization of the as-grown graphene is shown in Supplementary Figs 8 and 9. X-ray photoelectron spectroscopy is used to check the phase purity of face-to-face transferred graphene film on a $SiO_2$/Si wafer (Supplementary Fig. 11). There is no detectable Cu on the surface within the sensitivity of the technique. Residual N can be detected on the wafer (at 398 eV) but this is distinct from N dopants in graphene[29] and is attributed to implanted N species in the silicon oxynitride phase[21].

The electrical properties of face-to-face transferred graphene films are investigated by standard micro/nanofabrication processes. To test the robustness of this transfer process with regard to crack-free and continuous films, long graphene ribbons were fabricated (up to 1 m long and 15 μm wide in Fig. 4c; 5 cm long and 10 μm wide in Fig. 4d; details in Supplementary Information). The resistivity is measured by a standard four-probe station after applying a 50 nm Ni layer as metal electrodes. As shown in Fig. 4e, the measured resistances exhibit linearity with the length of the ribbon, and a conductivity of ~4,000 S $cm^{-1}$ can be achieved, thus attesting to the long-distance continuity of the film after transfer. To the best of our knowledge, continuous, uninterrupted graphene ribbons with length/width ratios of more than $10^5$ are seldom demonstrated, because cracks are inevitable in the wet or dry transfer processes; thus, the face-to-face transfer method developed here is unique in enabling high-fidelity, crack-free transfer. To evaluate the electronic quality of the face-to-face transferred graphene grown by ICP-CVD at 750 °C, standard electron-beam lithography is used to fabricate graphene Hall bars on a Si wafer with a 290-nm $SiO_2$ layer. A typical optical image of these graphene Hall bars is shown Fig. 4f, inset. Four metal electrodes of Cr/Au (5/50 nm) are used to eliminate the contact resistance, and the transport characteristics of the device are measured under ambient conditions. The transport result for this device is shown in Fig. 4f, and the extracted carrier mobilities of electrons and holes for this device are both ~3,800 $cm^2 V^{-1} s^{-1}$, which is comparable to the properties of thermal CVD graphene grown at much higher temperature[4,6,17,30]. The slightly negative Dirac point indicates that the transferred graphene is weakly n-type doped[17].

All the characterization above shows that the face-to-face transferred films maintain good crystalline integrity and long-distance continuity without cracks. The copper catalyst can be effectively removed and the carrier mobility of the film is comparable to that of thermal CVD-grown film prepared at higher temperatures. The key advantages of this face-to-face transfer are that it is relatively easy and requires only a simple pretreatment step followed by immersion in a suitable etchant; it resembles a spontaneous transfer process because there is no recovery of floating graphene needed; and, most importantly, the non-manual and wafer-compatible nature of the method suggests that it is automation-compatible and industrially scalable. Interestingly, we have found that water has the ability to infiltrate between the graphene and the wafer substrate, thus allowing the addition of different surfactants for modifying the interfacial tension and reducing the corrugations in graphene film. Although there are many potential applications of the roll-to-roll transfer method[6] owing to its applicability to flexible devices, it must be noted that so far most devices operate on 'stiff' substrates such as silicon, and that a non-manual, batch-processed transfer method serving this technology segment is definitely needed. The face-to-face transfer

method will be very useful as an enabler for rapidly emerging graphene-on-silicon platforms that have shown excellent promise for devices such as a gate-controlled Schottky barrier triode device[8] and an optical modulator[9]. Finally, the face-to-face transfer method should be applicable to all CVD growth on a metal-catalyst-coated wafer, such as hexagonal BN and transition-metal chalcogenide films.

## METHODS SUMMARY

The catalyst preparation and ICP-CVD growth are both performed in a customized sputter/electron-beam/ICP-CVD cluster. First, 8-inch or 4-inch $SiO_2$/Si (or quartz) wafers are sputtered with Cu films at 100–200 °C. For face-to-face transfer, the wafers are pre-treated with $N_2$ plasma (1,000 W, 50 mtorr). The conditions for ICP-CVD growth are as follows: the wafer with a sputtered Cu film is treated with a $H_2$ plasma (150 W) for 5 min with the substrate heated to 750 °C at 50 mtorr. A mixture of $H_2$ and $CH_4$ ($H_2$:$CH_4$ = 150:10) is introduced into the chamber for graphene growth (150 W plasma power, 50 mtorr, 5 min). After growth, the graphene/Cu/wafer is spin-coated with PMMA (996,000 relative molecular mass, 4 wt% in ethyl lactate, 3,000 r.p.m. for 1 min) for protection. A 0.1 M ammonium persulphate ($(NH_4)_2S_2O_8$) aqueous solution or 1 M iron chloride ($FeCl_3$) is used as the etchant. Baking at 150 °C for 10 min is needed to evaporate the water layer. Finally, face-to-face transferred graphene on the wafer is realized by removing the PMMA with acetone.

Raman spectra of graphene films are collected using a 532-nm laser under ambient conditions (WITec alpha 300 R). The AFM measurements were performed in tapping mode (Bruker Dimension FastScan), and a liquid AFM was used to measure the thickness of PMMA/graphene in water (Agilent 5420). All coloured optical images are captured with a Nikon Eclipse LV100D. Long graphene ribbons are fabricated by a laser writer (MicroTech LW405B), and graphene Hall bars are patterned by electron beam lithography (Nova NanoSEM 230). Electron-beam-evaporated 50-nm Ni films are used as contacts for the long ribbons, and thermally evaporated 5-nm Cr and 50-nm Au films are used as metal contacts for the Hall bars. Current–voltage curves are measured by a four-probe station under ambient conditions (Keithley 4200 SCS and 6430).

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 8 July; accepted 10 October 2013.**
**Published online 11 December 2013.**

1. Novoselov, K. S. *et al.* Electric field effect in atomically thin carbon films. *Science* **306,** 666–669 (2004).
2. Novoselov, K. S. *et al.* Two-dimensional atomic crystals. *Proc. Natl Acad. Sci. USA* **102,** 10451–10453 (2005).
3. Novoselov, K. S. *et al.* A roadmap for graphene. *Nature* **490,** 192–200 (2012).
4. Li, X. S. *et al.* Large-area synthesis of high-quality and uniform graphene films on copper foils. *Science* **324,** 1312–1314 (2009).
5. Gao, L. B. *et al.* Efficient growth of high-quality graphene films on Cu foils by ambient pressure chemical vapor deposition. *Appl. Phys. Lett.* **97,** 183109 (2010).
6. Bae, S. K. *et al.* Roll-to-roll production of 30-inch graphene films for transparent electrodes. *Nature Nanotechnol.* **5,** 574–578 (2010).
7. Kang, J. *et al.* Efficient transfer of large-area graphene films onto rigid substrates by hot pressing. *ACS Nano* **6,** 5360–5365 (2012).
8. Yang, H. *et al.* Graphene barristor, a triode device with a gate-controlled Schottky barrier. *Science* **336,** 1140–1143 (2012).
9. Liu, M. *et al.* A graphene-based broadband optical modulator. *Nature* **474,** 64–67 (2011).
10. Chen, J. Y. *et al.* Oxygen-aided synthesis of polycrystalline graphene on silicon dioxide substrates. *J. Am. Chem. Soc.* **133,** 17548–17551 (2011).
11. Federle, W., Barnes, W. J. P., Baumgartner, W., Drechsler, P. & Smith, J. M. Wet but not slippery: boundary friction in tree frog adhesive toe pads. *J. R. Soc. Interface* **3,** 689–697 (2006).
12. Persson, B. N. J. Wet adhesion with application to tree frog adhesive toe pads and tires. *J. Phys. Condens. Matter* **19,** 376110 (2007).
13. Reina, A. *et al.* Large area, few-layer graphene films on arbitrary substrates by chemical vapor deposition. *Nano Lett.* **9,** 30–35 (2009).
14. Kim, K. S. *et al.* Large-scale pattern growth of graphene films for stretchable transparent electrodes. *Nature* **457,** 706–710 (2009).
15. Gao, L. B. *et al.* Repeated growth and bubbling transfer of graphene with millimetre-size single-crystal grains using platinum. *Nature Commun.* **3,** 699 (2012).
16. Li, X. S. *et al.* Large-area graphene single crystals grown by low-pressure chemical vapor deposition of methane on copper. *J. Am. Chem. Soc.* **133,** 2816–2819 (2011).
17. Sun, Z. Z. *et al.* Growth of graphene from solid carbon sources. *Nature* **468,** 549–552 (2010).
18. Levendorf, M. P. *et al.* Graphene and boron nitride lateral heterostructures for atomically thin circuitry. *Nature* **488,** 627–632 (2012).
19. Rafiee, J. *et al.* Wetting transparency of graphene. *Nature Mater.* **11,** 217–222 (2012).
20. Bunch, J. S. *et al.* Impermeable atomic membranes from graphene sheets. *Nano Lett.* **8,** 2458–2462 (2008).
21. Kobayashi, H., Mizokuro, T., Nakato, Y., Yoneda, K. & Todokoro, Y. Nitridation of silicon oxide layers by nitrogen plasma generated by low energy electron impact. *Appl. Phys. Lett.* **71,** 1978–1980 (1997).
22. Gao, L. B., Ren, W. C., Li, F. & Cheng, H. M. Total color difference for rapid and accurate identification of graphene. *ACS Nano* **2,** 1625–1633 (2008).
23. Janos, S. *Colorimetry: Understanding the CIE System* 25–88 (Wiley, 2007).
24. Greenwood, J. A. & Williamson, J. B. P. Contact of nominally flat surfaces. *Proc. R. Soc. Lond. A* **295,** 300–319 (1966).
25. Persson, B. N. J. Capillary adhesion between elastic solids with randomly rough surfaces. *J. Phys. Condens. Matter* **20,** 315007 (2008).
26. Ferrari, A. C. *et al.* Raman spectrum of graphene and graphene layers. *Phys. Rev. Lett.* **97,** 187401 (2006).
27. Das, A. *et al.* Monitoring dopants by Raman scattering in an electrochemically top-gated graphene transistor. *Nature Nanotechnol.* **3,** 210–215 (2008).
28. Gao, L. B. *et al.* Surface and interference coenhanced Raman scattering of graphene. *ACS Nano* **3,** 933–939 (2009).
29. Wei, D. C. *et al.* Synthesis of N-doped graphene by chemical vapor deposition and its electrical properties. *Nano Lett.* **9,** 1752–1758 (2009).
30. Yoon, T. *et al.* Direct measurement of adhesion energy of monolayer graphene as-grown on copper and its application to renewable transfer process. *Nano Lett.* **12,** 1448–1452 (2012).

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

**Acknowledgements** We thank C. T. Nai for help with X-ray photoelectron spectroscopy and B. K. Chong (Agilent Technologies) for the liquid AFM. This work was supported by MOE Tier 2 grant 'Interface engineering of graphene hybrids for energy conversion' (R-143-000-488-112) and by NRF-CRP grant 'Novel 2D materials with tailored properties: beyond graphene' (R-144-000-295-281).

**Author Contributions** L.G. and K.P.L. designed the experiments, interpreted the data and wrote the manuscript. L.G. performed graphene growth, transfer and calculations. L.G., G.-X.N. and B.L. fabricated the devices. Y.L. performed the X-ray photoelectron spectroscopy measurements. L.G., K.P.L. and A.H.C.N. discussed the data.

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

LETTER

doi:10.1038/nature12909

# A metal–free organic–inorganic aqueous flow battery

Brian Huskinson[1]*, Michael P. Marshak[1,2]*, Changwon Suh[2], Süleyman Er[2,3], Michael R. Gerhardt[1], Cooper J. Galvin[2], Xudong Chen[2], Alán Aspuru-Guzik[2], Roy G. Gordon[1,2] & Michael J. Aziz[1]

**As the fraction of electricity generation from intermittent renewable sources—such as solar or wind—grows, the ability to store large amounts of electrical energy is of increasing importance. Solid-electrode batteries maintain discharge at peak power for far too short a time to fully regulate wind or solar power output[1,2]. In contrast, flow batteries can independently scale the power (electrode area) and energy (arbitrarily large storage volume) components of the system by maintaining all of the electro-active species in fluid form[3–5]. Wide-scale utilization of flow batteries is, however, limited by the abundance and cost of these materials, particularly those using redox-active metals and precious-metal electrocatalysts[6,7]. Here we describe a class of energy storage materials that exploits the favourable chemical and electrochemical properties of a family of molecules known as quinones. The example we demonstrate is a metal-free flow battery based on the redox chemistry of 9,10-anthraquinone-2,7-disulphonic acid (AQDS). AQDS undergoes extremely rapid and reversible two-electron two-proton reduction on a glassy carbon electrode in sulphuric acid. An aqueous flow battery with inexpensive carbon electrodes, combining the quinone/hydroquinone couple with the $Br_2/Br^-$ redox couple, yields a peak galvanic power density exceeding 0.6 W cm$^{-2}$ at 1.3 A cm$^{-2}$. Cycling of this quinone–bromide flow battery showed >99 per cent storage capacity retention per cycle. The organic anthraquinone species can be synthesized from inexpensive commodity chemicals[8]. This organic approach permits tuning of important properties such as the reduction potential and solubility by adding functional groups: for example, we demonstrate that the addition of two hydroxy groups to AQDS increases the open circuit potential of the cell by 11% and we describe a pathway for further increases in cell voltage. The use of π-aromatic redox-active organic molecules instead of redox-active metals represents a new and promising direction for realizing massive electrical energy storage at greatly reduced cost.**

Solutions of AQDS in sulphuric acid (negative side) and $Br_2$ in HBr (positive side) were pumped through a flow cell as shown schematically in Fig. 1a. The quinone–bromide flow battery (QBFB) was constructed using a Nafion 212 membrane sandwiched between Toray carbon paper electrodes (six stacked on each side) with no catalysts; it is similar to a cell described elsewhere (see figure 2 in ref. 7). We report the potential–current response (Fig. 1b) and the potential–power relationship (Fig. 1c and d) for various states of charge (SOCs; measured with respect to the quinone side of the cell). As the SOC increased from 10% to 90%, the open-circuit potential increased linearly from 0.69 V to 0.92 V. In the galvanic direction, peak power densities were 0.246 W cm$^{-2}$ and 0.600 W cm$^{-2}$ at these same SOCs, respectively (Fig. 1c). To avoid significant water splitting in the electrolytic direction, we used a cut-off voltage of 1.5 V, at which point the current densities observed at 10% and 90% SOCs were −2.25 A cm$^{-2}$ and −0.95 A cm$^{-2}$, respectively, with corresponding power densities of −3.342 W cm$^{-2}$ and −1.414 W cm$^{-2}$.

In Fig. 2 we report the results of initial cycling studies for this battery, to test for consistent performance over longer timescales. Figure 2a shows cycling data at ±0.2 A cm$^{-2}$ using 50% of the total capacity of the battery. The cycles are highly reproducible and indicate that current efficiencies for the battery are around 95%. Figure 2b shows constant-current cycling data, collected at ±0.5 A cm$^{-2}$, using voltage cut-offs of 0 V and 1.5 V. These tests were done using the identical solutions used in the battery for Fig. 1b–d. The galvanic discharge capacity retention (that is, the number of coulombs extracted in one cycle divided by the number of coulombs extracted in the previous cycle) is above 99%, indicating the battery is capable of operating with minimal capacity fade and suggesting that current efficiencies are actually closer to 99%. Full characterization of the current efficiency will require slower cycling experiments and chemical characterization of the electrolyte solutions after extended cycling.

To gain a better understanding of the quinone half-reaction on carbon, AQDS was subjected to half-cell electrochemical measurements. Cyclic voltammetry of a 1 mM solution of AQDS in 1 M sulphuric acid on a glassy carbon disk working electrode shows current peaks corresponding to reduction and oxidation of the anthraquinone species[9–11] (Fig. 3d, solid trace). The peak separation of 34 mV is close to the value of 59 mV/$n$, where $n$ is the number of electrons involved, expected for a two-electron process. Rotation of this disk at a variety of rates yields mass-transport-limited currents (Fig. 3a) from which the AQDS diffusion coefficient ($D = 3.8(1) \times 10^{-6}$ cm$^2$ s$^{-1}$) can be determined; throughout this paper, the numbers reported in parentheses indicate the standard deviation in the last reported digit. Koutecký–Levich analysis at low overpotentials (Fig. 3b) can be extrapolated to infinite rotation rate and fitted to the Butler–Volmer equation (Extended Data Fig. 3a) to give the kinetic reduction rate constant $k_0 = 7.2(5) \times 10^{-3}$ cm s$^{-1}$. This rate constant is greater than that found for other species used in flow batteries such as $V^{3+}/V^{2+}$, $Br_2/Br^-$ and $S_4^{2-}/S_2^{2-}$ (see table 2 in ref. 3). It implies that the voltage loss due to the rate of surface electrochemical reactions is negligible. The high rate is apparently due to an outer-sphere two-electron reduction into the aromatic π system requiring little reorganizational energy. The electrochemical reversibility of the two-electron redox reaction was confirmed by fitting the slope to the Butler–Volmer equation (Extended Data Fig. 3a), giving the transfer coefficient $\alpha = 0.474(2)$, which is close to the value of 0.5 expected for an ideally reversible reaction. The Pourbaix diagram (Extended Data Fig. 4) confirms that a two-electron, two-proton reduction occurs in acidic solution, and yields approximate p$K_a$ values of 7 and 11 for the reduced AQDS species[11].

Functionalization of the anthraquinone backbone with electron-donating groups such as hydroxy can be expected lower the reduction potential of AQDS ($E^0$), thereby raising the cell voltage[12]. Hydroxy-substituted anthraquinones are synthesized through oxidation reactions that may be performed at minimal cost. They are also natural products that have been extracted for millennia from common sources such as rhubarb and could even provide a renewable source for future anthraquinone-based electrolyte solutions.

Quantum chemical calculations of un-substituted and hydroxy-substituted AQDS were performed to predict how substitution patterns would change both $E^0$ of the quinone/hydroquinone couples (Fig. 3c)

[1]Harvard School of Engineering and Applied Sciences, 29 Oxford Street, Cambridge, Massachusetts 02138, USA. [2]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA. [3]Molecular Materials and Nanosystems, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands.
*These authors contributed equally to this work.

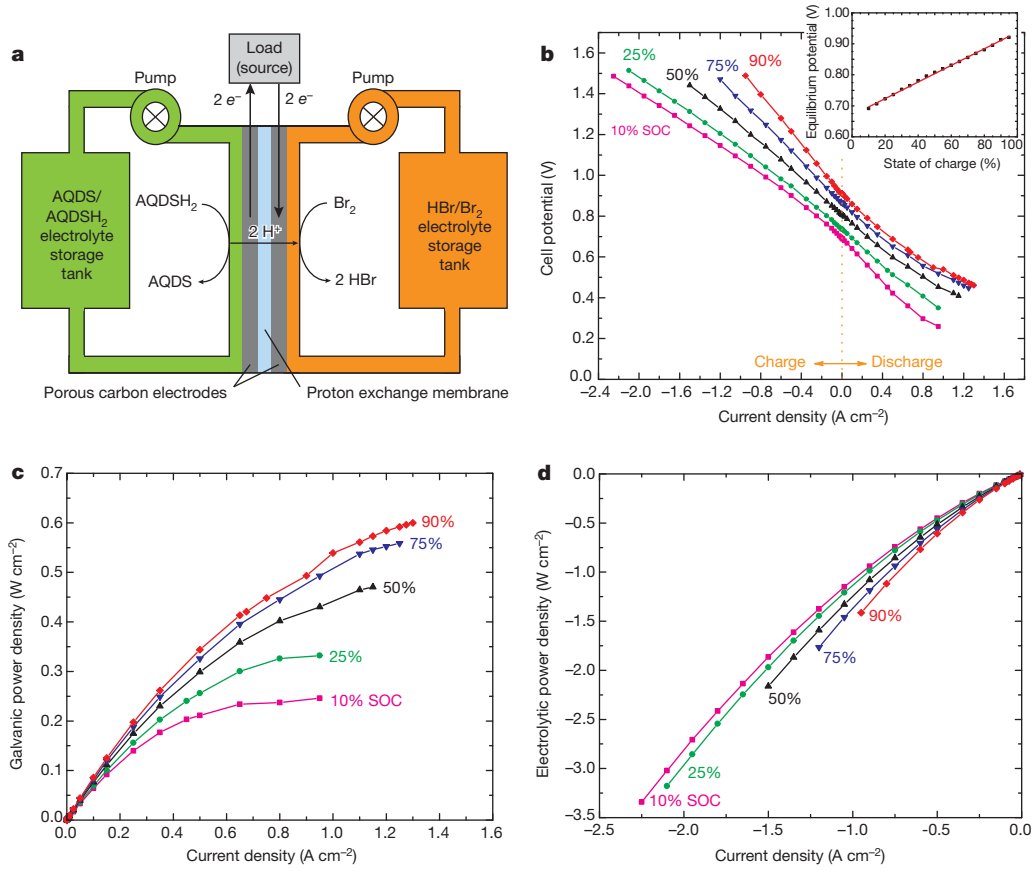



**Figure 1 | Cell schematic and cell performance in galvanic and electrolytic modes.** **a**, Cell schematic. Discharge mode is shown; the arrows are reversed for electrolytic/charge mode. $AQDSH_2$ refers to the reduced form of AQDS. **b**, Cell potential versus current density at five different states of charge (SOCs; average of three runs); inset shows the cell open circuit potential versus SOC with best-fit line superimposed ($E_{eq} = (0.00268 \times SOC) + 0.670$; $R^2 = 0.998$). **c**, Galvanic power density versus current density for the same SOCs. **d**, Electrolytic power density versus current density. All data here were collected at 40 °C using a 3 M HBr + 0.5 M $Br_2$ solution on the positive side and a 1 M AQDS + 1 M $H_2SO_4$ solution on the negative side.

and the solvation free energy ($G^0_{solv}$) in aqueous solution (Extended Data Table 1). The addition of −OH groups is calculated to lower the $E^0$ by an average of −50 mV per −OH and provide a wide window for tuning $E^0$ by almost 0.6 V. In addition, increasing numbers of hydroxy substituents are expected to raise the aqueous solubility due to hydrogen bonding.

In confirmation of the theory, the experimental reduction potential of 1,8-dihydroxy-9,10-anthraquinone-2,7-disulphonic acid (DHAQDS) was found to be 118 mV (versus the standard hydrogen electrode), which is very close to the 101 mV calculated for this species (Fig. 3c and d). The experimental $E^0$ of DHAQDS was 95 mV lower than AQDS, and would result in an 11% increase in QBFB cell potential.

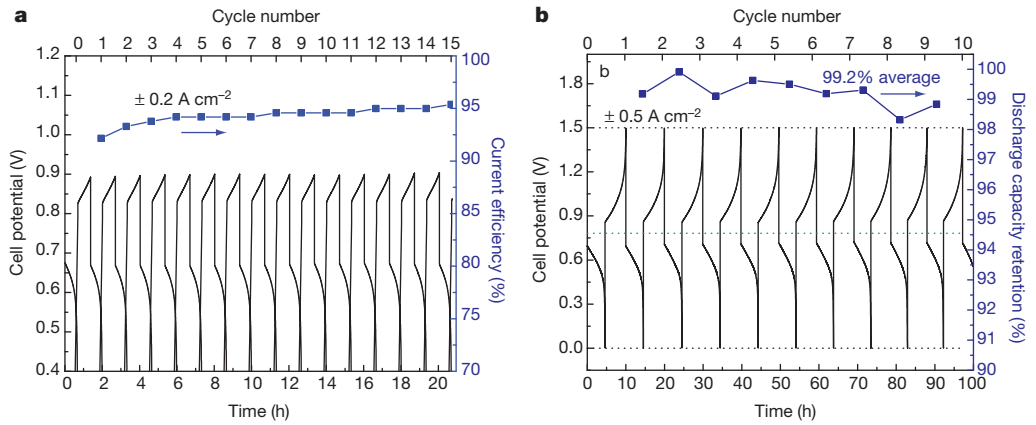



**Figure 2 | Cell cycling behaviour.** **a**, Constant-current cycling at 0.2 A $cm^{-2}$ at 40 °C using a 2 M HBr + 0.5 M $Br_2$ solution on the positive side and a 0.1 M AQDS + 2 M $H_2SO_4$ solution on the negative side; current efficiency is indicated for each complete cycle. **b**, Constant-current cycling at 0.5 A $cm^{-2}$ at 40 °C using a 3 M HBr + 0.5 M $Br_2$ solution on the positive side and a 1 M AQDS + 1 M $H_2SO_4$ solution on the negative side (same solution used in Fig. 1); discharge capacity retention is indicated for each cycle.

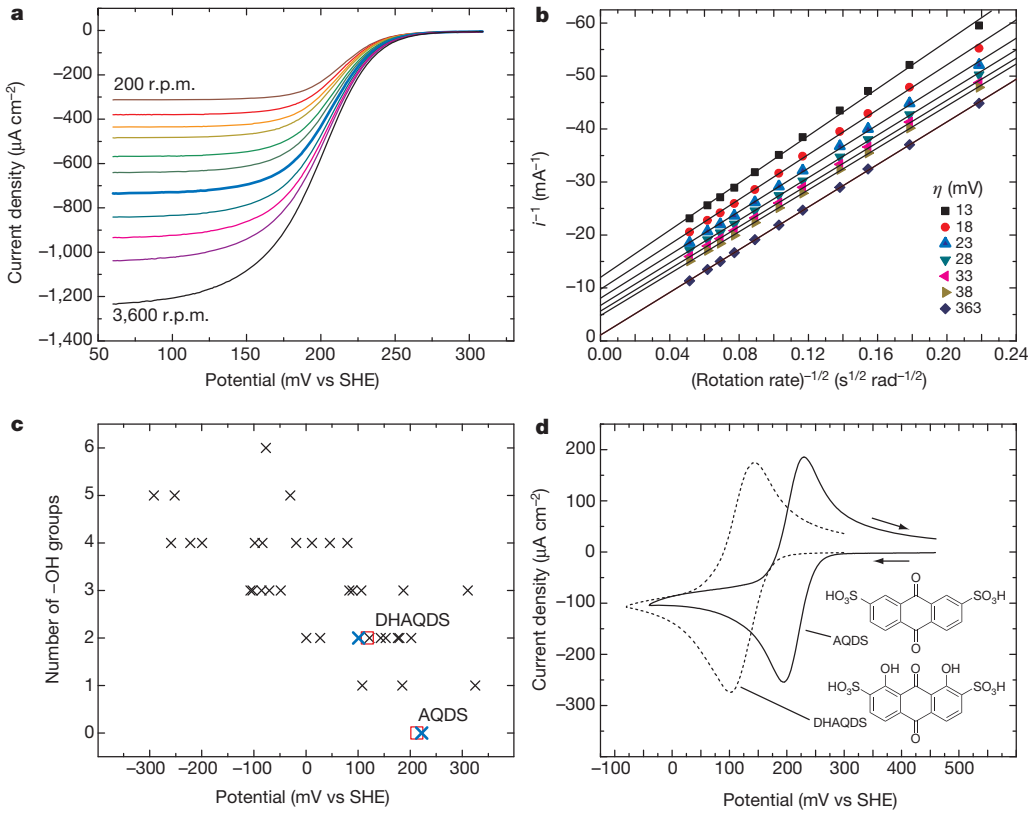



**Figure 3 | Half-cell measurements and theory calculations.** **a**, Rotating disk electrode (RDE) measurements of AQDS using a glassy carbon electrode in 1 M $H_2SO_4$ at 11 rotation rates ranging from 200 r.p.m. (red) to 3,600 r.p.m. (black). **b**, Koutecký–Levich plot (current$^{-1}$ versus rotation rate$^{-1/2}$) derived from **a** at seven different AQDS reduction overpotentials, $\eta$. **c**, Calculated reduction potentials of AQDS substituted with –OH groups (black), calculated AQDS and DHAQDS values (blue), and experimental values for AQDS and DHAQDS (red squares). **d**, Cyclic voltammogram of AQDS and DHAQDS (1 mM) in 1 M $H_2SO_4$ on a glassy carbon electrode (scan rate = 25 mV s$^{-1}$).

DHAQDS was also found to have faster reduction kinetics ($k_0 = 1.56(5) \times 10^{-2}$ cm s$^{-1}$), possibly due to intramolecular hydrogen bonding of the −OH to the ketone (Extended Data Fig. 3b).

The organic approach liberates battery redox chemistry from the constraints of the limited number of elemental redox couples of the periodic table. Although quinones have been used previously in batteries using redox-active solids[13–15], their incorporation into all-liquid flow batteries offers the following advantages over current flow-battery technologies. First, scalability: AQDS contains only the Earth-abundant atoms carbon, sulphur, hydrogen and oxygen, and can be inexpensively manufactured on large scales. Because some hydroxy-anthraquinones are natural products, there is also the possibility that the electrolyte material can be renewably sourced. Second, kinetics: quinones undergo extremely rapid two-electron redox on simple, inexpensive carbon electrodes and do not require a costly precious-metal catalyst. Furthermore, this electrode permits higher charging voltages by suppressing the parasitic water-splitting reactions. Third, stability: quinones should exhibit minimal membrane crossover owing to their relatively large size and charge in aqueous solution as a sulphonate anion. Furthermore, although bromine crossover is a known issue in zinc-bromine, vanadium-bromine and hydrogen-bromine cells, AQDS is stable to prolonged heating in concentrated $Br_2$/HBr mixtures (Extended Data Figs 5 and 6), and the QBFB can be cycled in HBr electrolyte solutions (Extended Data Fig. 9). Fourth, solubility: AQDS has an aqueous solubility greater than 1 M at pH 0, and the quinone solution can thus be stored at relatively high energy density—volumetric and gravimetric energy densities exceed 50 W h l$^{-1}$ and 50 W h kg$^{-1}$, respectively. Last, tunability: the reduction potential and solubility of AQDS can be further optimized by introduction of functional groups such as −OH. Use of DHAQDS is expected to lead to an increase in cell potential, performance and energy density.

These features lower the capital cost of storage chemicals per kilowatt hour, which sets a floor on the ultimate system cost per kilowatt hour at any scale. The precursor molecule anthracene is abundant in crude petroleum and is already oxidized on large scale to anthraquinone. Sulphonated anthraquinones are used on an industrial scale in wood pulp processing for paper[16], and they can be readily synthesized from the commodity chemicals anthraquinone and oleum[8]. In fact, a cyclic voltammogram of the crude sulphonation product of these two reagents is virtually identical to that of pure AQDS (Extended Data Fig. 8). Based on this simple electrolyte preparation that requires no further product separation, we estimate chemical costs of $21 per kilowatt hour for AQDS and $6 per kilowatt hour for bromine[17] (see Methods for information on cost calculations). The QBFB offers major cost improvements over vanadium flow batteries with redox-active materials that cost $81 per kilowatt hour (ref. 18). Optimization of engineering and operating parameters such as the flow field geometry, electrode design, membrane separator and temperature—which have not yet even begun—should lead to significant performance improvements in the future, as it has for vanadium flow batteries, which took many years to reach the power densities we report here[6]. The use of redox processes in π-aromatic organic molecules represents a new and promising direction for cost-effective, large-scale energy storage.

## METHODS SUMMARY

The QBFB comprised a mixture of commercially available and custom-made components. Pretreated 2 cm$^2$, stacked (×6) Toray carbon paper electrodes (each of which is about 7.5 μm uncompressed) were used on both sides of the cell. Nafion 212 (50 μm thick) was used as a proton-exchange membrane, and PTFE gasketing was used to seal the cell assembly. On the positive side of the cell, 120 ml of 3 M HBr and 0.5 M $Br_2$ were used as the electrolyte solution in the fully discharged state;

on the negative side, 1 M 2,7-AQDS in 1 M $H_2SO_4$ was used. AQDS disodium salt was flushed twice through a column containing Amberlyst 15H ion-exchange resin to remove the sodium ions. Half-cell measurements were conducted using a Ag/AgCl aqueous reference electrode (3 M KCl filling solution), a Pt wire counter electrode and a 3-mm-diameter glassy carbon disk electrode. For theoretical calculations, the total free energies of molecules were obtained from first-principles quantum chemical calculations within density functional theory at the level of the generalized gradient approximation (GGA) using the PBE functional. Three-dimensional conformer structures for each quinone/hydroquinone molecule were generated using the ChemAxon suite with up to 25 generated conformers per molecule using the Dreiding force field. Generated conformers were used as input structures for the density functional theory geometry optimization employed for determining the formation energy, which in turn is used to evaluate the reduction potential. In the QBFB cost calculation, a price of $4.74 per kilogram (eBioChem) was used for anthraquinone. To get the sulphonated form actually used here, anthraquinone must be reacted with oleum ($H_2SO_4/SO_3$), which adds a negligible cost at scale; this cost is not included here. The price of bromine was $1.76 per kilogram, based on estimates from the US Geological Survey[17]. The cell voltage used to calculate costs here was 0.858 V.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 26 June; accepted 25 November 2013.**


1. Rugolo, J. & Aziz, M. J. Electricity storage for intermittent renewable sources. *Energy Environ. Sci.* **5,** 7151–7160 (2012).
2. Yang, Z. *et al.* Electrochemical energy storage for green grid. *Chem. Rev.* **111,** 3577–3613 (2011).
3. Weber, A. Z. *et al.* Redox flow batteries: a review. *J. Appl. Electrochem.* **41,** 1137–1164 (2011).
4. Leung, P. *et al.* Progress in redox flow batteries, remaining challenges and their applications in energy storage. *RSC Adv.* **2,** 10125–10156 (2012).
5. Nguyen, T. & Savinell, R. F. Flow batteries. *Electrochem. Soc. Interface* **19,** 54–56 (2010).
6. Skyllas-Kazacos, M., Chakrabarti, M. H., Hajimolana, S. A., Mjalli, F. S. & Saleem, M. Progress in flow battery research and development. *J. Electrochem. Soc.* **158,** R55–R79 (2011).
7. Huskinson, B., Rugolo, J., Mondal, S. K. & Aziz, M. J. A high power density, high efficiency hydrogen–chlorine regenerative fuel cell with a low precious metal content catalyst. *Energy Environ. Sci.* **5,** 8690–8698 (2012).
8. Crossley, M. L. The separation of mono-β, 2,6- and 2,7-sulfonic acids of anthraquinone. *J. Am. Chem. Soc.* **37,** 2178–2181 (1915).
9. Conant, J. B., Kahn, H. M., Fieser, L. F. & Kurtz, S. S. An electrochemical study of the reversible reduction of organic compounds. *J. Am. Chem. Soc.* **44,** 1382–1396 (1922).
10. Kelsall, G. H. & Thompson, I. Redox chemistry of $H_2S$ oxidation by the British Gas Stretford process. Part III: Electrochemical behaviour of anthraquinone 2,7 disulphonate in alkaline electrolytes. *J. Appl. Electrochem.* **23,** 296–307 (1993).
11. Forster, R. J. & O'Kelly, J. P. Protonation reactions of anthraquinone-2,7-disulphonic acid in solution and within monolayers. *J. Electroanal. Chem.* **498,** 127–135 (2001).
12. Song, Y. & Buettner, G. R. Thermodynamic and kinetic considerations for the reaction of semiquinone radicals to form superoxide and hydrogen peroxide. *Free Radic. Biol. Med.* **49,** 919–962 (2010).
13. Xu, Y. *et al.* in *World Non-Grid-Connected Wind Power and Energy Conference* 1–4 (IEEE, 2009).
14. Wang, W. *et al.* Anthraquinone with tailored structure for a nonaqueous metal-organic redox flow battery. *Chem. Commun.* **48,** 6669–6671 (2012).
15. Yao, M. *et al.* High capacity organic positive-electrode material based on a benzoquinone derivative for use in rechargeable lithium batteries. *J. Power Sources* **195,** 8336–8340 (2010).
16. Gordon, O. W., Plattner, E. & Doppenberg, F. Production of pulp by the soda-anthraquinone process (SAP) with recovery of the cooking chemicals. US patent 5,595,628 (1997).
17. US Geological Survey *2010 Minerals Yearbook: Bromine* (2010); available at http://minerals.usgs.gov/minerals/pubs/commodity/bromine/myb1-2010-bromi.pdf.
18. US Geological Survey *Vanadium Mineral Commodities Summary* (2013); available at http://minerals.usgs.gov/minerals/pubs/commodity/vanadium/mcs-2013-vanad.pdf.


**Acknowledgements** This work was partially funded through US Department of Energy ARPA-E Award DE-AR0000348 and partially funded through the Harvard School of Engineering and Applied Sciences. Theoretical work was funded in part through the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575. B.H. was supported by an NSF Graduate Research Fellowship. S.E. performed work as part of the Fellowships for Young Energy Scientists programme of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organization for Scientific Research (NWO). We thank T. Betley, L. Hartle, R. Burton and R. Duncan for discussions.

**Author Contributions** B.H. and M.P.M. contributed equally to this work. B.H. and M.P.M. designed and tested the battery, with direction from M.J.A. Both M.P.M. and M.R.G. conducted electrochemistry experiments, with direction from M.J.A. M.P.M. and C.J.G. synthesized chemicals with direction from R.G.G. Theoretical calculations were done by C.S. and S.E., with input from M.P.M. and R.G.G. and direction from A.A.-G. X.C. contributed NMR results. B.H., M.P.M., C.S., M.R.G., S.E., A.A.G., R.G.G. and M.J.A. all contributed to writing the manuscript.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.A. (maziz@harvard.edu).

## METHODS

**Full cell measurements.** The QBFB comprised a mixture of commercially available and custom-made components. Circular endplates were machined out of solid aluminium. Current collectors were 3 inch × 3 inch pyrolytic graphite blocks with interdigitated flow channels (channel width = 0.0625 inch, channel depth = 0.08 inch, landing between channels = 0.031 inch, Fuel Cell Technologies). Pretreated 2 $cm^2$, stacked (×6) Toray carbon paper electrodes (each of which is about 7.5 μm uncompressed) were used on both sides of the cell. Pretreatment consisted of a 10 min sonication in isopropyl alcohol followed by a five hour soak in a hot (50 °C) mixture of undiluted sulphuric and nitric acids in a 3:1 volumetric ratio. Nafion 212 (50 μm thick) was used as a proton-exchange membrane (PEM, Alfa Aesar), and PTFE gasketing was used to seal the cell assembly. Membrane pretreatment was done according to previously published protocols[7]. Six bolts (3/8", 16 threads per inch) torqued to 10.2 N m completed the cell assembly, and PTFE tubing was used to transport reactants and products into and out of the cell. The cell was kept on a hot plate and wrapped in a proportional integral derivative (PID)-controlled heating element for temperature control. On the positive side of the cell, 120 ml of 3 M HBr and 0.5 M $Br_2$ were used as the electrolyte solution in the fully discharged state; on the negative side, 1 M AQDS in 1 M $H_2SO_4$ was used. HBr was used on the negative side instead of $H_2SO_4$ for stability testing results displayed in Extended Data Fig. 9. State-of-charge calculations are based on the composition of the quinone side of the cell. 2,7-Anthraquinone disulphonate disodium salt 98% (TCI) was flushed twice through a column containing Amberlyst 15H ion-exchange resin to remove the sodium ions. Measurements shown here were done at 40 °C. March centrifugal pumps were used to circulate the fluids at a rate of approximately 200 ml $min^{-1}$. For characterization, several instruments were used: a CH Instruments 1100C potentiostat (which can be used up to ±2 A), a DC electronic load (Circuit Specialists) for galvanic discharge, a DC regulated power supply (Circuit Specialists) for electrolytic characterization, and a standard multimeter for independent voltage measurements. The cell was charged at 1.5 V until a fixed amount of charge ran through the cell. During this process, the electrolyte colours changed from orange to dark green (AQDS to $AQDSH_2$) and from colourless to red ($Br^-$ to $Br_2$). Periodically, the open circuit potential was measured, providing the data inset in Fig. 1b. Also, at various SOCs, potential–current behaviour was characterized using the equipment described above: a fixed current was drawn from the cell, and the voltage, once stabilized, was recorded (Fig. 1b). For the cycling data in Fig. 2b, the potentiostat was used for constant-current (±0.5 A $cm^{-2}$) measurements with cut-off voltages of 0 V and 1.5 V. For the cycling data in Fig. 2a, a more dilute quinone solution (0.1 M as opposed to 1 M) was used. Here, the half-cycle lengths were programmed to run at constant current for a fixed amount of time, provided the voltage cut-offs were not reached, so that half of the capacity of the battery was used in each cycle. The voltage cut-offs were never reached during charging, but were reached during discharge. Current efficiencies are evaluated by dividing the discharge time by the charge time of the previous half-cycle.

As shown in Fig. 2, current efficiency starts at about 92% and climbs to about 95% over ~15 standard cycles. Note that these measurements are done near viable operating current densities for a battery of this kind. Because of this, we believe this number places an upper bound on the irreversible losses in the cell. In any case, 95% is comparable to values seen for other battery systems. For example, ref. 19 reports vanadium bromide batteries with current efficiencies of 50–90%, with large changes in current efficiency observed for varying membrane conditions. Our system will probably be less dependent on membrane conditions because we are storing energy in anions and neutral species as opposed to cations, which Nafion can conduct reasonably well.

In Fig. 2b we illustrate the capacity retention of the battery (that is, the number of coulombs available for discharge at the $n$th cycle divided by that available for discharge at the $(n-1)$th cycle) to be 99.2% on average, which is quite high and provides direct evidence that our irreversible losses are below 1%. If we attribute all of this loss (the 0.78% capacity fade per cycle) to some loss of redox-active quinone, it would be equivalent to losing 0.0006634 moles of quinone per cycle. If we attribute all of the loss to bromine crossover (which would react with the hydroquinone and oxidize it back to quinone), this corresponds to a crossover current density of 1.785 mA $cm^{-2}$, which is within the range of the widely varying crossover values reported in the literature[20]. Note that these crossover numbers can be very sensitive to membrane pretreatment conditions. It is also important to mention that, to determine very accurate current efficiencies, detailed chemical analyses of the electrolyte are necessary.

**Half-cell measurements.** These were conducted using a BASi Epsilon EC potentiostat, a BASi Ag/AgCl aqueous reference electrode (RE-5B, 3 M KCl filling solution) and a Pt wire counter electrode. Rotating disk electrode (RDE) measurements were conducted using a BASi RDE (RDE-2) and a 3 mm diameter glassy carbon disk electrode. Electrode potentials were converted to the standard hydrogen electrode (SHE) scale using $E(\mathrm{SHE}) = E(\mathrm{Ag/AgCl}) + 0.210\ \mathrm{V}$, where $E$(SHE) is the potential versus SHE and $E$(Ag/AgCl) is the measured potential versus Ag/AgCl. 2,7-Anthraquinone disulphonate disodium salt 98% was purchased from TCI and used as received. 1,8-Dihydroxy-anthraquinone-2,7-disulphonic acid was made according to the literature procedure[21]. The electrolyte solution was sulphuric acid (ACS, Sigma) in deionized $H_2O$ (18.2 MΩ cm, Millipore). The Pourbaix diagram (plot of $E^0$ versus pH) shown in Extended Data Fig. 4 was constructed using aqueous 1 mM solutions of AQDS in pH buffers using the following chemicals: sulphuric acid (1 M, pH 0), $HSO_4^-/SO_4^{2-}$ (0.1 M, pH 1–2), $AcOH/AcO^-$ (0.1 M, pH 2.65–5), $H_2PO_4^-/HPO_4^{2-}$ (0.1 M, pH 5.3–8), $HPO_4^{2-}/PO_4^{3-}$ (0.1 M, pH 9.28–11.52), and KOH (0.1 M, pH 13). The pH of each solution was adjusted with 1 M $H_2SO_4$ or 0.1 M KOH solutions and measured with an Oakton pH 11 Series pH meter (Eutech Instruments).

**RDE studies.** All RDE data represent an average of three runs. Error bars in Extended Data Figs 2 and 3 indicate standard deviations. Before each run, the glassy carbon disk working electrode was polished to a mirror shine with 0.05 μm alumina and rinsed with deionized water until a cyclic voltammogram of a solution of 1 mM AQDS in 1 M $H_2SO_4$ showed anodic and cathodic peak voltage separation of 34 to 35 mV (39 mV for DHAQDS) at a sweep rate of 25 mV $s^{-1}$. The electrode was then rotated at 200, 300, 400, 500, 700, 900, 1,200, 1,600, 2,000, 2,500 and 3,600 r.p.m. while the voltage was linearly swept from 310 to 60 mV (250 to −100 for DHAQDS) at 10 mV $s^{-1}$ (Extended Data Fig. 1). The currents measured at 60 mV (−100 for DHAQDS) (that is, the diffusion-limited current density) versus the square root of the rotation rate ($\omega$) is plotted in Extended Data Fig. 2. The data were fitted with a straight line, with the slope defined by the Levich equation as $0.620nFAC_OD^{2/3}v^{-1/6}$, where $n = 2$, Faraday's constant $F = 96{,}485$ C $mol^{-1}$, electrode area $A = 0.0707$ $cm^2$, AQDS concentration $C_O = 10^{-6}$ mol $cm^{-3}$, kinematic viscosity $v = 0.01$ $cm^2$ $s^{-1}$. This gives $D$ values of $3.8(1) \times 10^{-6}$ $cm^2$ $s^{-1}$ for AQDS and $3.19(7) \times 10^{-6}$ $cm^2$ $s^{-1}$ for DHAQDS. The reciprocal of the current at overpotentials of 13, 18, 23, 28, 33, 38 and 363 mV was plotted versus the reciprocal of the square root of the rotation rate (Fig. 3b and Extended Data Fig. 2). The data for each potential were fitted with a straight line; the intercept gives the reciprocal of $i_K$, the current in the absence of mass transport limitations (the extrapolation to infinite rotation rate). A plot of $\log_{10}(i_K)$ versus overpotential was linearly fitted with a slope of 62 mV (AQDS) and 68 mV (DHAQDS) defined by the Butler–Volmer equation as $2.3\alpha RT/nF$ (Extended Data Fig. 3), where $R$ is the universal gas constant, $T$ is temperature in kelvin and $\alpha$ is the charge transfer coefficient. This gives $\alpha = 0.474(2)$ for AQDS and 0.43(1) for DHAQDS. The $x$-intercept gives the log of the exchange current $i_0$, which is equal to $FAC_Ok_0$, and gives $k_0 = 7.2(5) \times 10^{-3}$ cm $s^{-1}$ for AQDS and $1.56(5) \times 10^{-2}$ cm $s^{-1}$ for DHAQDS.

**Stability studies.** AQDS (50 mg) was dissolved in 0.4 ml of $D_2O$, and treated with 100 μl of $Br_2$. The $^1H$ and $^{13}C$ NMR spectra (Extended Data Figs 5a, b and 6a, b) were unchanged from the starting material after standing for 20 h at 25 °C. AQDS (50 mg) was then treated with 1 ml of 2 M HBr and 100 μl of $Br_2$. The reaction was heated to 100 °C for 48 h and evaporated to dryness at that temperature. The resulting solid was fully dissolved in $D_2O$ giving unchanged $^1H$ and $^{13}C$ NMR (Extended Data Figs 5c and 6c); however, the $^1H$ NMR reference was shifted due to residual acid. These results imply that bromine crossover will not lead to irreversible destruction of AQDS.

**Sulphonation of anthraquinone and electrochemical study.** 9,10-Anthraquinone was treated with $H_2SO_4$ (20% $SO_3$) at 170 °C for 2 h according to a literature procedure[8]. The resulting red solution, containing roughly 37% AQDS, 60% 9,10-anthraquinone-2,6-disulphonic acid and 3% 9,10-anthraquinone-2-sulphonic acid, was diluted and filtered. A portion of this solution was further diluted with 1 M $H_2SO_4$ to ~1 mM total anthraquinone concentration. The cyclic voltammogram (Extended Data Fig. 8) is similar to that of pure AQDS, though the anodic/cathodic peak current density is broadened due to the presence of the multiple sulphonic acid isomers.

**Theory and methods.** We used a fast and robust theoretical approach to determine the $E^0$ of quinone/hydroquinone couples in aqueous solutions. We employed an empirical linear correlation of $\Delta H_f$, the heat of formation of hydroquinone at 0 K from the quinone and hydrogen gas, to the measured $E^0$ values[22]. Following the treatment of ref. 22, the linear correlation is described as $\Delta G = -nFE^0$, where $\Delta G$ is the difference in total free energy between quinone and hydroquinone, $n$ is the number of electrons involved in the reaction and $F$ is the Faraday constant. The entropy contributions to the total free energies of reaction have been neglected because the entropies of reduction of quinones are found to be very similar[22,23], and the $E^0$ of the oxidation–reduction system is linearly expressed as $(-nF)^{-1}\Delta H_f + b$, where $b$ is a constant. It was also assumed that the reduction of quinones takes place in a single-step reaction involving a two-electron two-proton process[9,24]. The total free energies of molecules were obtained from first-principles quantum chemical calculations within density functional theory (DFT) at the level of the generalized gradient approximation (GGA) using the PBE functional[25]. The projector augmented wave (PAW) technique and a plane-wave basis set[26,27] as implemented

in VASP[28,29] were employed. The kinetic energy cut-off for the plane-wave basis was set at 500 eV, which was sufficient to converge the total energies on a scale of 1 meV per atom. To obtain the ground-state structures of molecules in the gas phase, we considered multiple initial configurations for each molecule and optimized them in a cubic box of 25 Å using Γ-point sampling. The geometries were optimized without any symmetry constraints using the conjugate gradient (CG) algorithm, and the convergence was assumed to be complete when the total remaining forces on the atoms were less than 0.01 eV $Å^{-1}$.

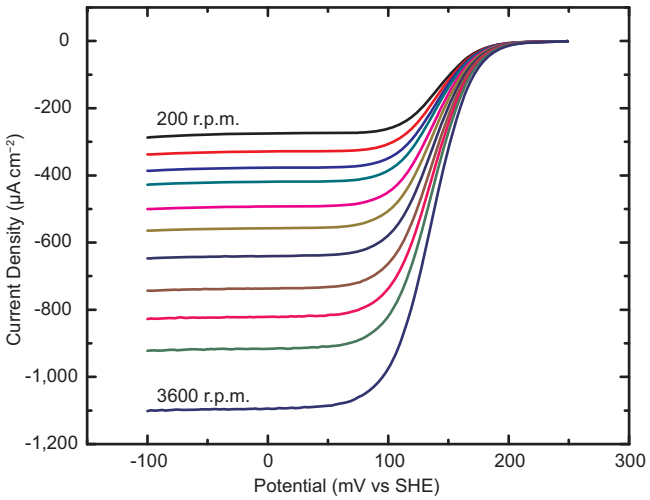
The search for conformational preference through theoretical calculations for each hydroxylated quinone is crucial because of the significant effects of intramolecular hydrogen bonds on the total free energies of the molecules[30]. Three-dimensional conformer structures for each quinone/hydroquinone molecule were generated using the ChemAxon suite (Marvin 6.1.0 by ChemAxon, http://www.chemaxon.com) with up to 25 conformers generated per molecule using the Dreiding force field[31]. The conformers generated were used as input structures for the DFT geometry optimization employed for determining $\Delta H_f$, which in turn is used to estimate $E^0$ and $G^0_{solv}$.

To calculate the $E^0$ of the hydroxy-substituted AQDS molecules (Fig. 3c), the correlation between $\Delta H_f$ and $E^0$ was calibrated from experimental data on six well-characterized quinones[32]. Specifically, we used the experimental values of the aqueous $E^0$ and computed $\Delta H_f$ of 1,2-benzoquinone, 1,4-benzoquione, 1,2-naphthoquinone, 1,4-naphthoquinone, 9,10-anthraquinone and 9,10-phenanthrene[33]. The training set ensures that the calibration plot addresses most classes and aspects of quinones, including two quinones each from 1-ring (benzoquinone), 2-ring (naphthoquinone) and 3-ring (anthraquinone and phenanthrene) structures. In addition, the experimental values of $E^0$ of the training set spanned 0.09 V (9,10-anthraquinone) to 0.83 V (1,2-benzoquinone), providing a wide range for $E^0$ (Extended Data Fig. 7). The linear calibration plot for $E^0$ yields an $R^2 = 0.97$ between the calculated $\Delta H_f$ and $E^0$ (Extended Data Fig. 7).

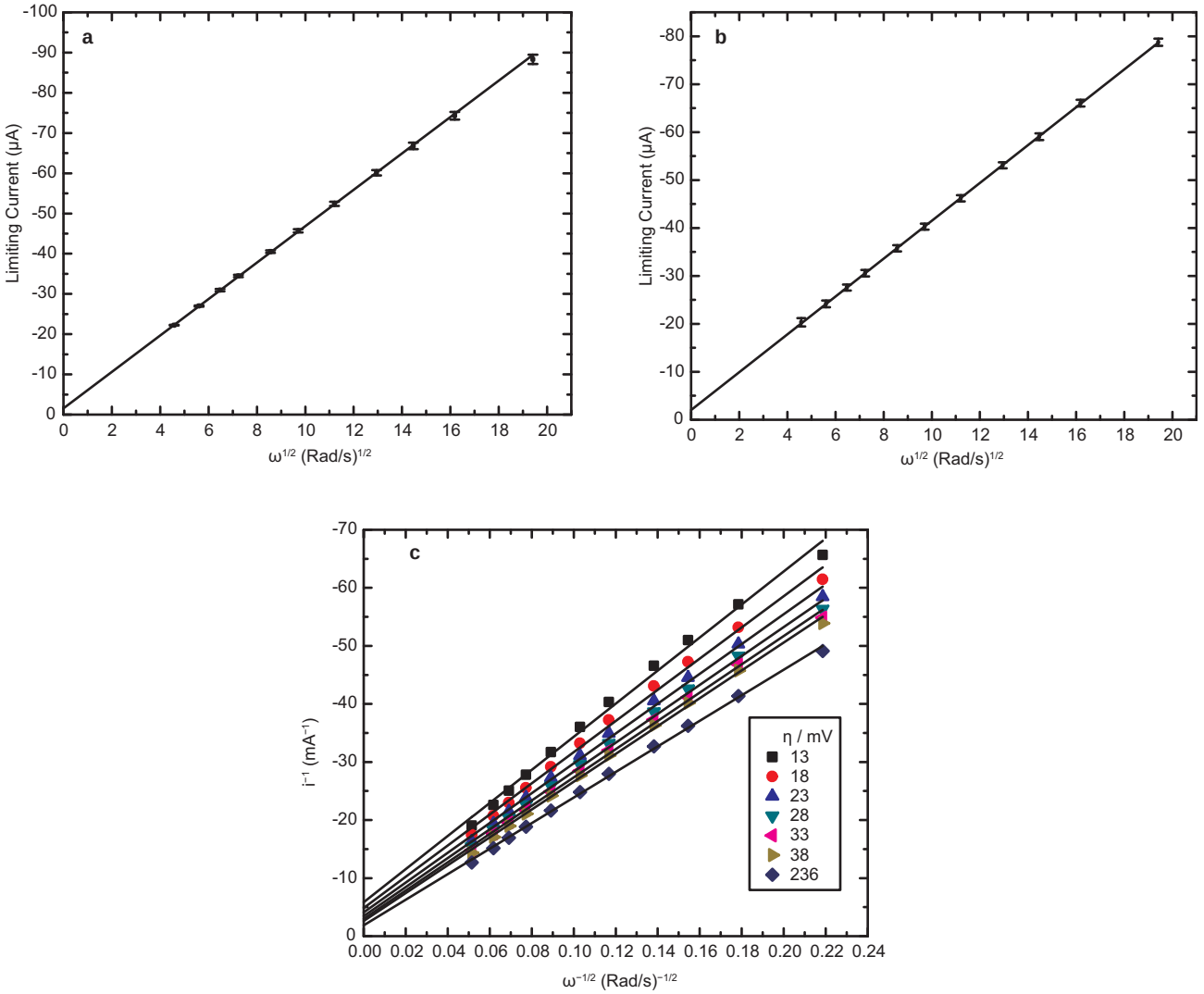
The $G^0_{solv}$ values of the quinones in water were calculated using the Jaguar 8.0 program in the Schrödinger suite 2012 (Jaguar, version 8.0, Schrödinger). The standard Poisson–Boltzmann solver was employed[34,35]. In this model, a layer of charges on the molecular surface represents the solvent. $G^0_{solv}$ was calculated as the difference between the total energy of the solvated structure and the total energy of the molecule in vacuum. A more negative value for $G^0_{solv}$ corresponds to a quinone that is likely to have a higher aqueous solubility. An absolute prediction of the solubility is not readily available, as the accurate prediction of the most stable forms of molecular crystal structures with DFT remains an open problem[36].

**Cost calculations.** These were done using the following formula: $C = (3.6 \times 10^3) \times (PM)/(nFE)$, where $C$ is the cost in US dollars of the compound per kilowatt hour, $P$ is the cost in US dollars per kilogram, $M$ is the molecular mass of the compound, $F$ is Faraday's constant (96,485 C $mol^{-1}$), $n$ is the number of moles of electrons transferred per mole of storage compound (two for the QBFB) and $E$ is the open-circuit voltage (V) of the storage device. In calculating the price for the anthraquinone-bromine battery, a price of $4.74 per kilogram (eBioChem) was used for anthraquinone (note that, to get the sulphonated form actually used here, anthraquinone must be reacted with oleum ($H_2SO_4/SO_3$), which adds a negligible cost at scale; this cost is not included here). The price of bromine was $1.76 per kilogram, based on estimates from the US Geological Survey[17]. The cell voltage used to calculate costs here was 0.858 V. For vanadium, costs were calculated from USGS prices from 2011[18] of vanadium pentoxide at $14.37 per kilogram, and the cell voltage used was 1.2 V. Balance-of-system costs have not been estimated because the technology is too immature.

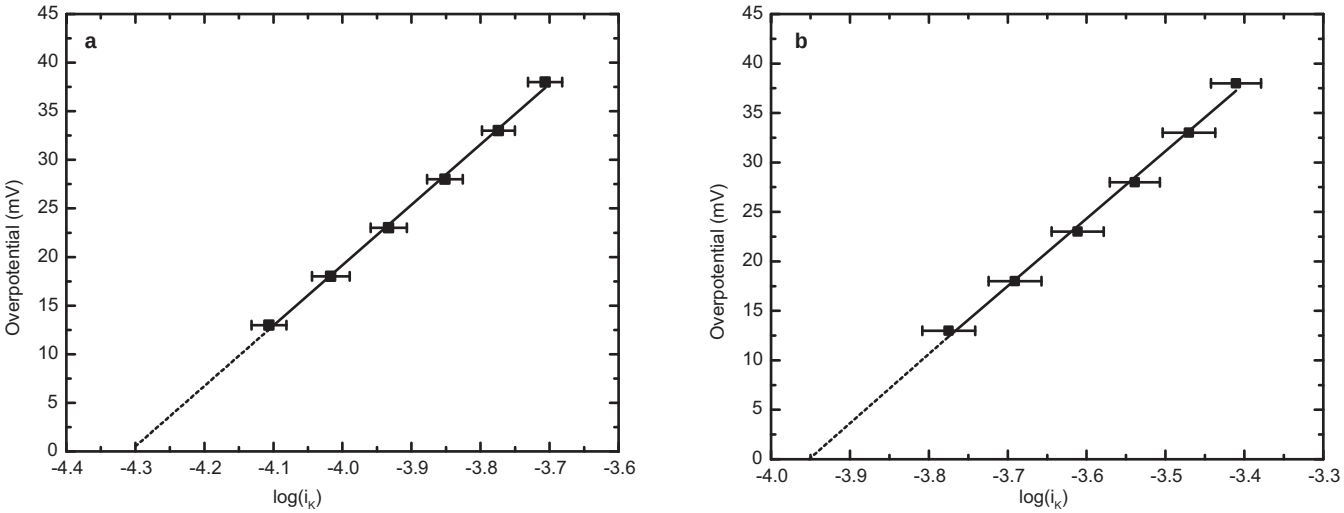

19. Skyllas-Kazacos, M., Milne, N. A. & Kazacos, G. C. Membrane properties and behavior in the Generation 2 vanadium bromide redox flow batteries. *Materials Forum* **32,** 72–77 (2008).
20. Yeo, R. S. & McBreen, J. Transport properties of Nafion membranes in electrochemically regenerative hydrogen/halogen cells. *J. Electrochem. Soc.* **126,** 1682–1687 (1979).
21. Haase, J., Baudys, J., Obruba, K. & Panek, J. Dihydroxyanthraquinone disulfonic acids. Czechoslovakian patent 148,977 (1973).
22. Dewar, M. J. S. & Trinajstic, N. Ground states of conjugated molecules-XIV: redox potentials of quinones. *Tetrahedron* **25,** 4529–4534 (1969).
23. Pullman, B. & Pullman, A. *Quantum Biochemistry* 475 (Interscience, 1963).
24. Guin, P. S., Das, S. & Mandal, P. C. Electrochemical reduction of quinones in different media: a review. *Int. J. Electrochem.* 816202 (2011).
25. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77,** 3865–3868 (1996).
26. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50,** 17953–17979 (1994).
27. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59,** 1758–1775 (1999).
28. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47,** 558–561 (1993).
29. Kresse, G. & Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54,** 11169–11186 (1996).
30. Qu, R., Liu, H., Feng, M., Yang, X. & Wang, Z. Investigation on intramolecular hydrogen bond and some thermodynamic properties of polyhydroxylated anthraquinones. *J. Chem. Eng. Data* **57,** 2442–2455 (2012).
31. Mayo, S. L., Olafson, B. D. & Goddard, W. A. III Dreiding: a generic force field for molecular simulations. *J. Phys. Chem.* **94,** 8897–8909 (1990).
32. Wang, J. & Hou, T. Recent advances on aqueous solubility prediction. *Comb. Chem. High Throughput Screen.* **14,** 328–338 (2011).
33. Johnsson Wass, J. R. T., Ahlberg, E., Panas, I. & Schiffrin, D. J. Quantum chemical modeling of the reduction of quinones. *J. Phys. Chem. A* **110,** 2005–2020 (2006).
34. Tannor, D. J. *et al.* Accurate first principles calculation of molecular charge distributions and solvation energies from ab initio quantum mechanics and continuum dielectric theory. *J. Am. Chem. Soc.* **116,** 11875–11882 (1994).
35. Marten, B. *et al.* New model for calculation of solvation free energies: correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects. *J. Phys. Chem.* **100,** 11775–11788 (1996).
36. Hongo, K., Watson, M. A., Sánchez-Carrera, R. S., Iitaka, T. & Aspuru-Guzik, A. Failure of conventional density functionals for the prediction of molecular crystal polymorphism: a quantum Monte Carlo study. *J. Phys. Chem. Lett.* **1,** 1789–1794 (2010).



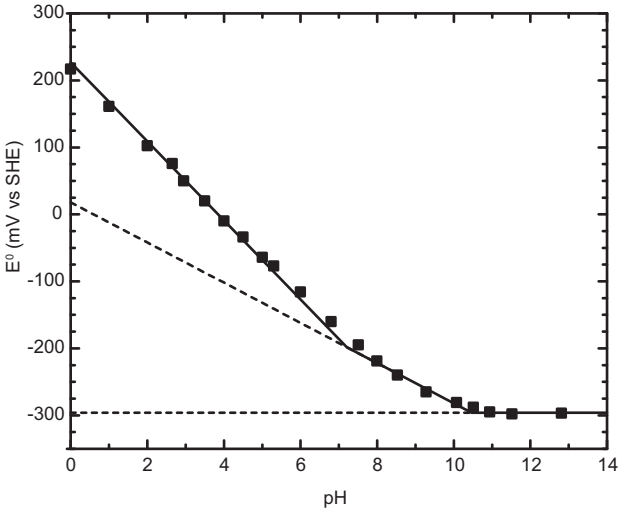

**Extended Data Figure 1 | Plot of potential versus current density at different rotation rates of the RDE.** The solution is 1 mM DHAQDS (1 mM in 1 M $H_2SO_4$), using a rotating disk electrode (RDE) of glassy carbon. Rotation rates are 200, 300, 400, 500, 700, 900, 1,200, 1,600, 2,000, 2,500 and 3,600 r.p.m.



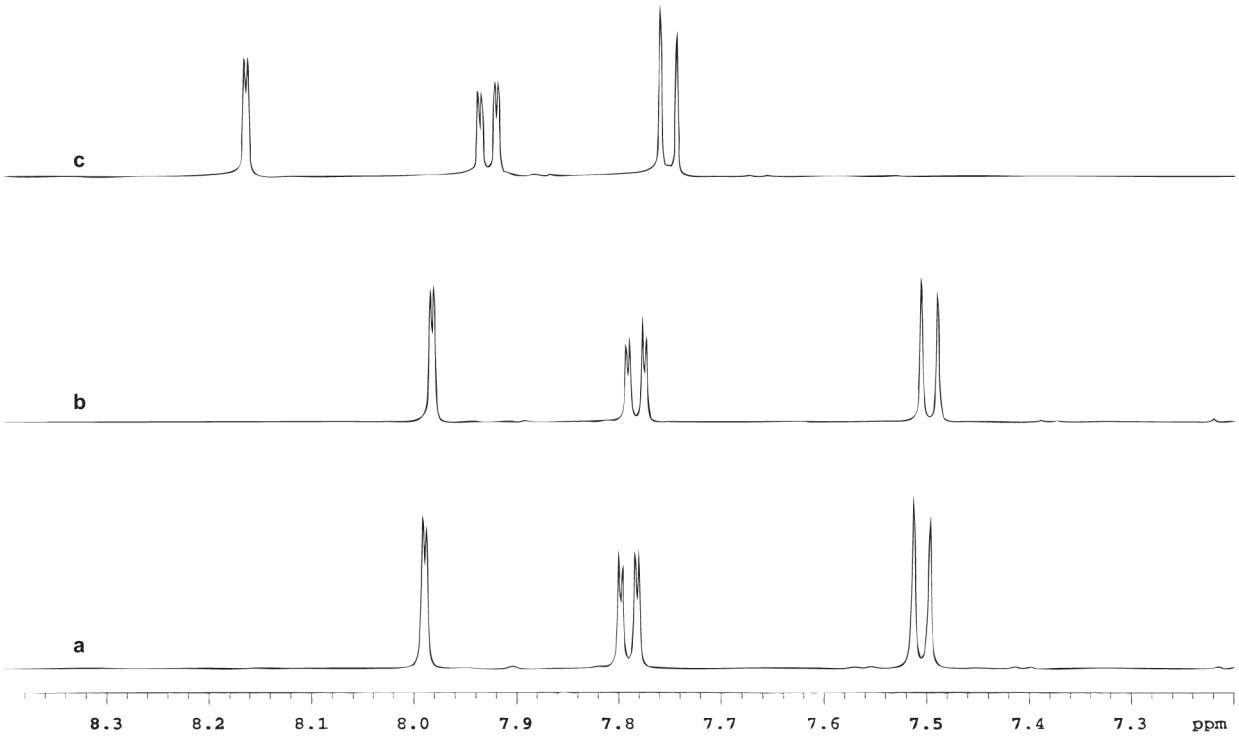

**Extended Data Figure 2 | Levich and Koutecký–Levich plots obtained using the RDE.** **a**, Levich plot (limiting current versus square root of rotation rate $\omega$) of 1 mM AQDS in 1 M $H_2SO_4$ (the fitted line has a slope of 4.53(2) μA $s^{1/2}$ $rad^{-1/2}$, giving $D = 3.8(1) \times 10^{-6}$ $cm^2$ $s^{-1}$). Data are an average of three runs; error bars indicate the standard deviation. **b**, As **a** but for DHAQDS in 1 M $H_2SO_4$ (slope of 3.94(6) μA $s^{1/2}$ $rad^{-1/2}$ gives $D = 3.19(7) \times 10^{-6}$ $cm^2$ $s^{-1}$). **c**, Koutecky–Levich plot ($i^{-1}$ versus $\omega^{-1/2}$) of 1 mM DHAQDS in 1 M $H_2SO_4$. The current response, $i$, is shown for seven different AQDS reduction overpotentials $\eta$.



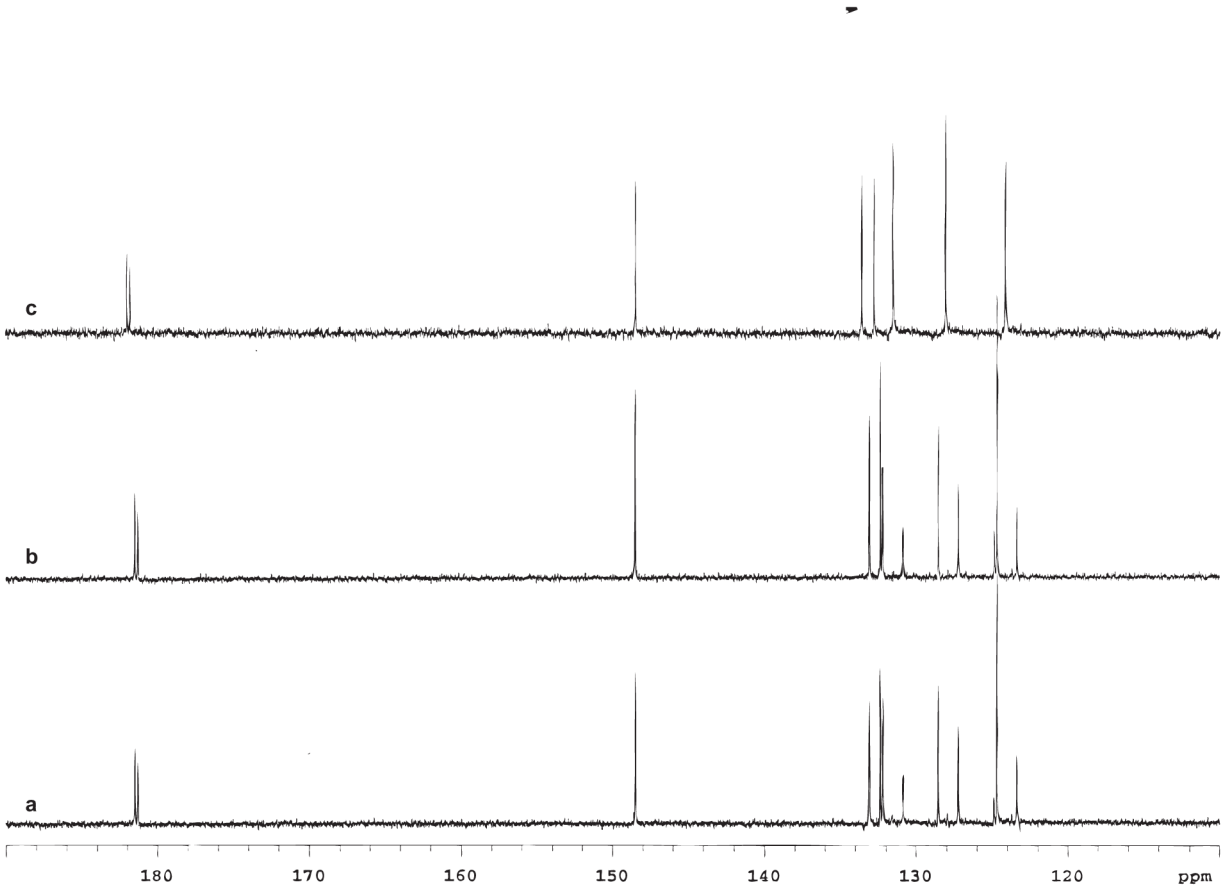

**Extended Data Figure 3 | Fit of Butler–Volmer equation.** Constructed using the current response in the absence of mass transport at low AQDS reduction overpotentials; $i_K$ is the current extrapolated from the zero-intercept of Fig. 3b and Extended Data Fig. 2c (infinite rotation rate). Data are an average of three runs; error bars indicate the standard deviation. **a**, AQDS: best-fit line has the equation $y = 62(x + 4.32)$. This yields $\alpha = 0.474(2)$ and $k_0 = 7.2(5) \times 10^{-3}\,\mathrm{cm\,s^{-1}}$. **b**, DHAQDS: best-fit line is the function $y = 68(x + 3.95)$. This yields $\alpha = 0.43(1)$ and $k_0 = 1.56(5) \times 10^{-2}\,\mathrm{cm\,s^{-1}}$.



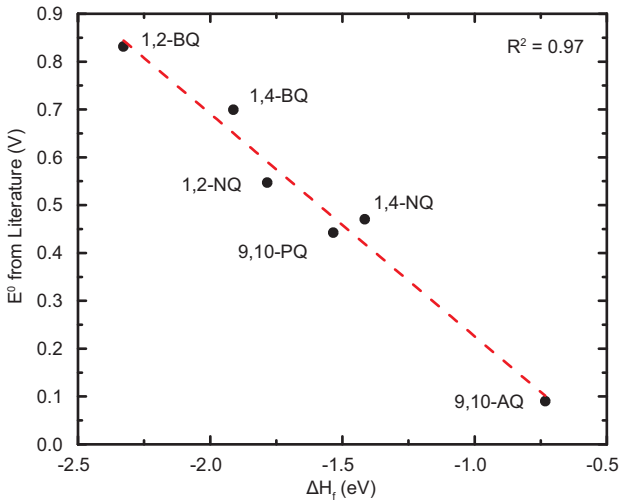

**Extended Data Figure 4 | Pourbaix diagram ($E^0$ vs pH) of AQDS.** Data are fitted to three solid lines indicating slopes of $-59\,\mathrm{mV\,pH^{-1}}$, $-30\,\mathrm{mV\,pH^{-1}}$ and $0\,\mathrm{mV\,pH^{-1}}$, corresponding to two-, one- and zero-proton processes, respectively. Dashed lines linearly extrapolate the one- and zero-proton processes to give $E^0$ values of 18 mV ($2e^-/1H^+$) and $-296$ mV ($2e^-/0H^+$).



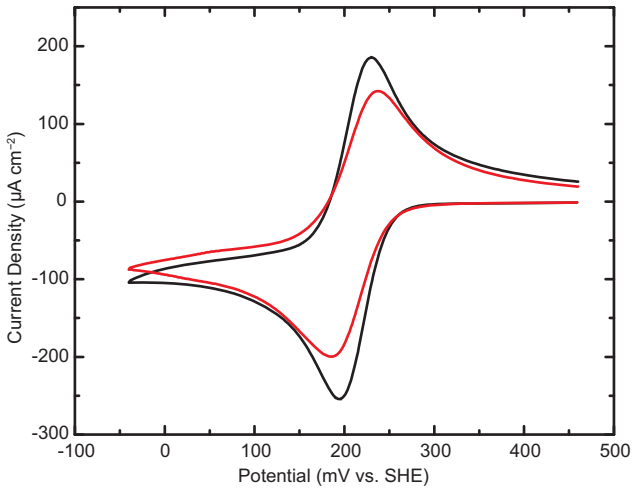

**Extended Data Figure 5 | $^1$H NMR (500 MHz, $D_2O$) spectra.** **a**, Spectrum of AQDS: chemical shift $\delta = 7.99$ p.p.m. versus tetramethylsilane (TMS) (doublet (d), coupling constant $J = 2$ Hz, 1,8 C–H), 7.79 p.p.m. (doublet of doublets, $J = 2$ and 8 Hz, 4,5 C–H), 7.50 p.p.m. (d, $J = 8$ Hz, 3,6 C–H). **b**, The same sample, 20 h after addition of $Br_2$. **c**, $^1$H NMR of AQDS treated with 2 M HBr and $Br_2$ and heated to 100 °C for 48 h. The peaks are shifted due to presence of trace HBr, which shifted the residual solvent peak due to increased acidity. Coupling constants for each peak are identical to **a**.



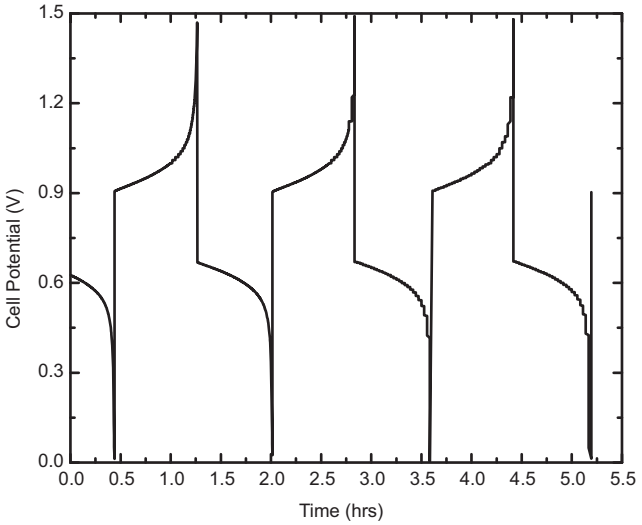

**Extended Data Figure 6 | $^{13}$C NMR (500 MHz, $D_2O$) spectra.** **a**, AQDS, $\delta$ = 181.50 p.p.m. versus TMS (C 9), 181.30 p.p.m. (C 10), 148.51 p.p.m. (C 2,7), 133.16 p.p.m. (C 11), 132.40 p.p.m. (C 12), 130.86 p.p.m. (C 3,6), 128.59 p.p.m. (C 4,5), 124.72 p.p.m. (C 1,8). **b**, The same sample, 24 h after addition of $Br_2$. **c**, $^{13}$C NMR of AQDS treated with 2 M HBr and $Br_2$ and heated to 100 °C for 48 h.




**Extended Data Figure 7 | Calibration model for $\Delta H_f$ and experimental $E^0$.** This shows a linear relationship (red dashed line; $R^2 = 0.97$) between calculated $\Delta H_f$ (this work) and experimental $E^0$ (from the literature) of six quinones in aqueous solutions: BQ, benzoquinone; NQ, naphthoquinone; AQ, anthraquinone; and PQ, phenanthraquinone.

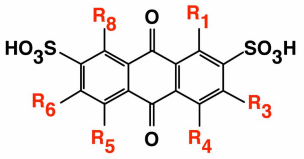



**Extended Data Figure 8 | AQDS cyclic voltammograms.** Black curve, obtained for a 1 mM solution of AQDS in 1 M $H_2SO_4$ on a stationary glassy carbon working electrode. Red curve, obtained for a crude anthraquinone sulphonation solution containing a mixture of AQDS, 9,10-anthraquinone-2,6-disulphonic acid and 9,10-anthraquinone-2-sulphonic acid diluted to 1 mM total anthraquinone in 1 M $H_2SO_4$.




**Extended Data Figure 9 | Flow-battery cycling behaviour with HBr electrolyte on both sides.** Data collected by cycling the current at 0.2 A $cm^{-2}$ at 40 °C using a 2 M HBr + 0.5 M $Br_2$ solution on the positive side and a 2 M HBr + 0.1 M AQDS solution on the negative side; cell potential versus time performance is comparable to data in Fig. 2.

**Extended Data Table 1 | AQDS screened by theoretical calculations**




| ID | -OH substituted | $R_1$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | $R_8$ | $E^0$ (V) | $G^0_{solv}$ (kJ mol$^{-1}$) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Non- | H | H | H | H | H | H | 0.222 | -81.5 |
| 2 | Mono- | OH | H | H | H | H | H | 0.185 | -81.5 |
| 3 | | H | OH | H | H | H | H | 0.325 | -111.7 |
| 4 | | H | H | OH | H | H | H | 0.108 | -88.2 |
| 5 | Di- | OH | OH | H | H | H | H | 0.176 | -110.3 |
| 6 | | OH | H | OH | H | H | H | 0.027 | -85.6 |
| 7 | | OH | H | H | OH | H | H | 0.122 | -96.7 |
| 8 | | OH | H | H | H | OH | H | 0.143 | -85.7 |
| 9 | | OH | H | H | H | H | OH | 0.101 | -83.2 |
| 10 | | H | OH | OH | H | H | H | 0.153 | -105.4 |
| 11 | | H | OH | H | OH | H | H | 0.179 | -119.1 |
| 12 | | H | OH | H | H | OH | H | 0.202 | -112.0 |
| 13 | | H | H | OH | OH | H | H | 0.000 | -95.6 |
| 14 | Tri- | OH | OH | OH | H | H | H | -0.070 | -101.7 |
| 15 | | OH | OH | H | OH | H | H | 0.083 | -116.2 |
| 16 | | OH | OH | H | H | OH | H | 0.187 | -114.3 |
| 17 | | OH | OH | H | H | H | OH | 0.310 | -120.9 |
| 18 | | OH | H | OH | OH | H | H | -0.102 | -91.4 |
| 19 | | OH | H | OH | H | OH | H | 0.089 | -114.0 |
| 20 | | OH | H | OH | H | H | OH | -0.085 | -87.1 |
| 21 | | OH | H | H | OH | OH | H | -0.048 | -102.8 |
| 22 | | H | OH | OH | OH | H | H | -0.107 | -107.8 |
| 23 | | H | OH | OH | H | OH | H | 0.106 | -136.8 |
| 24 | Tetra- | OH | OH | OH | OH | H | H | -0.098 | -109.0 |
| 25 | | OH | OH | OH | H | OH | H | 0.012 | -108.4 |
| 26 | | OH | OH | OH | H | H | OH | -0.222 | -102.3 |
| 27 | | OH | OH | H | OH | OH | H | -0.019 | -132.3 |
| 28 | | OH | OH | H | OH | H | OH | 0.046 | -114.6 |
| 29 | | OH | OH | H | H | OH | OH | 0.080 | -111.1 |
| 30 | | OH | H | OH | OH | OH | H | -0.259 | -99.0 |
| 31 | | OH | H | OH | OH | H | OH | -0.199 | -91.9 |
| 32 | | H | OH | OH | OH | OH | H | -0.083 | -120.6 |
| 33 | Penta- | OH | OH | OH | OH | OH | H | -0.252 | -117.1 |
| 34 | | OH | OH | OH | OH | H | OH | -0.292 | -108.3 |
| 35 | | OH | OH | OH | H | OH | OH | -0.030 | -111.6 |
| 36 | Hexa- | OH | OH | OH | OH | OH | OH | -0.077 | -121.0 |

The effect of –OH substitution on reduction potential and solvation energy.

LETTER

doi:10.1038/nature12761

# Merging allylic carbon–hydrogen and selective carbon–carbon bond activation

Ahmad Masarwa[1], Dorian Didier[1], Tamar Zabrodski[1], Marvin Schinkel[2], Lutz Ackermann[2] & Ilan Marek[1]

**Since the nineteenth century, many synthetic organic chemists have focused on developing new strategies to regio-, diastereo- and enantioselectively build carbon–carbon and carbon–heteroatom bonds in a predictable and efficient manner[1–3]. Ideal syntheses should use the least number of synthetic steps, with few or no functional group transformations and by-products, and maximum atom efficiency. One potentially attractive method for the synthesis of molecular skeletons that are difficult to prepare would be through the selective activation of C–H and C–C bonds[4–8], instead of the conventional construction of new C–C bonds. Here we present an approach that exploits the multifold reactivity of easily accessible substrates[9] with a single organometallic species to furnish complex molecular scaffolds through the merging of otherwise difficult transformations: allylic C–H and selective C–C bond activations[10–12]. The resulting bifunctional nucleophilic species, all of which have an all-carbon quaternary stereogenic centre, can then be selectively derivatized by the addition of two different electrophiles to obtain more complex molecular architecture from these easily available starting materials.**

The direct transformation of C–H bonds into C–C and C–heteroatom bonds by means of C–H activation[4] has had an enormous impact on advancing the field of chemical synthesis, because it led to the design of novel, selective, economic and efficient processes for the utilization of hydrocarbons[5,6]. Carbon–carbon bond activation by soluble metal complexes, an alternative strategy, is based on either oxidative addition or β-carbon and radical cleavage[7]. The former approach, requiring the introduction of pincer-type ligands, has expanded the field by bringing the metal centre close to the otherwise 'hidden' C–C bond through intramolecular chelation[8]. However, this system puts some limitations on synthesis and, except for hydrogenation and the silylation of unactivated C–C bonds[7], no other applications were described. The latter approach relies on strained substrates that do not require such pre-coordination, but the selectivity of the C–C bond activation remains a major problem when C–C bonds are differentially substituted[9].

Selective C–H bond activation and selective C–C bond activation are cutting-edge methods that, in combination, may be ideal for accessing complex molecular architectures. Indeed, the ability to design the preparation of complex molecular skeletons through a succession of allylic C–H and selective C–C bond activations would pave the way for new directions in organic synthesis[10–12]. Unifying these two independent strategies into a single approach would necessitate perfect control of both modes of activation. Here we describe the development and study of reactions in which the challenges of selective C–H bond activation and selective C–C bond activation are overcome, demonstrating the efficacy of a synthetic approach that could be converted to the preparation of all-carbon quaternary stereogenic centres in acyclic systems. We predicted that a direct and expedient approach to such complex molecular frameworks containing several stereogenic centres in acyclic systems could result from either ω-ene-cyclopropanes **1** or alkylidenecyclopropanes **2**. Ideally, it could proceed through a series of allylic C–H activations ('metal-walk' reactions) followed by a selective C–C bond activation of the three-membered ring to give intermediate **3**. Then selective reactions with two different electrophiles would functionalize **3** into the linear adduct **4**, which possesses new stereogenic centres including the original all-carbon quaternary stereocentre (path A; Fig. 1a). Alternatively, ω-ene-cyclopropanes possessing a leaving group **5**, treated under the same conditions, would undergo the allylic C–H activation followed by a selective C–C fragmentation to lead to non-conjugated dienes **6** (path B; Fig. 1b).

This approach would require the use of a single organometallic species that could mediate a combination of allylic C–H activation, selective C–C bond activation and selective functionalization with two different

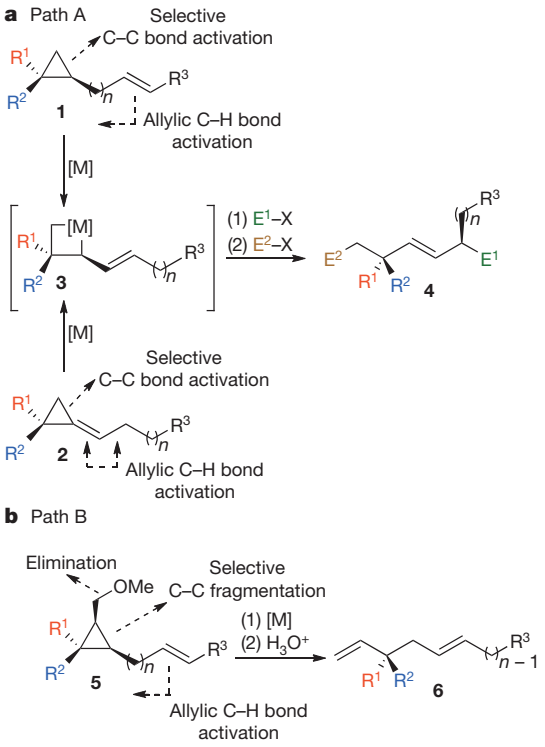



**Figure 1 | General approach combining allylic C–H and selective C–C bond activation towards functionalized adducts. a**, On path A, ω-ene-cyclopropanes **1** undergo a metal-mediated allylic C–H activation followed by a selective C–C bond activation to lead to bismetalated species **3**. Because the reactivity of an allylmetal species to electrophiles is higher than that of the alkylmetal species, the first electrophile ($E^1$–X) reacts selectively to functionalize the γ-carbon whereas the second electrophile ($E^2$–X) reacts with the remaining $sp^3$ organometallic species to give the functionalized acyclic adduct **4**. When alkylidenecyclopropane species **2** are used, the chemistry proceeds similarly and **4** is formed. **b**, Alternatively (on path B), when ω-ene-cyclopropane **5** is used, the sequence goes through the combined allylic C–H activation/selective C–C bond cleavage promoted by an elimination reaction to give the non-conjugated dienes **6** possessing the all-carbon quaternary stereocentre. Me, methyl.

[1]The Mallat Family Laboratory of Organic Chemistry, Schulich Faculty of Chemistry and the Lise Meitner-Minerva Center for Computational Quantum Chemistry, Technion-Israel Institute of Technology, Haifa 32000, Israel. [2]Institut Für Organische und Biomolekulare Chemie, Georg-August-Universität, Göttingen, Tammannstrasse 2, 37077 Göttingen, Germany.

electrophiles. Our solution to this salient problem is based on the Zr-walk chemistry we developed for the transformation of simple ω-unsaturated fatty alcohols[13] and ω-non-conjugated enol ether derivatives[14] into allyl- and dienylzirconocene derivatives, respectively[15]. Recently, a related enantioselective Heck-type arylation of alkenyl alcohols using a Pd walk as a redox-relay strategy was reported[16,17].

Our mechanistic rationale was to transform the model ω-alkenyl cyclopropane species **1a** by means of a zirconocene-mediated migration of the double bond with the Negishi reagent[18]. Indeed, $C_4H_8ZrCp_2$ would react with the remote double bond of **1a** to give the corresponding zirconacyclopropane, **7**. Then, through an allylic C–H bond activation, the $\eta^3$-allyl intermediate **8** would be generated as a transient species and, after hydride insertion, would give rise to the new zirconacyclopropane, **9**, possessing a C–Zr bond in the β-position of the three-member ring[19,20] (Fig. 2). When **9** is formed, an irreversible C–C bond activation proceeding through ring cleavage should take place. If the C–C bond activation occurs through the $C_2$–$C_3$ bond, a primary acyclic organometallic species, **3a**, could be obtained. Alternatively, ring opening through the activation of the $C_1$–$C_2$ bond (not shown in Fig. 2), would result in an unfavourable trisubstituted organometallic species. On the assumption that primary organometallic species are kinetically and thermodynamically more stable than tertiary organometallic species[9], we anticipate an exclusive formation of the constitutional isomer **3a** in the activation reaction. Hence, the presence of substituents on the cyclopropyl unit furnishes a highly selective C–C bond activation[21].

Importantly, even if our hypotheses regarding selective bond activation were correct, the synthetic problem would not yet be fully solved, because it is necessary to differentiate the reactivity of the two C–metal bonds in **3a** (or **3b**) with two different electrophiles. It occurred to us that in **3a** (or **3b**), the allylzirconocene moiety [$C_2$–Zr] might be significantly more reactive than the alkylzirconocene moiety [$C_3$–Zr] (ref. 22). Therefore, the addition of the first electrophile should selectively react to give **10** via a $S_E2'$ reaction, and **10** will then go on to react with a second electrophile to form **4** (Fig. 2). On addition of the Negishi reagent to **1a** ($n = 1$) and subsequent hydrolysis, we isolated **4a** as the *E*-isomer in excellent yield after hydrolysis (Fig. 2). When $D_2O$ was added at the end of the sequence, the bisdeuterated compound **4a(D)** was similarly afforded with an exclusive introduction of deuterium both at the allylic and primary alkyl positions. This tandem allylic C–H activation/selective C–C activation was not limited to ω-ene-cyclopropanes with a one-carbon tether. Compounds **1b**–**1d** (with respective carbon tethers $n = 2$, 3 and 6) also underwent this transformation via an initial migration of the zirconocene through the alkyl chain with formation of **4b**–**4d** in good yields[23]. This reaction sequence always led to a selective C–C activation of the 1,1-disubstituted ω-ene-cyclopropanes **4a**–**4e** ($R^1$, $R^2$ = alkyl or aryl) because it led to the preferential formation of a primary alkylzirconacyclobutane over a tertiary alkyl-zirconacyclobutane. Because

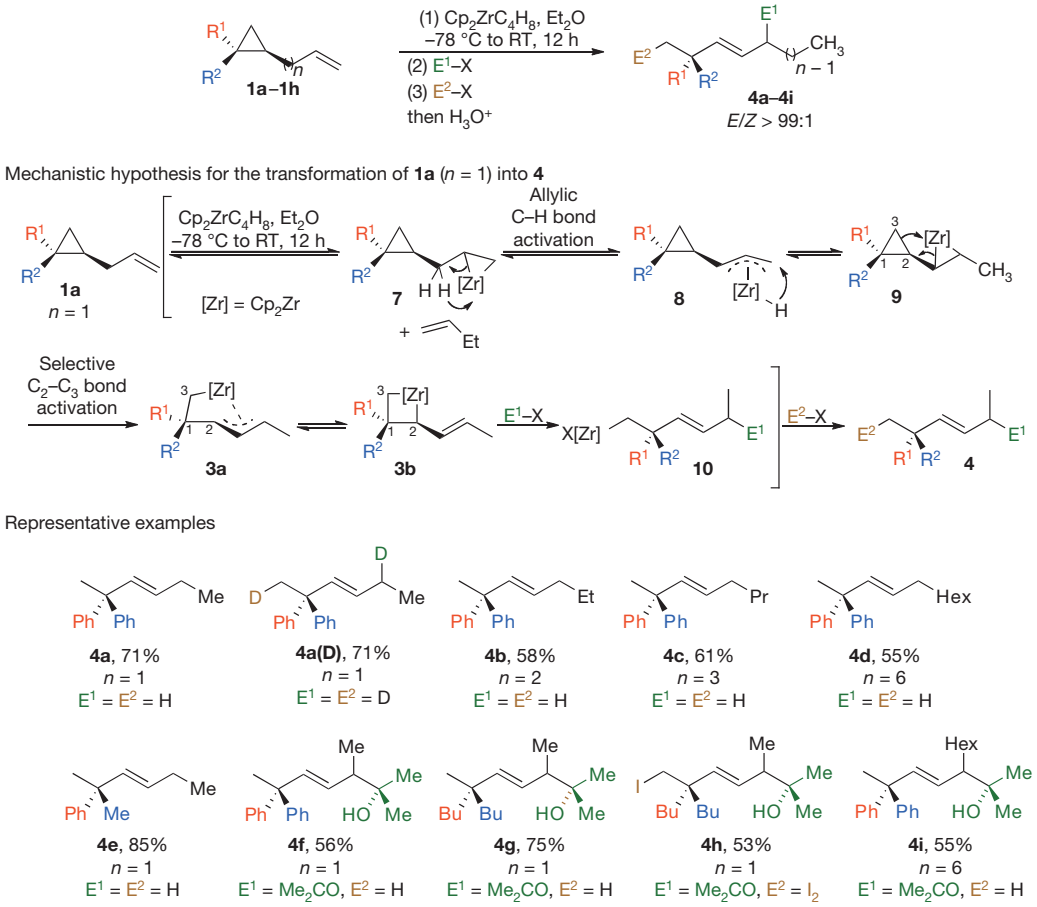



**Figure 2 | Zirconocene-promoted allylic C–H and C–C bond activation of ω-ene-cyclopropanes.** $C_4H_8ZrCp_2$ reacts with the remote double bond of **1a** to give **7** and free butene. Then, through an allylic C–H bond activation, the $\eta^3$-allyl intermediate **8** is generated and, after hydride insertion, zirconacyclopropane **9** is formed. Then an irreversible step occurs transforming the zirconacyclopropane **9** into allyl-alkyl zirconocene species **3a** or **3b**. From the two possible C–C bonds that could be activated ($C_1$–$C_2$ versus $C_2$–$C_3$ in **9**), only the $C_2$–$C_3$ bond is cleaved, leading to the primary and allyl-bismetalated organozirconocene derivative **3a**, which may be represented as its σ-allylic zirconocene **3b**. Because the $C_2$–Zr fragment is more reactive to electrophiles than is the $C_3$–Zr fragment, the addition of the first electrophile reacts selectively at room temperature (RT = 25 °C) with the allylzirconocene moiety. Also shown are representative examples of the scope of the combined allylic C–H activation, selective activation and bisfunctionalization of the resulting organozirconocene species en route to acyclic fragments. Bu, butyl; Et, ethyl; Hex, hexane.

the allylic [$C_2$–Zr] bond is more reactive to electrophiles than is the primary alkyl-[$C_3$–Zr] bond, the addition of a carbonyl electrophile ($E^1$–X = MeCOMe) selectively functionalizes the allylmetal fragment, whereas the second electrophile ($E^2$–X = $H_3O^+$, $I_2$) reacts with the remaining alkylzirconocene species to give **4f**–**4i** as *E*-isomers in good yields.

This multifold functionalization of a single molecule was also successfully applied to alkylidenecyclopropane derivatives (ACPs) **2** to furnish linear adducts **4**. Alkylidenecyclopropanes are highly reactive substrates, often giving rise to complex mixtures of products, and development of mild and selective transformations for this class of compounds is very useful[9,24]. In this context, the Zr walk could lead to particularly interesting transformations of ACPs (Fig. 3). To further our mechanistic understanding, we used $CD_3$-labelled ACP **2a** as a substrate. The migration of a deuterium atom into the vinylic position in the product **4j** corroborated our hypothesis of allylic C–H activation (Fig. 3), as follows.

When ACP **2a** is treated with the Negishi reagent, the zirconocene-mediated allylic C–D bond activation releases the initial strain of the alkylidenecyclopropane system to furnish zirconacyclopropane $\mathbf{9_{d3}}$. It then undergoes a cyclopropane C–C bond activation to give zirconacyclobutane $\mathbf{3_{d3}}$, which is functionalized selectively at the allylic position to give **4j** in 82% yield by addition of acetaldehyde. In all cases, the combined allylic C–H bond activation followed by the selective $C_2$–$C_3$ bond activation of alkylidenecyclopropane proceeded smoothly and only the linear products **4k**–**4w** were formed with no traces of activation along the $C_1$–$C_2$ bond, even though the activation could have otherwise occurred at the benzylic position (that is, **4l** and **4n**).

Because enantiomerically enriched alkylidenecyclopropanes can easily be prepared[25,26], and because the chiral quaternary centre is at no risk of racemization in the process, the optical purity can be assumed to be identical throughout the rest of the synthesis. Indeed, when enantiomerically enriched **2b** ($R^1$ = Bu, $R^2$ = Et, $R^3$ = H; enantiomeric ratio, 98:2) was treated with the Negishi reagent, **4k** was obtained with

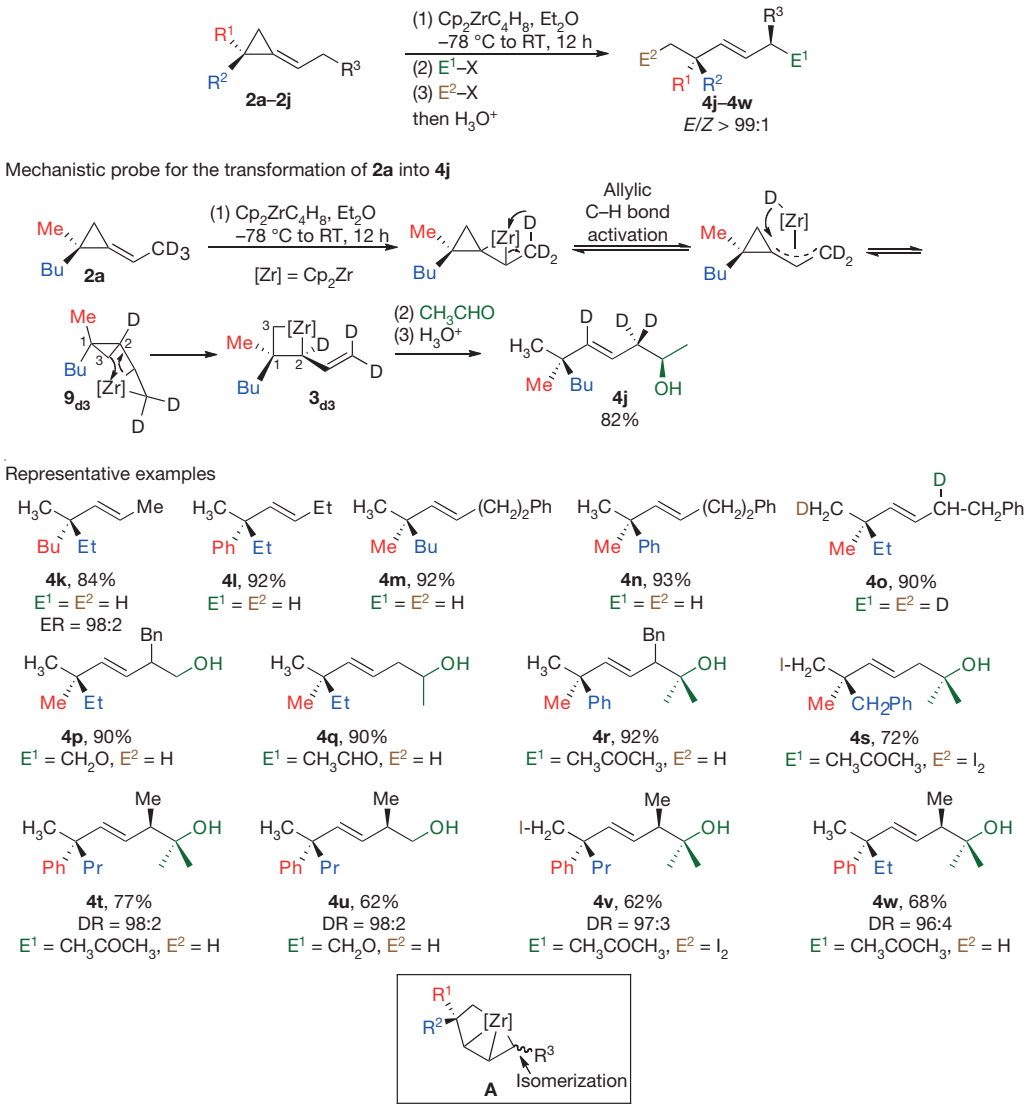



**Figure 3 | Zirconocene-promoted allylic C–H and C–C bond activation of alkylidenecyclopropanes.** When alkylidenecyclopropane **2a** is treated with $C_4H_8ZrCp_2$, an allylic C–H activation reaction occurs to give the intermediate **9**, which undergoes a selective C–C bond activation leading to the intermediate **3** (Fig. 2). A mechanistic probe and representative examples of substrate scope for the reaction of alkylidenecyclopropanes with $Cp_2ZrC_4H_8$ are shown. We note that intermediate **3** reacts, after heating at 60 °C in THF, with various carbonyl derivatives as first electrophiles to generate a new stereogenic centre at a remote position with very high diastereoselectivity before addition of the second electrophiles. If the reaction is performed without heating before the addition of the first electrophile, two diastereomers are obtained in a 3:1 ratio. Therefore, we proposed that the heating allows the isomerization of the substituted allyl fragment into the *E*-isomer (**A**). Bn, $PhCH_2$; DR, diastereomeric ratio; ER, enantiomeric ratio.

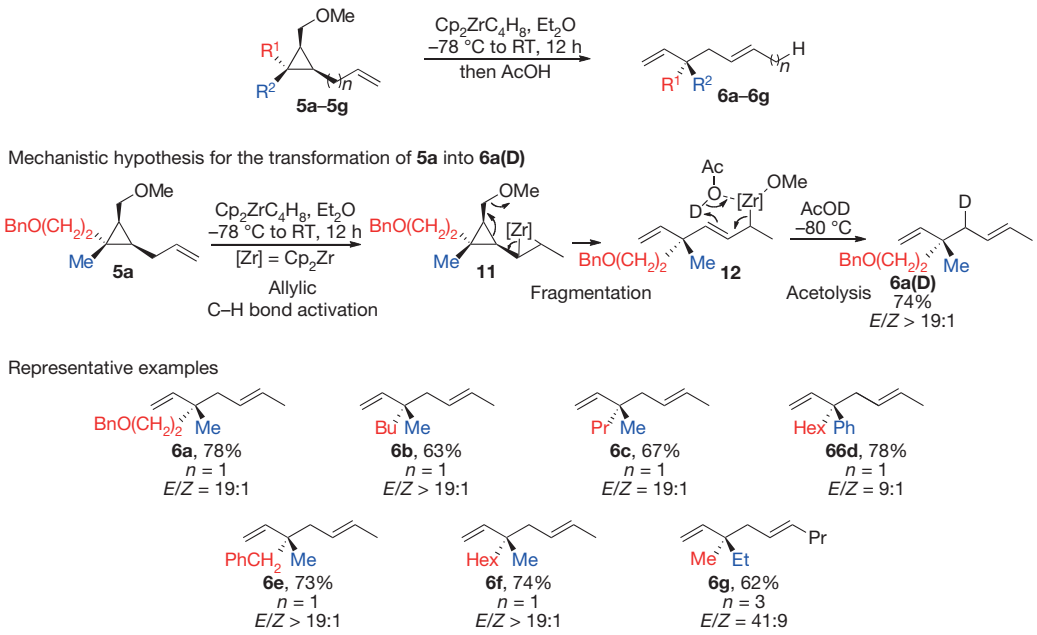



**Figure 4 | Merging allylic C–H activation and fragmentation en route to non-conjugated dienes.** $C_4H_8ZrCp_2$ reacts with the remote double bond of ω-ene-cyclopropane methyl ether **5** to induce an allylic C–H activation followed by a fragmentation of the three-membered ring. The resulting allylzirconocene, **12**, reacts selectively with AcOH to give only the non-conjugated diene **6** with an excellent *E*/*Z* selectivity. After the allylic C–H activation, the zirconacyclopropane **11** undergoes a fragmentation reaction to lead to the allylzirconocene species **12**. Acetolysis leads to non-conjugated dienes **6** possessing the all-carbon quaternary stereocentres. Ac, acetyl.

the same enantiomeric ratio (98:2). Regardless of the *E*/*Z* ratio of the starting ACPs, **2**, only the *E*-isomer was formed in the opened adducts (*E*/*Z* > 98:2). When an excess of $D_3O^+$ was added to the ring-opened product resulting from the treatment of alkylidenecyclopropane **2e** ($R^1$ = Me, $R^2$ = Et, $R^3$ = $CH_2$Ph) with the Negishi reagent, **4o** was isolated in 90% yield with a quantitative incorporation of deuterium at the allylic and terminal methyl positions. Consistent with our previous results in which the allylzirconocene [$C_2$–Zr] is more reactive to electrophiles than is alkylzirconocene [$C_3$–Zr], the addition of a first electrophile ($E^1$–X = HCHO) led to selective functionalization of the allyl fragment that was then hydrolysed under acidic work-up to give alcohol **4p** in 90% yield (Fig. 3). This selective reactivity of the first electrophile can be extended to other aldehydes and ketones without altering the yield of the overall transformation (that is, **4q** and **4r**, respectively). The remaining primary organozirconocene species can then be trapped by a second electrophile ($E^2X = I_2$) to give the bifunctionalized adduct **4s**.

This strategy holds potential for induction of 1,4-diastereoselectivity in acyclic systems through the transfer of chiral information from remote positions. To generate such remote diastereoselectivity, alkylidenecyclopropane **2i** ($R^1$ = Ph, $R^2$ = Pr, $R^3$ = Me) was treated under the same conditions with the Negishi reagent and the reaction mixture was heated to 60 °C in tetrahydrofuran (THF) for a few hours before addition of the first electrophile (Supplementary Information). In this case, the open adduct **4t** was obtained, after hydrolysis, with a diastereomeric ratio of 98:2 in good isolated yield. Interestingly, without heating the reaction mixture, a modest 3:1 ratio is obtained and may be attributed to conformational isomers of the substituted allylzirconocene fragment before reaction with the aldehyde; by heating to 60 °C, the most stable thermodynamically *E*-isomer is quantitatively formed (**A**; Fig. 3). This sequence can be extended to other versatile acyclic adducts while diastereomeric ratios remain excellent (formation of **4t**–**4w**). Finally, we applied this combined Zr-walk/C–C bond fragmentation strategy to ω-ene-cyclopropanes that possess a strategically placed leaving group on the cyclopropane ring (Fig. 4). When cyclopropanes **5a**–**5g**, which are easily prepared by carbometalation reaction of cyclopropenyl esters[27], were treated with the Negishi reagent, dienes **6a**–**6g** were obtained in good yields and with excellent *E*/*Z* ratios after hydrolysis. Mechanistically, the allylic C–H activation leads to zirconacyclopropane **11**, which undergoes fragmentation to give the allylzirconocene species **12**. Owing to the potential steric interactions between the zirconocene and the quaternary stereocentre, we proposed that only the isomer **12** would be present, with no evidence for metalotropic equilibrium, under its σ-bond form. Hence, **12** reacted with deuterated acetic acid at low temperature, by means of an $S_E2'$ reaction, to give **6a(D)** with an excellent positional selectivity. This sequence was then applied to several other differentially substituted cyclopropanes, **5a**–**5f**, to give non-conjugated dienes in all cases. This reaction is not restricted to the one-carbon tether; for example, **6g**, which has a three-carbon tether (*n* = 3), was also obtained in good yields. Because enantiomerically enriched cyclopropenyl esters can be easily prepared[27], the formation of non-conjugated dienes possessing an all-carbon quaternary stereocentre is therefore possible through this combined sequence of allylic C–H activation-ring fragmentation (Fig. 4).

We have shown that any cyclopropane species possessing a remote double bond such as ω-ene-cyclopropanes **1** and **5** or alkylidenecyclopropanes **2** can be easily transformed into acyclic molecular fragments possessing two stereogenic centres including the original challenging all-carbon quaternary stereocentre[28–30]. This process goes through zirconocene-mediated allylic C–H activations followed by highly selective C–C bond activations or fragmentations. Owing to the presence of the quaternary centre, the resulting bifunctional nucleophilic species are further derivatized with two different electrophiles to give more complex molecular architectures arising from simple starting materials. Overall, this work emphasizes the feasibility of synthetic approaches to exploit the multifold reactivity of a single substrate to furnish advanced molecular scaffolds through the combination of transformations that are otherwise difficult to perform.

## METHODS SUMMARY

To a solution of zirconocene dichloride (496.5 mg, 1.70 mmol) in THF (45 ml) cooled to −78 °C was added dropwise *n*-BuLi (1.40 M in hexane, 2.4 ml, 3.4 mmol). The resulting solution was stirred for 1 h at −78 °C. A solution of alkylidenecyclopropane

**2** (1.00 mmol) in THF (5 ml) was added, and the reaction mixture was slowly warmed to 10 °C. After stirring for 13 h, the reaction was heating to 60 °C for 6 h. After cooling the reaction to 10 °C, carbonyl derivative (4 mmol) was added at 10 °C and the reaction mixture was stirred for 40 min at the same temperature. Then the second electrophile (5 mmol) was added and the reaction mixture was stirred for 3 h. An aqueous solution of HCl (1 M) was then added at 0 °C and after extraction with $Et_2O$, the organic extracts were washed with water, $NaHCO_3$ and brine, dried ($MgSO_4$), filtered, and concentrated *in vacuo* to give a crude oil. The product **4** was obtained pure by silica-gel chromatography using gradient mixtures of ethyl acetate and *n*-hexane as eluents.

**Received 22 August; accepted 8 October 2013.**
**Published online 8 December 2013.**

1. Baran, P. S., Maimone, T. J. & Richter, J. M. Total synthesis of marine natural products without using protecting groups. *Nature* **446,** 404–408 (2007).
2. Wender, P. A. & Miller, B. L. Synthesis at the molecular frontier. *Nature* **460,** 197–201 (2009).
3. Trost, B. M. The atom economy: a search for synthetic efficiency. *Science* **254,** 1471–1477 (1991).
4. Bergman, R. G. Activation of alkanes with organotransition metal complexes. *Science* **223,** 902–908 (1984).
5. Engle, K. M., Mei, T.-S., Wasa, M. & Yu, J.-Q. Weak coordination as a powerful means for developing broadly useful C-H functionalization reactions. *Acc. Chem. Res.* **45,** 788–802 (2012).
6. Ackermann, L. Carboxylate-assisted ruthenium-catalyzed alkyne annulations by C-H/Het-H bond functionalizations. *Acc. Chem. Res.* http://dx.doi.org/10.1021/ar3002798 (4 February 2013).
7. Rybtchinski, B. & Milstein, D. Metal insertion into C–C bonds in solution. *Angew. Chem. Int. Ed.* **38,** 870–883 (1999).
8. Gozin, M., Weisman, A., Ben-David, Y. & Milstein, D. Activation of a carbon-carbon bond in solution by transition-metal insertion. *Nature* **364,** 699–701 (1993).
9. Masarwa, A. & Marek, I. Selectivity in metal-catalyzed carbon-carbon bond cleavage of alkylidenecyclopropanes. *Chem. Eur. J.* **16,** 9712–9721 (2010).
10. Seiser, T., Roth, O. A. & Cramer, N. Enantioselective synthesis of indanols from *tert*-cyclobutanols using a rhodium-catalyzed C–C/C-H activation sequence. *Angew. Chem. Int. Ed.* **48,** 6320–6323 (2009).
11. Aïssa, C. & Fürstner, A. A rhodium-catalyzed C−H activation/cycloisomerization tandem. *J. Am. Chem. Soc.* **129,** 14836–14837 (2007).
12. Aloise, A. D., Layton, M. E. & Shair, M. D. Synthesis of cyclooctenones using intramolecular hydroacylation. *J. Am. Chem. Soc.* **122,** 12610–12611 (2000).
13. Chinkov, N., Levin, A. & Marek, I. Unsaturated fatty alcohol derivatives as a new source of substituted allylzirconocenes. *Angew. Chem. Int. Ed.* **45,** 465–468 (2006).
14. Chinkov, N., Majumbar, S. & Marek, I. Stereoselective preparation of dienyl zirconocene complexes via a tandem allylic C-H bond activation-elimination sequence. *J. Am. Chem. Soc.* **125,** 13258–13264 (2003).
15. Chinkov, N., Levin, A. & Marek, I. A zirconium promenade – an efficient tool in organic synthesis. *Synlett* 501–514 (2006).
16. Werner, E. W., Mei, T.-S., Burckle, A. J. & Sigman, M. S. Enantioselective Heck arylations of acyclic alkenyl alcohols using a redox-relay strategy. *Science* **338,** 1455–1458 (2012).
17. Mei, T.-S., Werner, E. W., Burckle, A. J. & Sigman, M. S. Enantioselective redox-relay oxidative Heck arylations of acyclic alkenyl alcohols using boronic acids. *J. Am. Chem. Soc.* **135,** 6830–6833 (2013).
18. Negishi, E. & Takahashi, T. Organozirconium compounds in organic synthesis. *Synthesis* 1–18 (1988).
19. Harada, S., Kiyono, H., Nishio, R., Taguchi, T. & Hanzawa, Y. Complexation of vinylcyclopropanes with zirconocene−1-butene complex: application to the stereocontrolled synthesis of steroidal side chains. *J. Org. Chem.* **62,** 3994–4001 (1997).
20. Dimmock, P. W. & Whitby, R. J. Zirconium-mediated ring opening of cyclopropanes. *J. Chem. Soc. Chem. Commun.* 2323–2324 (1994).
21. Simaan, S., Goldberg, A. F. G., Rosset, S. & Marek, I. Metal-catalyzed ring-opening of alkylidenecyclopropanes as a new approach to the preparation of enantiomerically pure all-carbon quaternary stereocenters. *Chem. Eur. J.* **16,** 774–778 (2010).
22. Kliman, L. T., Mlynarski, S. N., Ferris, G. E. & Morken, J. P. Catalytic enantioselective 1,2-diboration of 1,3-dienes: versatile reagents for stereoselective allylation. *Angew. Chem. Int. Ed.* **51,** 521–524 (2012).
23. Chinkov, N., Majumbar, S. & Marek, I. New approach to the stereoselective synthesis of metallated dienes via an isomerization-elimination sequence. *J. Am. Chem. Soc.* **124,** 10282–10283 (2002).
24. Nakamura, I. & Yamamoto, Y. Transition metal-catalyzed reactions of methylenecyclopropanes. *Adv. Synth. Catal.* **344,** 111–129 (2002).
25. Marek, I., Simaan, S. & Masarwa, A. Enantiomerically enriched cyclopropene derivatives: versatile building blocks in asymmetric synthesis. *Angew. Chem. Int. Ed.* **46,** 7364–7376 (2007).
26. Simaan, S., Masarwa, A., Bertus, P. & Marek, I. Chiral cyclopropenylcarbinols as a new source of chiral alkylidenecyclopropane derivatives. *Angew. Chem. Int. Ed.* **45,** 3963–3965 (2006).
27. Delaye, P.-O., Didier, D. & Marek, I. Carbon-carbon bond activation en route to the formation of all-carbon quaternary stereogenic centers in acyclic system. *Angew. Chem. Int. Ed.* **52,** 5333–5337 (2013).
28. Minko, Y., Pasco, M., Lercher, L., Botoshansky, M. & Marek, I. Forming all-carbon quaternary stereogenic centres in acyclic systems from alkynes. *Nature* **490,** 522–526 (2012).
29. Das, J. P., Chechik, H. & Marek, I. Enantioselective construction of all-carbon quaternary stereogenic carbons in aldol reactions. *Nature Chem.* **1,** 128–132 (2009).
30. Das, J. P. & Marek, I. Enantioselective synthesis of all-carbon quaternary stereogenic centers in acyclic systems. *Chem. Commun.* **47,** 4593–4623 (2011).

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

**Acknowledgements** This research was supported by the Niedersachsen-Technion Research Cooperation Program, by the Israel Science Foundation administrated by the Israel Academy of Sciences and Humanities (140/12) and by the European Research Council under the European Community's Seventh Framework Programme (ERC grant agreement no. 338912). I.M. is holder of the Sir Michael and Lady Sobell Academic Chair.

**Author Contributions** A.M., D.D., T.Z., M.S., L.A. and I.M. planned, conducted and analysed experiments. I.M. had the idea for and directed the project, and wrote the manuscript with contributions from A.M. and D.D. All authors contributed to discussions.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.M. (chilanm@tx.technion.ac.il).

LETTER

doi:10.1038/nature12778

# Primitive layered gabbros from fast-spreading lower oceanic crust

Kathryn M. Gillis[1], Jonathan E. Snow[2], Adam Klaus[3], Natsue Abe[4], Álden B. Adrião[5], Norikatsu Akizawa[6], Georges Ceuleneer[7], Michael J. Cheadle[8], Kathrin Faak[1,9], Trevor J. Falloon[10], Sarah A. Friedman[11], Marguerite Godard[12], Gilles Guerin[13], Yumiko Harigane[14], Andrew J. Horst[15], Takashi Hoshide[16], Benoit Ildefonse[12], Marlon M. Jean[17], Barbara E. John[8], Juergen Koepke[18], Sumiaki Machi[6], Jinichiro Maeda[19], Naomi E. Marks[20], Andrew M. McCaig[21], Romain Meyer[22], Antony Morris[23], Toshio Nozaka[24], Marie Python[19], Abhishek Saha[25] & Robert P. Wintsch[26]

**Three-quarters of the oceanic crust formed at fast-spreading ridges is composed of plutonic rocks whose mineral assemblages, textures and compositions record the history of melt transport and crystallization between the mantle and the sea floor. Despite the importance of these rocks, sampling them *in situ* is extremely challenging owing to the overlying dykes and lavas. This means that models for understanding the formation of the lower crust are based largely on geophysical studies[1] and ancient analogues (ophiolites)[2–5] that did not form at typical mid-ocean ridges. Here we describe cored intervals of primitive, modally layered gabbroic rocks from the lower plutonic crust formed at a fast-spreading ridge, sampled by the Integrated Ocean Drilling Program at the Hess Deep rift. Centimetre-scale, modally layered rocks, some of which have a strong layering-parallel foliation, confirm a long-held belief that such rocks are a key constituent of the lower oceanic crust formed at fast-spreading ridges[3,6]. Geochemical analysis of these primitive lower plutonic rocks—in combination with previous geochemical data for shallow-level plutonic rocks, sheeted dykes and lavas—provides the most completely constrained estimate of the bulk composition of fast-spreading oceanic crust so far. Simple crystallization models using this bulk crustal composition as the parental melt accurately predict the bulk composition of both the lavas and the plutonic rocks. However, the recovered plutonic rocks show early crystallization of orthopyroxene, which is not predicted by current models of melt extraction from the mantle[7] and mid-ocean-ridge basalt differentiation[8,9]. The simplest explanation of this observation is that compositionally diverse melts are extracted from the mantle and partly crystallize before mixing to produce the more homogeneous magmas that erupt.**

The gabbroic rocks that make up the lowermost oceanic crust formed at fast-spreading ridges, such as the East Pacific Rise (EPR), have long been assumed to be modally layered and primitive in composition[4,6,10,11]. Igneous layering, and a layering-parallel foliation, are nearly ubiquitous in the lower plutonic sections of many ophiolites[3,5,10], and explaining the formation of these layered rocks has become central to models for the accretion of the plutonic crust at fast-spreading ridges[2,4,5]. Accretion models have evolved from describing layers accumulating along the floors of large magma bodies[3,6] to describing layers developing in sill-like magma bodies focused at the top of the crystal mush within axial magma chambers or distributed throughout the crustal mush zone, or both[2,4,5]. Until now, however, no significant cored intervals of layered gabbros have been recovered from the lower plutonic section at a modern fast-spreading ridge.

Integrated Ocean Drilling Program (IODP) Expedition 345 was conceived as a test of whether modern fast-spreading crust shows layering similar to many ophiolites and as a test of models for the transport of parental melts into the crust and the differentiation of these melts within the crust. To sample the generally inaccessible lower plutonic crust, the expedition took advantage of the tectonic window at the Hess Deep rift (HDR) in the equatorial Pacific Ocean[11,12] (Fig. 1). This site is unique in that it is the only place where the lower crust and the upper crust have been extensively sampled by submersible or remotely operated underwater vehicle[11–14] and drilling[15] (Ocean Drilling Program (ODP) Leg 147), and previous studies of known sea-floor exposures of lower plutonic rocks have suggested that layering exists[11,12]. At IODP Site U1415, primitive olivine gabbros and troctolites were recovered at one 35-m-deep hole (U1415I) and two ~110-m-deep holes (U1415J and U1415P), located within 100 m of each other (Extended Data Fig. 1). Sampling of primitive layered gabbro and troctolite series at Site U1415 thus provides the final part of the most complete composite section of fast-spreading EPR crust so far.

Simple modally layered and irregularly banded rocks, collectively called the layered gabbro series, were recovered from all three drill holes, comprising >50% of the core. The layered gabbro series from holes U1415I and U1415J show simple modal layering, with or without concurrent grain size variations, on a scale of centimetres to decimetres (Fig. 2a and Extended Data Fig. 2). Layers include troctolite, olivine gabbro, gabbro and gabbronorite with local variations in texture (for example the presence or absence of clinopyroxene oikocrysts). Layering is reminiscent of 'dynamic' layering resulting from magmatic flow[16] commonly found in layered mafic intrusions and some ophiolites. A layering-parallel foliation exists throughout these rocks that is commonly

[1]School of Earth and Ocean Sciences, University of Victoria, PO Box 1700 Station CSC, Victoria, British Columbia V8W 2Y2, Canada. [2]Earth and Atmospheric Sciences, University of Houston, 312 Science and Research Building 1, Houston, Texas 77204-5007, USA. [3]United States Implementing Organization, Integrated Ocean Drilling Program, Texas A&M University, 1000 Discovery Drive, College Station, Texas 77845, USA. [4]Institute for Research on Earth Evolution, Japan Agency for Marine-Earth Science and Technology, 2-15 Natsushima-cho, Yokosuka 237-0061, Japan. [5]Institute of Geoscience, Rio Grande do Sul Federal University, Avenida Bento Gonçalves, 9500 Bloco I Prédio, 43113 Sala 207, Porto Alegre/RS 91501-970, Brazil. [6]Department of Earth Sciences, Kanazawa University, Kakuma-machi, Kanazawa Ishikawa 920-1192, Japan. [7]Observatoire Midi-Pyrénées (UMS 831), CNRS, Université Paul Sabatier, 14 Avenue Edouard Belin, Toulouse Cedex 31400, France. [8]Department of Geology and Geophysics, University of Wyoming, 1000 University Avenue, Department 3006, Laramie, Wyoming 82071, USA. [9]Institut für Geologie, Mineralogie, and Geophysik, Ruhr-Universität Bochum, Universitätsstrasse 150, 44780 Bochum, Germany. [10]Institute for Marine and Antarctic Studies and School of Earth Sciences, University of Tasmania, Hobart, Tasmania 7001, Australia. [11]Department of Geology, Southern Illinois University at Carbondale, Carbondale, Illinois 62901, USA. [12]Géosciences Montpellier, Université Montpellier 2, CNRS UMR5243, CC 60, 34095 Montpellier Cedex 5, France. [13]Borehole Research Group, Lamont-Doherty Earth Observatory of Columbia University, PO Box 1000, 61 Route 9W, Palisades, New York 10964, USA. [14]Institute of Geology and Geoinformation, Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology, AIST Tsukuba Central 7, 1-1-1 Higashi, Tsukuba Ibaraki 305-8567, Japan. [15]Department of Geology, Oberlin College, Oberlin, Ohio 44074, USA. [16]Graduate School of Science, Tohoku University, Aoba-ku, Sendai 980-8578, Japan. [17]Department of Geology and Environmental, Geosciences, Northern Illinois University, Davis Hall 312, Normal Road, DeKalb, Illinois 60115, USA. [18]Institut für Mineralogie, University of Hannover, Callinstrasse 3, Hannover 30167, Germany. [19]Department of Natural History Sciences, Hokkaido University, North 10, West 8, Kita-ku, Sapporo, Hokkaido 060-0810, Japan. [20]Chemistry and Material Sciences Department, Lawrence Livermore National Laboratory, PO Box 808, L-231, Livermore, California 94551, USA. [21]School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK. [22]Centre for Geobiology and Department of Earth Science, University of Bergen, Allegaten 41, Bergen 5007, Norway. [23]School of Geography, Earth and Environmental Sciences, Plymouth University, Drake Circus, Plymouth PL4 8AA, UK. [24]Department of Earth Sciences, Okayama University, 3-1-1 Tsushima-naka, Okayama 700-8530, Japan. [25]Department of Geology, University of Calcutta, 35 Ballygunge Circular Road, Kolkata 700 019, India. [26]Department of Geological Sciences, Indiana University, 1001 East 10th Street, Bloomington, Indiana 47405, USA.

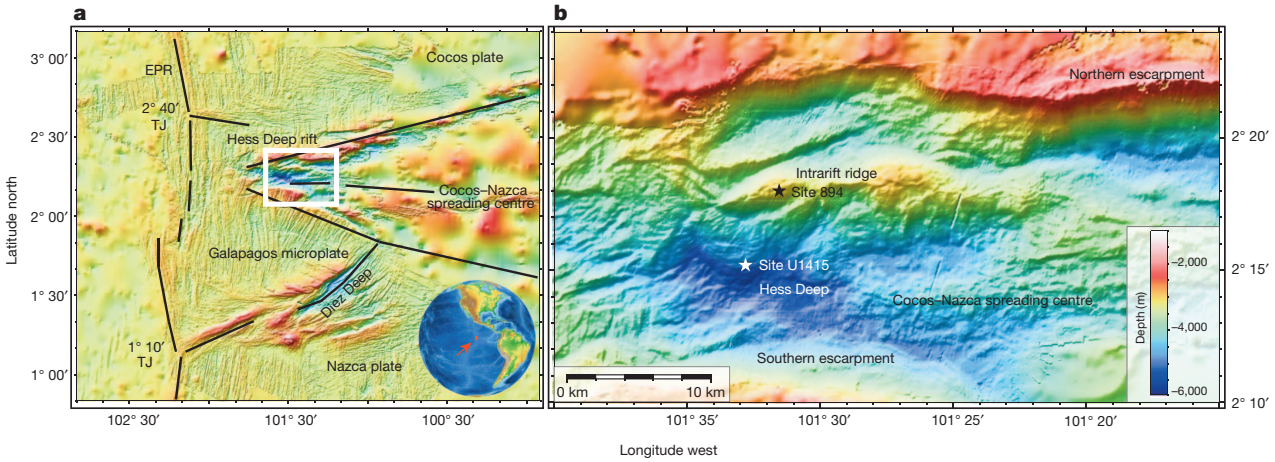



**Figure 1 | Tectonic setting of the HDR and the location of IODP Site U1415.** The HDR formed by deep lithospheric extension in front of the westwards-propagating Cocos–Nazca spreading centre, exposing oceanic crust that formed at the fast-spreading (130 mm $yr^{-1}$) EPR. Upper crustal lavas and dykes are exposed along the northern and southern escarpments, shallow-level gabbros along the northern escarpment and the western intrarift ridge, and lower-level gabbros along the southern slope of the intrarift ridge[11–14]. **a**, Map of the Galapagos triple junction (TJ) in the eastern equatorial Pacific Ocean showing major tectonic boundaries. The white box indicates the location of the map in **b**. **b**, Regional bathymetric map of the HDR showing key morphological features and the locations of IODP Site U1415 and ODP Site 894. Maps show bathymetry derived from satellite altimetry data and archived multibeam bathymetry data available from the Global Multi-Resolution Topography Data Portal at the Lamont-Doherty Earth Observatory.

strong (Fig. 2c) and locally anastomoses around large oikocrysts. In contrast, the layered gabbro series in Hole U1415P shows irregular banding that is identified by modal and grain size variations, with all of the same lithologies present in the simple layers of holes U1415I and U1415J, but also includes rare anorthositic bands (Fig. 2b). Grain size variation is much more extreme and heterogeneous than in the simple layers, bands can be discontinuous, and one lithology can enclose another. Additionally, the boundaries between bands are generally less planar than the simple layers and show abrupt changes in mineralogy leading to asymmetric distributions of distinct leucocratic and melanocratic bands (Fig. 2b), and mineral foliations are weak or absent. This banding is reminiscent of non-dynamic layering in layered mafic intrusions that is the result of varying rates of nucleation and growth, and postcumulus processes[17,18]. The troctolite series at the base of holes U1415J and U1415P contain melanocratic to leucocratic troctolite with little or no layering or banding and a weak-to-moderate foliation (Extended Data Fig. 1). Evolved lithologies such as FeTi oxide gabbros and felsic veins, which are prevalent in the upper gabbros[15], are strikingly absent throughout the cores, indicating that evolved residual melt was efficiently extracted from the lower plutonic crust. Also absent are ultramafic rocks, suggesting the recovered lithologies are not part of the mantle transition zone.

The foliation and layering in the layered gabbro series provide important constraints on the processes of crustal accretion. There is little subsolidus crystal plastic deformation, meaning that the foliations were formed early, while the rocks were still partly molten. In addition, olivine commonly exhibits skeletal morphologies, which limits the amount of grain-scale strain that some of the rocks may have experienced at low melt fractions. The abundance of layering in the material recovered from Site U1415, along with the absence of intermixed evolved lithologies, distinguishes the HDR lower gabbroic crust from crustal sections recovered from slow-spreading ridges (see, for example, ref. 19). This supports models that invoke a strong spreading rate and thermal control on magma chamber processes at mid-ocean ridges[20]. Furthermore, the occurrence of layering that resembles both dynamic[16] and non-dynamic[17] layering in layered mafic intrusions suggests multiple mechanisms of crustal accretion and melt differentiation. This variation in style of layering and banding, and the diversity of lithologies, differs from the mid-ocean-ridge basalt (MORB)-like, southern portions of the Oman ophiolite, which has been used as a fast-spreading-ridge analogue[3,5]. The primary difference is that simple modal layering dominates and irregular banding is very rare in this region of Oman[3].

The Site U1415 cores are much more primitive (high MgO and high Mg number, Mg# = 100Mg/(Mg + Fe)) than previously recovered samples of the overlying upper gabbros (Fig. 3). These new samples allow estimation of the bulk composition of fast-spreading oceanic crust because the thickness of the crust and its component parts can be estimated along with uncertainties from field relationships determined from four previous submersible surveys[11,13,14] and ODP Leg 147 drilling[15] (Methods). The relative proportions of the upper crust (lavas and dykes form 22.5% ± 4.5% of the crust) and shallow-level gabbros (32.5% ± 7.5% of the plutonics) constrain the plausible fraction of deeper gabbros. For these calculations, we use the mean compositions of the upper crust (lavas and sheeted dykes) from the northern escarpment, the shallow-level gabbros from the northern escarpment and ODP Site 894, and the lower gabbros from IODP Site U1415 (see Fig. 1b for locations, and Extended Data Table 1).

The calculated bulk composition of the HDR crust (Extended Data Table 1) contains 12.1 ± 1.0 wt% MgO, 7.7 ± 0.4 wt% $FeO^{Total}$ (all iron as FeO) and has a Mg number of 74 ± 1, which falls at the Mg-rich end of the range of experimentally determined parental melts for MORBs[21]. The calculated liquidus temperature is ~1,290–1,300 °C, and fractional crystallization models follow the expected MORB sequence of olivine (Fo ≈ 90–91 (Fo = 100Mg/Mg + Fe)) followed by plagioclase (An ≈ 84–87 (An = 100Ca/Ca + Na)) then clinopyroxene (Mg# ≈ 86–88), with orthopyroxene (Mg# ≈ 75–82) saturation at ~1,165–1,180 °C (refs 22, 23) with 46–35% melt remaining[22,23] (Fig. 3 and Methods). The first olivine to be produced in these models has a similar Fo content to the olivine in dunites and harzburgites recovered from Hess Deep by previous drilling (Fo content, 89.4–91.3; refs 24, 25); this is consistent with the bulk-crustal composition being representative of a primary mantle melt extracted either directly from the harzburgites or through dunite channels[7,24].

An unexpected finding is that cumulus orthopyroxene commonly occurs in primitive (bulk-rock $Mg\#_{80–85}$) plutonic rocks from the deep portion of the crust at the HDR (Fig. 2d). Orthopyroxene is a common minor cumulus phase and an early intercumulus phase (<5%) in olivine gabbro, gabbro, gabbronorite and troctolite in the layered gabbro series. In contrast, the virtual absence of orthopyroxene as a phenocryst in MORB globally (including HDR), as well as experimental studies of MORB differentiation (see, for example, ref. 8) and modelling of the differentiation using our bulk-crustal composition (Fig. 3), all indicate that orthopyroxene should not be a liquidus phase until >50% crystallization has occurred with a substantial interval of clinopyroxene

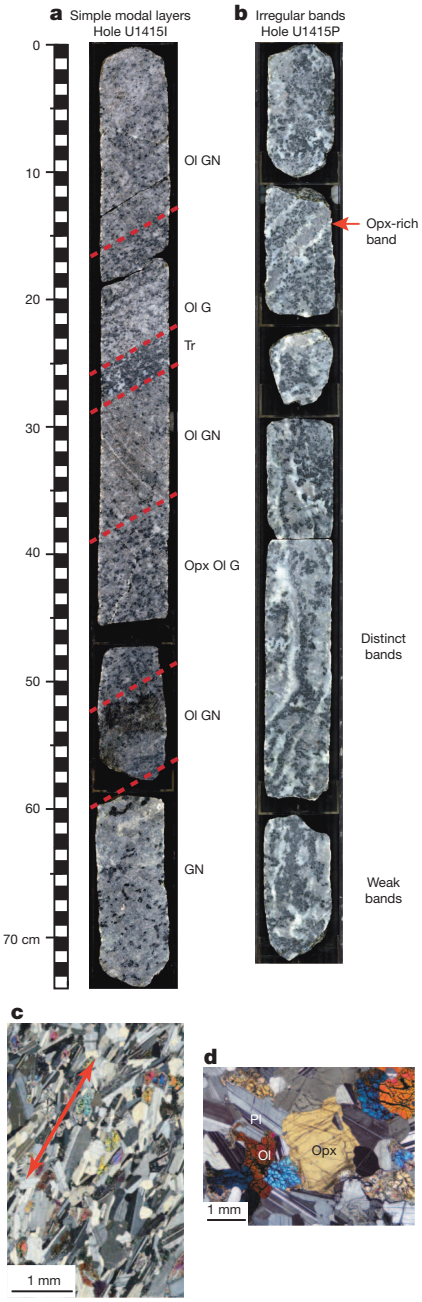



**Figure 2 | Typical gabbroic rocks at Site U1415.** **a**, Simple, centimetre-scale modal layers of alternating lithology (labelled to right of core) (Hole U1415I, section 4R-1, 47–115 cm); layer boundaries (red dashed lines) are sharp and planar, and a layering-parallel foliation is seen throughout. Additional examples are given in Extended Data Fig. 2. **b**, Orthopyroxene-bearing olivine gabbro showing irregular, steeply dipping leucocratic and melanocratic bands that range from distinct to weak (Hole U1415P, section 8R-1, 86–144 cm). Banding is defined by modal, grain size and textural variations. Also of note is an orthopyroxene-rich band. **c**, Photomicrograph, under cross-polarized light, of a troctolite in the layered gabbro series showing a strong magmatic foliation indicated by the red arrow (sample U1415I, section 4R-2, 0–4 cm, piece 2). **d**, Photomicrograph, under cross-polarized light, showing cumulus orthopyroxene in an orthopyroxene-bearing olivine gabbro (sample U1415P, section 4G-1, 5–7 cm, piece 2). Ol G, olivine gabbro; Tr, troctolite; GN, gabbronorite; Ol GN, olivine-bearing gabbronorite; Opx Ol G, orthopyroxene-bearing olivine gabbro; Ol, olivine; Opx, orthopyroxene; Pl, plagioclase.

crystallization preceding orthopyroxene saturation. Such late-stage orthopyroxene is commonly found in more evolved gabbros, including the shallow-level gabbros from the HDR[15].

Orthopyroxene is ubiquitous in the upper mantle, where its coexistence with olivine buffers the silica activity in primary mantle melts.

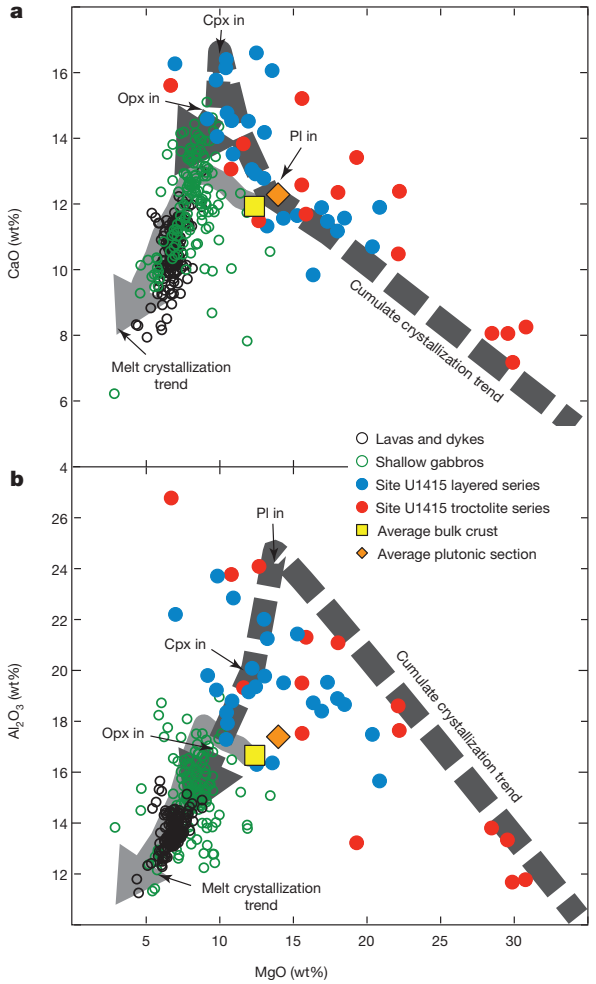



**Figure 3 | Variations in whole-rock CaO and $Al_2O_3$ with MgO for different parts of the crust at the HDR.** **a**, CaO; **b**, $Al_2O_3$. The lower plutonics recovered from Site U1415 are much more primitive than the shallow plutonics. The compositions of the lower-gabbro cumulates and lavas and dykes can be broadly explained using a simple fractional crystallization model (grey arrows) of the bulk-crustal composition, but the shallow gabbros clearly contain substantial trapped melt in the bulk composition (that is, they are mixtures of cumulate and melt compositions). The uncertainties for the bulk-crustal and plutonic-section compositions are smaller than their symbol size (Extended Data Table 1). Fractional crystallization trends (showing instantaneous compositions) for the melt (thick solid light-grey arrow) and cumulates (thick dashed dark-grey arrow) calculated using output of the PETROLOG program[23] schematically encompass the range of models considered (Methods). The first appearance of each mineral modelled is shown for the cumulate crystallization trend: olivine is the liquidus phase, plagioclase appears after ~7% crystallization, clinopyroxene appears after ~32% to 35% and orthopyroxene appears after ~55% to 65%.

The expected late crystallization of orthopyroxene in MORB is due to the generation of MORB by means of polybaric, near-fractional melting[26,27], with an average melting pressure of about 10 kbar (ref. 27). Decompression of melts aggregated from throughout the melting column leads to an expansion of the olivine stability field and shrinking of the orthopyroxene stability field, leaving the low-pressure melt far from orthopyroxene saturation[28]. Although several processes could explain the occurrence of orthopyroxene in the deep primitive gabbros at the HDR, most seem unlikely. For example, the parental magmas could be more oxidized than typical MORB. This would lead to less of the Fe in the melt being divalent and, hence, available to partition into mafic phases[29], and could also lead to early oxide saturation driving an increase in silica activity, both of which could lead to early orthopyroxene saturation. However, this model is difficult to reconcile with either the normal

differentiation trends observed in the overlying lavas and dykes, including Fe-enrichment trends[30], or the virtual absence of FeTi oxides (typically ≤0.1 modal per cent) in the Site U1415 rocks. Another possibility is that orthopyroxene saturation is influenced by the addition and removal of $H_2O$ from the system[31]. The low water content of primitive MORB[32] and the observation that the Site U1415 cumulates contain almost no magmatic amphibole suggest very limited $H_2O$ in the system, making this possibility unlikely. The most reasonable explanation is that orthopyroxene was precipitated from a primitive melt that had undergone little decompression since being in (major-element) equilibrium with shallow mantle orthopyroxene. This can be explained if this melt either was generated by shallow mantle melting, or re-equilibrated with the shallow mantle as it was transported through it, and crystallized within the crust without first mixing with aggregated MORB melts in the crust[33]. Partial re-equilibration of melt during transport through the shallow mantle is supported by the relatively high Sr content of the primitive cumulates, which suggests that their parental melt was not depleted in incompatible elements. If this model is correct, it indicates that diverse melt compositions feed the crust and that the lower crust acts as an efficient filter for mixing these before the eruption of their homogenized and differentiated products.

Overall, our findings demonstrate that although the bulk oceanic crust at the HDR has a similar composition to that predicted for parental MORB, the diversity in parental melts added into the crust is greater than expected. Partial crystallization of these diverse melts occurs before mixing, something that is not considered in models of MORB differentiation. However, such melts are not erupted, indicating that melt transport through the lower crust acts as an efficient mechanism to homogenize the Moho-crossing melts. The heterogeneity in the lithologies, bulk compositions, layering types and foliation strength observed within the Site U1415 core suggests complex melt differentiation and crustal accretion processes in the lower crust at fast-spreading ridges.

## METHODS SUMMARY

The bulk composition of the EPR crust exposed at the HDR and its uncertainty were calculated using new data and published compositions and relative mass fractions of the main crustal lithologies: lavas and dykes, shallow-level gabbros, and deep-level gabbros (Extended Data Table 1). The mass fractions of the crustal lithologies and their uncertainties are derived from field observations[11–14]. A series of models[22,23] was used to investigate whether differentiation of a parental melt with the composition of the bulk crust would produce cumulates and residual melts of similar composition to the observed plutonics and upper crustal rocks at the HDR. A fuller description is given in Methods.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 19 April; accepted 10 October 2013.**

**Published online 1 December 2013.**

1. Detrick, R. S. *et al.* Multichannel seismic imaging of a crustal magma chamber along the East Pacific Rise. *Nature* **326,** 35–41 (1987).
2. Kelemen, P. B. & Aharanov, E. in *Faulting and Magmatism at Mid-Ocean Ridges* (eds Buck, W. R., Delaney, P. T., Karson, J. A. & Lagrabrielle, Y.) 267–289 (Geophys. Monogr. Ser. 106, American Geophysical Union, 1998).
3. Pallister, J. S. & Hopson, C. A. Samail ophiolite plutonic suite: field relations, phase variation, cryptic variation and layering, and a model of a spreading ridge magma chamber. *J. Geophys. Res.* **86,** 2593–2644 (1981).
4. Quick, J. E. & Denlinger, R. P. Ductile deformation and the origin of layered gabbro in ophiolites. *J. Geophys. Res.* **98,** 14015–14027 (1993).
5. Boudier, F., Nicolas, A. & Ildefonse, B. Magma chambers in the Oman ophiolite: fed from the top and the bottom. *Earth Planet. Sci. Lett.* **144,** 239–250 (1996).
6. Cann, J. R. A model for oceanic crustal structure developed. *Geophys. J. R. Astron. Soc.* **39,** 169–187 (1974).
7. Kelemen, P. B., Shimizu, N. & Salters, V. J. M. Extraction of mid-ocean-ridge basalt from the upwelling mantle by focused flow of melt in dunite channels. *Nature* **375,** 747–753 (1995).
8. Grove, T. L., Kinzler, R. J. & Bryan, W. B. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Phipps Morgan, J., Blackman, D. K. & Sinton, J. M.) 281–311 (Geophys. Monogr. Ser. 71, American Geophysical Union, 1992).
9. O'Neill, H., St C. & Jenner, F. E. The global pattern of trace-element distributions in ocean floor basalts. *Nature* **491,** 698–704 (2012).
10. Nicolas, A., Boudier, F. & Ceuleneer, G. Mantle flow patterns and magma chambers at ocean ridges: evidence from the Oman ophiolite. *Mar. Geophys. Res.* **9,** 293–310 (1988).
11. Francheteau, J. *et al.* 1 Ma East Pacific Rise oceanic crust and uppermost mantle exposed by rifting in Hess Deep (equatorial Pacific Ocean). *Earth Planet. Sci. Lett.* **101,** 281–295 (1990).
12. Hékinian, R., Bideau, D., Francheteau, J., Lonsdale, P. & Blum, N. Petrology of the East Pacific Rise crust and upper mantle exposed in the Hess Deep (eastern equatorial Pacific). *J. Geophys. Res.* **98,** 8069–8094 (1993).
13. Karson, J. A. *et al.* Structure of uppermost fast-spread oceanic crust exposed at the Hess Deep Rift: implications for subaxial processes at the East Pacific Rise. *Geochem. Geophys. Geosyst.* **3,** 2001GC000155 (2002).
14. Lissenberg, C. J., MacLeod, C. J., Howard, K. A. & Godard, M. Pervasive reactive melt migration through fast-spreading lower oceanic crust (Hess Deep, equatorial Pacific Ocean). *Earth Planet. Sci. Lett.* **361,** 436–447 (2013).
15. Gillis, K. M., Mével, C. & Allan, J. (eds). *Proc. Ocean Drilling Program, Initial Reports* Vol. 147, 45–108 (Ocean Drilling Program, 1993).
16. McBirney, A. R. & Nicolas, A. The Skaergaard layered series. Part II. Magmatic flow and dynamic layering. *J. Petrol.* **38,** 569–580 (1997).
17. Boudreau, A. E. & McBirney, A. R. The Skaergaard layered series. Part III. Non-dynamic layering. *J. Petrol.* **38,** 1003–1020 (1997).
18. Naslund, H. R. & McBirney, A. R. in *Layered Igneous Intrusions* (ed. Cawthron, R. G.) 1–43 (Dev. Petrol. 15, Elsevier, 1996).
19. Blackman, D. K. *et al.* Drilling constraints on lithospheric accretion and evolution at Atlantis Massif, Mid-Atlantic Ridge 30 degrees N. *J. Geophys. Res.* **116,** B07103 (2011).
20. Henstock, T. J., Woods, A. W. & White, R. S. The accretion of oceanic crust by episodic sill intrusion. *J. Geophys. Res.* **98,** 4143–4161 (1993).
21. Kinzler, R. J. & Grove, T. L. Primary magmas of mid-ocean ridge basalts. 2. Applications. *J. Geophys. Res.* **97,** 6907–6926 (1992).
22. Ghiorso, M. S. & Sack, R. O. Chemical mass transfer in magmatic processes IV. A revised and internally consistent thermodynamic model for the interpolations of liquid-solid equilibria in magmatic systems at elevated temperatures and pressures. *Contrib. Mineral. Petrol.* **119,** 197–212 (1995).
23. Danyushevsky, L. V. & Plechov, P. Petrolog3: integrated software for modeling crystallization processes. *Geochem. Geophys. Geosyst.* **12,** Q07021 (2011).
24. Dick, H. J. B. & Natland, J. H. in *Proc. Ocean Drilling Program, Scientific Results* Vol. 147 (eds Mével, C., Gillis, K. N., Allan, J. F. & Meyer, P. S.) 103–134 (Ocean Drilling Program, 1996).
25. Arai, S. & Matsukage, K. in *Proc. Ocean Drilling Program, Scientific Results* Vol. 147 (eds Mével, C., Gillis, K. N., Allan, J. F. & Meyer, P. S.) 135–155 (Ocean Drilling Program, 1996).
26. Johnson, K. T. M., Dick, H. J. B. & Shimizu, N. Melting in the oceanic upper mantle: an ion microprobe study of diopsides in abyssal peridotites. *J. Geophys. Res.* **95,** 2661–2678 (1990).
27. Langmuir, C. H., Klein, E. M. & Plank, T. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Phipps Morgan, J., Blackman, D. K. & Sinton, J. M.) 183–280 (Geophys. Monogr. Ser. 71, American Geophysical Union, 1992).
28. O'Hara, M. J. Are ocean floor basalts primary magma? *Nature* **220,** 683–686 (1968).
29. Grove, T. L. & Juster, T. C. Experimental investigations of low-Ca pyroxene stability and olivine pyroxene liquid equilibria at 1-Atm in natural basaltic and andesitic liquids. *Contrib. Mineral. Petrol.* **103,** 287–305 (1989).
30. Stewart, M. A., Klein, E. M. & Karson, J. Geochemistry of dikes and lavas from the north wall of the Hess Deep Rift: insights into the four-dimensional character of crustal construction at fast-spreading mid-ocean ridges. *J. Geophys. Res.* **107,** 2238 (2002).
31. Berndt, J., Koepke, J. & Holtz, F. An experimental investigation of the influence of water and oxygen fugacity on differentiation of MORB at 200 MPa. *J. Petrol.* **46,** 135–167 (2005).
32. Saal, A. E., Hauir, E. H., Langmuir, C. H. & Perfit, M. R. Vapour undersaturation in primitive mid-ocean-ridge basalt and the volatile content of Earth's upper mantle. *Nature* **419,** 451–455 (2002).
33. Coogan, L. A., Gillis, K. M., MacLeod, C. J., Thompson, G. M. & Hékinian, R. Petrology and geochemistry of the lower ocean crust formed at the East Pacific Rise and exposed at Hess Deep: a synthesis and new results. *Geochem. Geophys. Geosyst.* **3,** 8604 (2002).

**Acknowledgements** This research used samples and data provided by the IODP. We thank the USIO staff and the Siem Offshore crew for recovering hard rock core in an unsedimented environment at >4,850-m water depth, and for their invaluable assistance during the expedition. We gratefully acknowledge the contributions of the drilling proposal proponents and the leaders (C. MacLeod and D. Teagle) and participants of the site survey cruise (JC21) in making IODP Expedition 345 possible. We thank E. Klein and H. Dick for reviews.

**Author Contributions** All authors were shipboard participants on IODP Expedition 345, contributed to the shipboard data collection, and discussed the results and their implications. K.M.G. wrote the first draft of the manuscript, K.M.G. and J.E.S. were co-chief scientists on the expedition, and A.K. was the staff scientist.

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

## METHODS

**Calculation of the bulk composition of the HDR crust.** Calculation of the bulk composition of the crust requires knowledge of the compositions and relative mass fractions of the crustal lithologies. We divide the crust into three sections: (i) the upper crust (lavas and dykes), (ii) shallow-level gabbros and (iii) deep-level gabbros. Seismic velocity–depth models for undisrupted EPR crust north of the HDR indicate a crustal thicknesses of ~5.6 km (G. Christeson, personal communication). Field relationships constrain the thickness of the upper crust to be ~1.25 ± 0.25 km (1 s.d.; ref. 13), and the subjacent plutonic sequence is therefore 4.1–4.6 km thick. The respective volumes of the upper crust and plutonic sequence were converted to mass fractions on the basis of their density differences.

The composition of the upper crust is based on an extensive sample suite collected by submersible along vertical transects through the lava and dyke sections of the northern escarpment of the HDR between longitudes 101° 13.5′ and 101° 28.5′ west (Fig. 1b). This representative data set includes whole-rock lava and dyke compositions[30] ($n = 157$) and glass compositions[30,34] ($n = 18$). A series of test calculations were performed to assess how using either just the glasses or an average of the glass and whole-rock data impact the resulting bulk-crustal composition; minimal difference in bulk-crustal composition was found and the mean of the two data sets was therefore used to define the bulk composition of the upper crust and its uncertainty. The bulk-crustal modelling assumes the upper crust comprises ~22.5% ± 4.5% (1 s.d.) of the total crustal thickness.

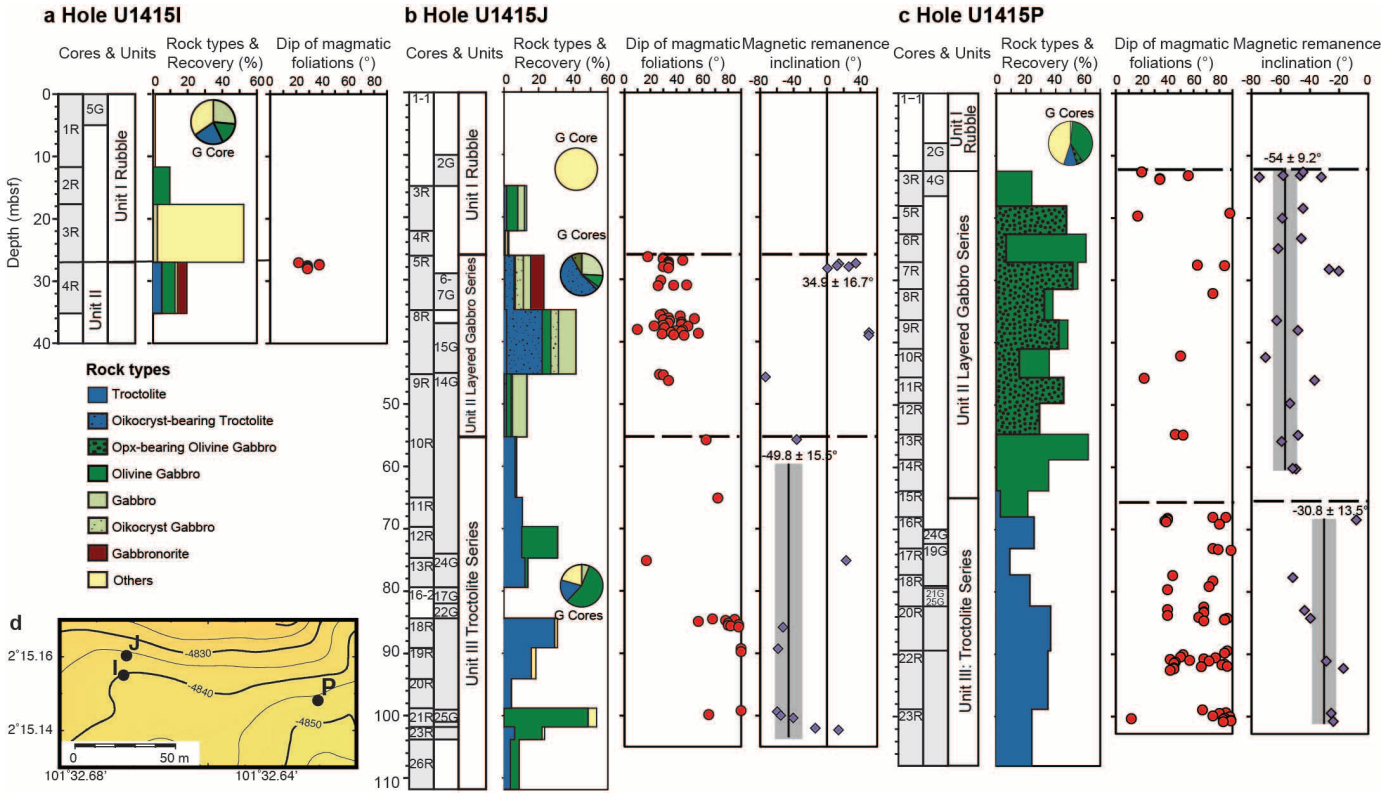
The shallow-level gabbro compositions are also based on an extensive sample suite, including samples collected by submersible from the northern escarpment where the upper 1 km of the plutonic section is well exposed subjacent to the sheeted dyke complex, across a horizontal distance of 3 km (ref. 13); by drilling a 150-m-long section from the upper 1 km of the plutonic sequence at ODP Site 894[15]; and by submersible from the western end of the intrarift ridge in the vicinity of ODP Site 894[11,12]. The mean composition of the samples from the northern escarpment[35,36] ($n = 56$), ODP Site 894[15,37] ($n = 76$) and the western end of the intrarift ridge[38] ($n = 7$) was used to define the bulk composition of the shallow gabbros and its uncertainty. The proportion of the shallow-level gabbros in the plutonic section is estimated using field relationships from the HDR (see above) and the Oman ophiolite[39], which show them to comprise >20%–25% and 20%–50% of the plutonic section, respectively. The bulk-crustal modelling assumes that the shallow-level gabbros comprise ~32.5% ± 7.5% (1 s.d.) of the plutonic sequence.

The lower-level gabbro compositions are calculated from the Site U1415 samples, using the compositions of the layered gabbro series ($n = 28$) and troctolite series ($n = 15$). The proportions of the layered gabbro and troctolite series are approximately equal at IODP holes U1415J and U1415P (the two >100-m-deep drill holes), and we thus model their relative proportions when calculating the bulk crustal composition and its uncertainty as 50% ± 20%.

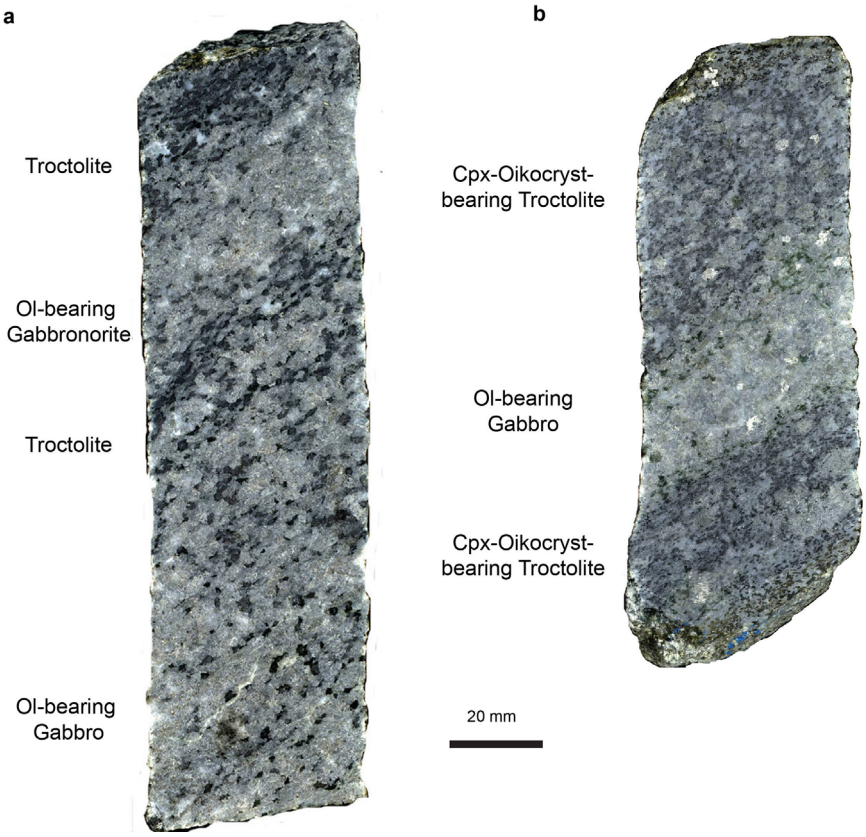
The mean compositions of the different sections of the crust calculated as explained above are given in Extended Data Table 1, along with the calculated bulk-crustal and bulk-plutonic-sequence compositions. The uncertainties in the mass fractions of each portion of the crust, along with the uncertainties in their average compositions, were propagated into the uncertainty in the bulk-crustal composition using a Monte Carlo error propagation assuming all errors are Gaussian.

**Modelling melt differentiation.** A series of models was run to investigate whether differentiation of a parental melt with the composition of the bulk crust would produce cumulates and residual melts of similar composition to the observed plutonics and upper crustal rocks at the HDR. Both the MELTS[22] model and the PETROLOG[23] model were used to test how sensitive the results are to the model calibration. The models all assumed perfect fractional crystallization at 1 kbar and oxygen fugacities between the quartz–magnetite–fayalite buffer and one log unit below this buffer. Although perfect fractional crystallization is unlikely in oceanic crust (see, for example, ref. 9 and references therein), comparison of these trends with observed compositions from the HDR provides a first-order test of whether the calculated bulk-crustal composition is an appropriate parental melt composition. Although the choice of model used has a small effect on the result, all models are broadly consistent in predicting a liquidus temperature of ~1,300 °C, a crystallization sequence of olivine → olivine + plagioclase → olivine + plagioclase + clinopyroxene, and that orthopyroxene is not saturated until after a substantial interval of cotectic olivine + plagioclase + clinopyroxene crystallization. The PETROLOG models used the model of ref. 40 to divide Fe into FeO and $Fe_2O_3$ on the basis of the defined $f_{O_2}$ and the following combinations of mineral models: olivine, refs 41, 42; plagioclase, refs 27, 42; clinopyroxene, refs 27, 42; orthopyroxene, refs 43, 44. The model crystallization trends shown in Fig. 3 outline the compositional range for the melt and instantaneous cumulates with progressive fractional crystallization using these mineral models in various combinations.


34. Nilsson, K. *Oxidation State, Sulfur Speciation, and Sulfur Concentration in Basaltic Magmas: Examples from Hess Deep and the Lau Basin* 58. PhD thesis, Scripps Inst. Oceanogr., Univ. California (1993).
35. Hanna, H. D. *Geochemical Variations in Basaltic Glasses from an Incipient Rift and Upper Level Gabbros from Hess Deep, Eastern Equatorial Pacific* 56, 57. MSc thesis, Duke Univ. (2004).
36. Natland, J. H. & Dick, H. J. B. Paired melt lenses at the East Pacific Rise and the pattern of melt flow through the gabbroic layer at a fast-spreading ridge. *Lithos* **112,** 73–86 (2009).
37. Pedersen, R. B., Malpas, J. & Falloon, T. in *Proc. Ocean Drilling Program, Scientific Results* Vol. 147 (eds Mével, C., Gillis, K. N., Allan, J. F. & Meyer, P. S.) 3–19 (Ocean Drilling Program, 1996).
38. Blum, N. *Structure and Composition of Oceanic Crust and Upper Mantle Exposed in Hess Deep of the Galapagos Microplate (Equatorial East Pacific)* 232–291. PhD thesis, Univ. Karlsruhe (1991).
39. Nicolas, A., Boudier, F. & Ildefonse, B. Variable crustal thickness in the Oman ophiolite: implication for oceanic crust. *J. Geophys. Res.* **101,** 17941–17950 (1996).
40. Kress, V. C. & Carmichael, I. S. E. Stoichiometry of iron oxidation reaction in silicate melts. *Am. Min.* **73,** 1267–1274 (1988).
41. Ford, C. E., Russell, D. G., Craven, J. A. & Fisk, M. R. Olivine-liquid equilibria: temperature, pressure and composition dependence of the crystal/liquid cation partition coefficients for Mg, $Fe^{2+}$, Ca and Mn. *J. Petrol.* **24,** 256–266 (1983).
42. Danyushevsky, L. V. The effect of small amounts of $H_2O$ on crystallisation of mid-ocean ridge and backarc basin magmas. *J. Volcanol. Geotherm. Res.* **110,** 265–280 (2001).
43. Beattie, P. Olivine-melt and orthopyroxene-melt equilibria. *Contrib. Mineral. Petrol.* **115,** 103–111 (1993).
44. Ariskin, A. A., Frenkel, M. Y., Barmina, G. S. & Nielsen, R. L. COMAGMAT: a Fortran program to model magma differentiation processes. *Comput. Geosci.* **19,** 1155–1170 (1993).
45. Ferrini, V. L, *et al.* Evidence of mass failure in the Hess Deep rift from multi-resolutional bathymetry data. *Mar. Geol.* **339,** 13–21 (2013).




**Extended Data Figure 1 | Summary lithostratigraphic columns of the gabbroic rocks recovered at IODP holes.** **a**, U1415I; **b**, U1415J; **c**, U1415P. Columns show recovery, lithological units, major rock types, dip of magmatic foliations and well-constrained magnetic remanence inclination values (mean and 1 s.d. listed). Lithological units were identified on the basis of similarities in rock types, magmatic textures and foliations. Palaeomagnetic remanence directions and the dip of the magmatic foliations and layers (not shown) for units II and III in holes U1415J and U1415P are most easily interpreted as blocks that probably formed by slumping and were rotated relative to each other. Ghost cores (G cores) are intervals drilled during hole cleaning operations. In **a**, Unit II refers to the Unit II layer gabbro series. **d**, Map showing the relative locations of holes U1415I, U1415J and U1415P; microbathymetry from ref. 45.




**Extended Data Figure 2 | Core images showing examples of simple, centimetre-scale modal layering and a moderate-to-strong magmatic foliation.** **a**, 345-U1415J-5R-2, piece 1, 2.0–17.5 cm; **b**, 345-U1415J-8R-2, piece 9, 105.5–121.0 cm.

**Extended Data Table 1 | Bulk compositions of crustal sections used to calculate the bulk-crustal composition and the bulk composition of the HDR crust and plutonic section**

| | | $SiO_2$ | $TiO_2$ | $Al_2O_3$ | $FeO^T$ | MnO | MgO | CaO | $Na_2O$ | $K_2O$ | Total | Mg#* | n |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Upper** | mean | 50.64 | 2.12 | 13.75 | 11.79 | 0.21 | 6.83 | 10.57 | 2.77 | 0.13 | 99.87 | 50.2 | 175 |
| **crust** | s.e.m | 0.05 | 0.03 | 0.06 | 0.12 | 0.00 | 0.05 | 0.07 | 0.02 | 0.01 | 0.05 | 0.00 | |
| **Shallow** | mean | 51.22 | 1.15 | 15.05 | 9.09 | 0.17 | 8.09 | 12.02 | 2.41 | 0.05 | 99.32 | 61.8 | 139 |
| **gabbros** | s.e.m. | 0.15 | 0.09 | 0.13 | 0.22 | 0.00 | 0.12 | 0.13 | 0.04 | 0.00 | 0.07 | 0.01 | |
| **Lower gabbros**† | | | | | | | | | | | | | |
| Site U1415 | mean | 47.22 | 0.16 | 19.50 | 4.95 | 0.09 | 13.40 | 13.39 | 1.46 | 0.10 | 100.78 | 82.6 | 28 |
| Layered series | s.e.m. | 0.33 | 0.01 | 0.40 | 0.20 | 0.00 | 0.66 | 0.37 | 0.07 | 0.01 | 0.18 | 0.01 | |
| Site U1415 | mean | 44.56 | 0.08 | 18.23 | 5.45 | 0.09 | 19.27 | 11.58 | 0.85 | 0.13 | 100.83 | 85.7 | 15 |
| Troctolite series | s.e.m. | 0.47 | 0.02 | 1.31 | 0.41 | 0.01 | 2.09 | 0.69 | 0.19 | 0.02 | 0.32 | 0.01 | |
| **Bulk crust** | | 48.32 | 0.83 | 16.71 | 7.66 | 0.14 | 12.12 | 11.95 | 1.83 | 0.11 | 99.73 | 73.8 | |
| **composition** | s.d. | 0.51 | 0.10 | 0.50 | 0.37 | 0.01 | 1.04 | 0.31 | 0.13 | 0.01 | 0.80 | 2.2 | |
| **Plutonic section** | | 47.65 | 0.46 | 17.57 | 6.47 | 0.12 | 13.66 | 12.35 | 1.56 | 0.10 | 100.30 | 79.0 | |
| **bulk composition** | s.d. | 0.63 | 0.08 | 0.60 | 0.36 | 0.01 | 1.28 | 0.39 | 0.15 | 0.01 | 1.03 | 0.02 | |

$FeO^T$, all Fe as FeO. s.d., standard deviation; s.e.m., standard error on the mean.
* Mg# = 100Mg/(Mg + Fe).
† Includes data for IODP holes U1415I, U1415J and U1415P.

LETTER

doi:10.1038/nature12772

# The effects of genetic variation on gene expression dynamics during development

Mirko Francesconi[1,2] & Ben Lehner[1,2,3]

**The development of a multicellular organism and physiological responses require massive coordinated changes in gene expression across several cell and tissue types[1–3]. Polymorphic regions of the genome that influence gene expression levels have been identified by expression quantitative trait locus (eQTL) mapping in many species[4–6], including loci that have cell-dependent[7,8], tissue-dependent[9] and age-dependent[10] effects. However, there has been no comprehensive characterization of how polymorphisms influence the complex dynamic patterns of gene expression that occur during development and in physiology. Here we describe an efficient experimental design to infer gene expression dynamics from single expression profiles in different genotypes, and apply it to characterize the effect of local (*cis*) and distant (*trans*) genetic variation on gene expression at high temporal resolution throughout a 12-hour period of the development of *Caenorhabditis elegans*. Taking dynamic variation into account identifies >50% more *cis*-eQTLs, including more than 900 that alter the dynamics of expression during this period. Local sequence polymorphisms extensively affect the timing, rate, magnitude and shape of expression changes. Indeed, many local sequence variants both increase and decrease gene expression, depending on the time-point profiled. Expression dynamics during this 12-hour period are also influenced extensively *in trans* by distal loci. In particular, several *trans* loci influence genes with quite diverse dynamic expression patterns, but they do so primarily during a common time interval. *Trans* loci can therefore act as modifiers of expression during a particular period of development. This study provides the first characterization, to our knowledge, of the effect of local and distant genetic variation on the dynamics of gene expression throughout an extensive time period. Moreover, the approach developed here should facilitate the genetic dissection of other dynamic processes, including potentially development, physiology and disease progression in humans.**

To characterize the effects of genetic variation on gene expression dynamics, we developed an approach that reconstructs dynamic changes in gene expression from the static genome-wide expression profiles of several individual genotypes. For both technical and biological reasons, individual samples collected for gene expression profiling will normally not be perfectly synchronized[10,11]. Rather than considering this a disadvantage, our method exploits this variation to infer the dynamics of gene expression at high temporal resolution (Fig. 1). Our approach has two stages: first, we rank the samples according to their relative developmental progression inferred from their gene expression profiles (Fig. 1b). Second, we use multivariate dynamic time warping to derive the precise physiological time point of each sample, relative to a reference time series (Fig. 1c). The dependence of these dynamics on genetic variation in a population is then evaluated (see Methods).

As a model process, we considered the late larval and early adult development of the nematode *C. elegans*, which represent a series of complex, multi-tissue developmental transitions that occur over many hours. Genome-wide gene expression measurements have been made on more than 200 recombinant inbred advanced intercross lines (RIAILs) derived from a cross between two divergent strains of *C. elegans*—the standard Bristol (N2) laboratory strain, and an isolate from Hawaii (CB4856), with each RIAIL genotyped at 1,455 markers[12] (Fig. 1a). The genomes of the parental strains differ at a similar level of polymorphism as two humans[12]. As a reference gene expression time series, we used a data set consisting of 12 genome-wide expression profiles made every 3 h and in triplicate from the mid-L3 larval stage to reproductive maturity in the Bristol strain[3].

To determine the relative physiological age of each sample, we used a multivariate technique, canonical correlation analysis, to search for mutually uncorrelated linear combinations of gene expression (canonical variates) in the reference time series[3] that best explain (that is, are maximally correlated with) gene expression trends in the RIAIL data set. The first two canonical variates are highly enriched for oogenesis and spermatogenesis genes, respectively (Extended Data Fig. 1a), and the trajectory of the reference samples (Extended Data Fig. 1b) reveals that these germline expression signatures can be used to sort the RIAIL samples by their physiological age (Fig. 1b).

After sorting the RIAILs, we next inferred their absolute physiological ages using a multivariate dynamic time-warping algorithm[13] that searches for the optimal match between the profiles of the first six RIAIL canonical variates and those of the reference time series (Fig. 1c, d and Extended Data Fig. 1f–h). This reveals that the physiological ages of the 206 RIAIL samples are non-uniformly distributed over a 12-h time interval centred on the late L4 developmental stage (Fig. 1d, inset).

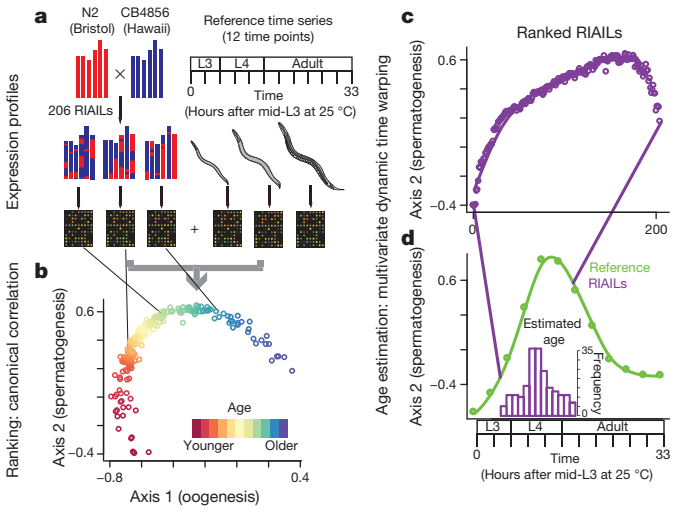



**Figure 1 | Reconstructing dynamics from static expression profiles.** **a**, Overview of the experimental design. **b**, Projection of the RIAILs on the first two canonical variates. **c**, Expression of the second canonical variate, representing spermatogenesis, in each RIAIL ranked according to age. **d**, Dynamic time warping onto the reference time series. Inset, distribution of the inferred physiological ages of the RIAILs.

[1]EMBL-CRG Systems Biology Unit, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain. [2]Universitat Pompeu Fabra (UPF), Dr. Aiguader 88, 08003 Barcelona, Spain. [3]Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain.


**Figure 2 | High-resolution view of gene expression during 12 h of *C. elegans* development.** Hierarchically clustered expression of 15,855 genes in the 206 RIAILs ordered by their inferred physiological age. Tissue-specific expression signatures enriched in each cluster are listed (tissues accounting for at least 30% of annotated genes, and Fisher's exact test $P < 0.0134$ that corresponds to FDR = 0.1).

The 206 ordered RIAIL samples provide an unprecedentedly high temporal resolution picture of gene expression dynamics during this period. In total, 14,926 out of 15,855 expressed genes (94%) change in relative expression during this 12-h time interval (false discovery rate (FDR) = 0.1, Supplementary Table 1), with hierarchical clustering revealing a range of different dynamic patterns of expression deriving from genes expressed in different tissues within the animal (Fig. 2). For example, multiple time-shifted waves of gene expression are observed from the hypodermal cells, reflecting the molting cycle, and even spermatogenesis-related expression falls into two temporal phases (Fig. 2). Intestine and neuronal expression can also be identified from the high-resolution temporal dynamics (Fig. 2).

To analyse how genetic variation influences expression dynamics, we performed eQTL mapping[6]. We first focused on local regulatory variation. Without considering time, regression on local genetic markers identified 2,958 genes with expression levels significantly affected by local sequence variation (*cis*-eQTLs, FDR = 0.1). However, adding time as a covariate yielded 4,246 *cis*-eQTLs, 44% more than when time is ignored. Moreover, 905 genes were found to have expression that is significantly better explained by a model that includes an interaction between local genetic variation and time than by an additive model (at FDR = 0.1). Indeed, for 300 genes, a significant *cis*-eQTL could only be detected when considering an interaction with time, resulting in a total of 4,546 *cis*-eQTLs (54% more than when time is ignored and methods that regress out hidden confounding factors[14,15] have not been applied; Extended Data Fig. 1i).

We classified the 905 'dynamic' *cis*-eQTLs into five non-mutually exclusive classes (Fig. 3 and Extended Data Fig. 1j). A total of 174 *cis*-eQTLs affected the magnitude of an expression response, with the Hawaii allele, for example, showing stronger induction than the Bristol allele owing to local sequence variation (Fig. 3A, a), or with an increased amplitude of a periodic expression change (Fig. 3A, b). By contrast, for 275 genes it was the timing of expression that was altered, with expression often showing faster dynamics for one of the two alleles (Fig. 3B). For 148 genes with monotonic dynamics, it was the rate of expression change that was altered, with faster or slower induction of expression depending on whether the allele derived from the Bristol or Hawaii parental strain (Fig. 3C). Finally, in 245 cases, the situation was more complex, with the Bristol and Hawaii alleles showing a different shape of temporal expression dynamics during this 12-h period (Fig. 3D).

Notably, for 230 genes, the effect of the *cis*-eQTL changed sign (that is, from positive to negative) one or more times even during this limited 12-h time window, meaning that the Hawaii allele would have been reported as increasing or decreasing expression depending on the time-point considered (Fig. 3E), similar to what has been observed when comparing between cell types[16,17]. Thus, even during a relatively short time window, local genetic variation between two individuals can have diverse and complex effects on gene expression.

As previously reported[12], genes with significant *cis*-eQTLs have a higher level of local sequence polymorphism and they are enriched on the more polymorphic arms of *C. elegans* autosomes (Extended Data Fig. 1k). This is also true for genes with dynamic *cis*-eQTLs (Extended Data Fig. 1k). However, genes with different classes of dynamic *cis*-eQTLs differ in the density of polymorphisms in different genic regions (see d in Fig. 3, and Extended Data Fig. 1l). In particular, genes with a significant *cis*-eQTL altering the rate or the shape of expression have a higher level of polymorphism in their 5′ untranslated regions (UTRs) (Fig. 3C, D), suggesting that causal variants might be enriched in the 5′ UTRs and that changes in the rate of expression may be mediated by changes in post-transcriptional regulation.

We next analysed the effect of distant (*trans*) genetic variation on the dynamics of gene expression using a multivariate approach, random forest regression[18], including physiological age as a covariate (see Methods). In total, we identified 3,164 significant *trans*-eQTLs (FDR = 0.1; Fig. 4A), indicating that the expression of a large number of genes is influenced by non-local polymorphisms between the Bristol and Hawaii genomes.

We found that 773 of the genes with *trans*-eQTLs were influenced by variation in 10 *trans* hotspots (Fig. 4A, B), including three previously identified loci[12]. We clustered the expression of the genes influenced by each of these loci and analysed their tissue-specificity (Fig. 4 and Extended Data Fig. 2). Two hotspots target hypodermally expressed genes (Fig. 4C, b and c, 37% of annotated genes, $P = 1.55 \times 10^{-4}$ and 39%, $P = 1 \times 10^{-3}$, respectively), one is specific for sperm-expressed genes (Fig. 4C, d, 80% of annotated genes, $P = 1.39 \times 10^{-5}$) and another for intestine-expressed genes (Extended Data Fig. 2d). The hotspot containing a known loss-of-function polymorphism in the neuropeptide receptor *npr-1* in the Bristol strain targets neuronal- (28% of annotated genes, $P = 1.06 \times 10^{-5}$) and body-wall-muscle-expressed genes (28% of annotated genes, $P = 6.28 \times 10^{-7}$) (Extended Data Fig. 2a). Thus, one set of *trans*-eQTL hotspots probably acts by influencing the development of individual cell types within the animal.

Notably, however, although a hotspot may target a single tissue, the target genes can have diverse temporal dynamics. For example, the hotspot in the middle of chromosome I alters the expression of genes expressed in the hypodermis with three different dynamic patterns of expression (Fig. 4C, b). This promiscuity is also apparent for hotspots that influence expression in several tissues. For example, the hotspot on the left arm of chromosome V alters the expression of genes with five different dynamic patterns of expression, including genes that increase, decrease or oscillate in expression during this 12-h window (Fig. 4C, a). Only one of the top ten hotspots primarily influences the expression of genes with a single dynamic pattern (Fig. 4C, d).

What connects the expression of genes expressed in different tissues and with different time trends that are influenced by a common *trans* hotspot? In some cases, at least, it seems that the hotspot region influences expression levels during a restricted time window. For example, the hotspot on chromosome V primarily affects expression at the start of the

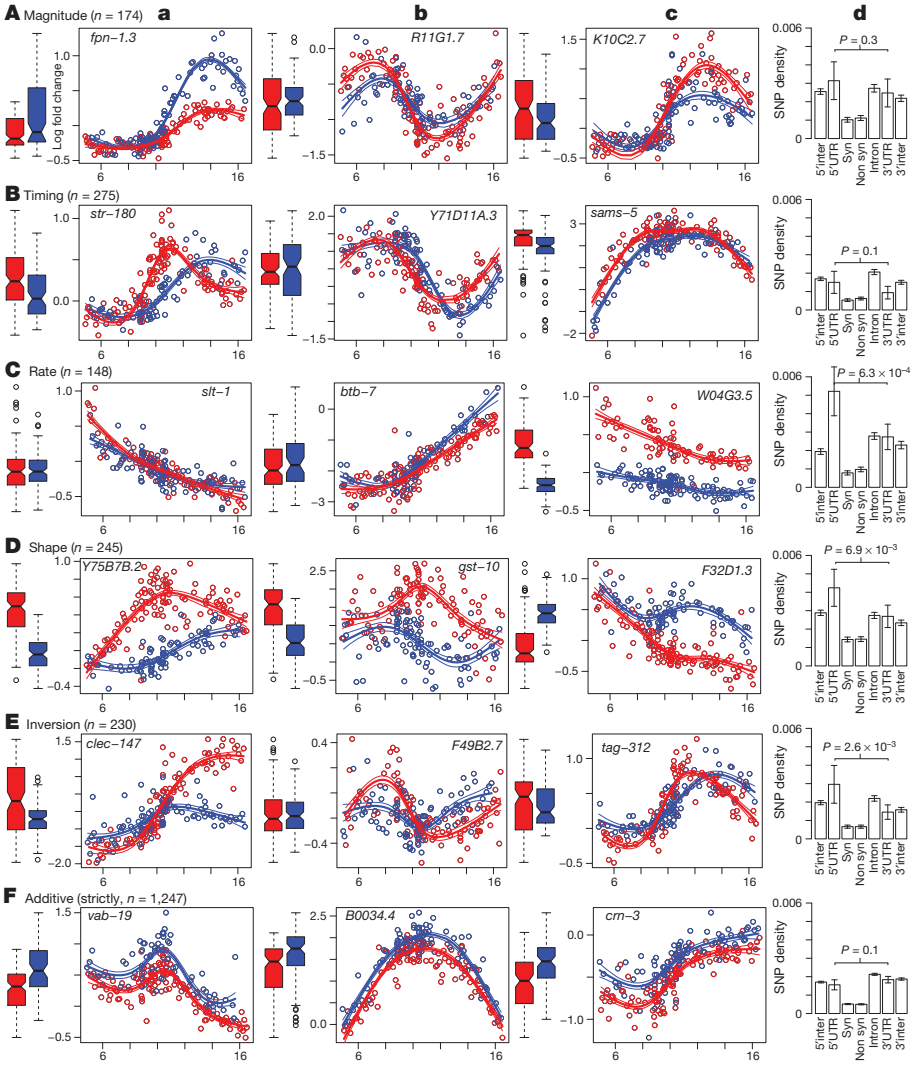



**Figure 3 | A total of 905 *cis*-eQTLs affect the dynamics of gene expression.** **A**–**F**, Example *cis*-eQTLs that affect the magnitude (**A**), timing (**B**), rate (**C**), and shape (**D**) of expression changes, or that have opposite effects depending on the time point (**D**) and strictly additive effects (**F**). Bristol alleles are red, Hawaii alleles are blue, box plots (left) include all time points. **A**–**F**, **d**, Density of polymorphisms in different regions of the genes in each class. inter, intergenic region; SNP, single nucleotide polymorphism; syn, synonymous sites. *P* values calculated using Fisher's exact test. Error bars represent 95% binomial confidence intervals.

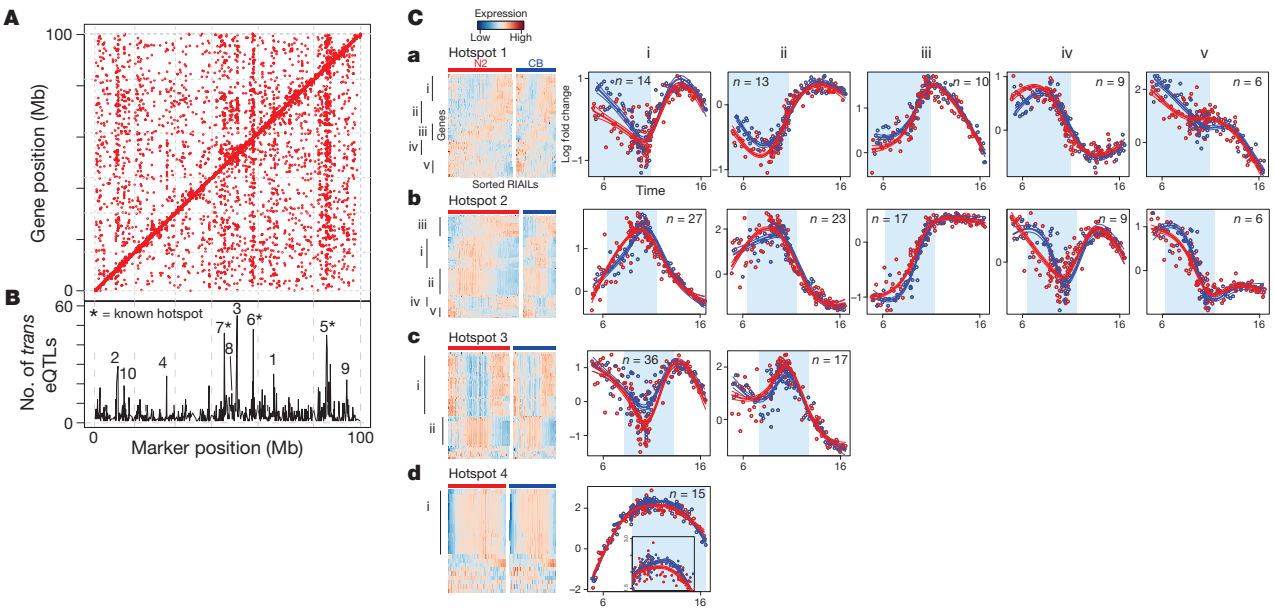



**Figure 4 | *trans*-eQTLs.** **A**, Gene expression (*y* axis) significantly influenced by marker genotypes (*x* axis). **B**, The number of genes influenced *in trans* by each marker identify *trans*-eQTL hotspots, including three previously identified regions (asterisks). **C**, Left, clustered heat maps of the time-ordered expression of genes regulated by *trans*-eQTL hotspots; right, expression of the means of each cluster (labelled i–v, as in the heat maps) when the hotspot carries the Bristol (red) or Hawaii (blue) allele. Three hotspots affecting expression during a limited time interval (pale blue background) are shown, together with a hotspot specific for sperm expression. See also Extended Data Fig. 2.

12-h time interval, altering the expression of genes with quite diverse expression trends during this period (Fig. 4C, a). By contrast, the hotspot on chromosome I has the largest effects on expression slightly later in development (Fig. 4C, b), and the hotspot in the middle chromosome IV affects expression a few hours later in development (Fig. 4C, c). Thus, sequence variation in these regions alters the expression of multiple genes including those with different expression trends, but during a specific period of development (Fig. 4).

In summary, we have performed a comprehensive analysis of how local and distant genetic variation influences the dynamics of gene expression. Considering the dynamics of gene expression identified more eQTLs, but, more importantly, revealed that local sequence variation can have diverse and complex effects, altering the timing, rate, magnitude and shape of expression changes. Natural sequence variation also extensively alters gene expression dynamics *in trans*, with hotspot regions altering expression in individual tissues or across multiple tissues, often affecting genes expressed with quite diverse dynamic trends, and, in some cases, primarily influencing expression during a particular temporal period of development.

The approach that we have used here represents an efficient experimental design, exploiting technical, biological or deliberate variation in timing to infer dynamics from a single static profile of expression in each genotype. We envisage, therefore, that it could be quite widely applied for the genetic analysis of other complex dynamic systems, including the potential to characterize the effect of natural variation on development, physiology and disease progression in humans and other economically important species.

## METHODS SUMMARY

The physiological age of each RIAIL sample was estimated using a canonical correlation analysis of the RIAIL expression profiles and a reference gene expression time course followed by multivariate dynamic time warping on the pairs of canonical variates. Single nucleotide polymorphisms and insertions and deletions between the CB4856 and N2 genomes were obtained from the million mutation project[19]. *cis*-eQTLs were detected using a linear model that includes RIAILs age as a covariate, while *trans*-eQTLs were detected using random forest regression including age as a covariate. Tissue specific gene expression was defined using published data[20].

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 27 June; accepted 14 October 2013.**
**Published online 24 November 2013.**

1. Arbeitman, M. N. *et al.* Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297,** 2270–2275 (2002).
2. White, K. P., Rifkin, S. A., Hurban, P. & Hogness, D. S. Microarray analysis of *Drosophila* development during metamorphosis. *Science* **286,** 2179–2184 (1999).
3. Reinke, V., Gil, I. S., Ward, S. & Kazmer, K. Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. *Development* **131,** 311–323 (2004).
4. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296,** 752–755 (2002).
5. Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422,** 297–302 (2003).
6. Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. *Nature Rev. Genet.* **7,** 862–872 (2006).
7. Gerrits, A. *et al.* Expression quantitative trait loci are highly sensitive to cellular differentiation state. *PLoS Genet.* **5,** e1000692 (2009).
8. Ackermann, M., Sikora-Wohlfeld, W. & Beyer, A. Impact of natural genetic variation on gene expression dynamics. *PLoS Genet.* **9,** e1003514 (2013).
9. Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6,** e107 (2008).
10. Viñuela, A., Snoek, L. B., Riksen, J. A. G. & Kammenga, J. E. Genome-wide gene expression regulation as a function of genotype and age in *C. elegans*. *Genome Res.* **20,** 929–937 (2010).
11. Gutteling, E. W., Riksen, J. A. G., Bakker, J. & Kammenga, J. E. Mapping phenotypic plasticity and genotype-environment interactions affecting life-history traits in *Caenorhabditis elegans*. *Heredity* **98,** 28–37 (2007).
12. Rockman, M. V., Skrovanek, S. S. & Kruglyak, L. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science* **330,** 372–376 (2010).
13. Tormene, P., Giorgino, T., Quaglini, S. & Stefanelli, M. Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation. *Artif. Intell. Med.* **45,** 11–34 (2009).
14. Stegle, O., Parts, L., Durbin, R. & Winn, J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLOS Comput. Biol.* **6,** e1000770 (2010)
15. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3,** 1724–1735 (2007).
16. Fu, J. *et al.* Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. *PLoS Genet.* **8,** e1002431 (2012).
17. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nature Genet.* **44,** 502–510 (2012).
18. Michaelson, J. J., Alberts, R., Schughart, K. & Beyer, A. Data-driven assessment of eQTL mapping methods. *BMC Genomics* **11,** 502 (2010).
19. Thompson, O. *et al.* The million mutation project: A new approach to genetics in *Caenorhabditis elegans*. *Genome Res.* **23,** 1749–1762 (2013).
20. Spencer, W. C. *et al.* A spatial and temporal map of *C. elegans* gene expression. *Genome Res.* **21,** 325–341 (2011).

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

**Acknowledgements** This work was funded by ERASysBio+ ERANET 'GRAPPLE', an ERC Starting Grant, MICINN BFU2008-00365 and BFU2011- 26206, AGAUR, the EMBO Young Investigator Program, EU Framework 7 project 277899 4DCellFate, and by the EMBL-CRG Systems Biology Program. We thank J. Kammenga, B. Snoek, M. Sterken and F. Supek for discussions and/or comments on the manuscript.

**Author Contributions** M.F. performed all analyses. M.F. and B.L. designed analyses and 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 the paper. Correspondence and requests for materials should be addressed to B.L. (email: ben.lehner@crg.eu).

## METHODS

**Recombinant inbred lines.** The RIAILs derive from a cross between the N2 (Bristol) and CB4856 (Hawaii) strains of *C. elegans*, followed by ten generations of random pair mating and ten generations of selfing[21]. Each RIAIL line was previously genotyped at 1,455 SNP markers[22].

**Gene expression profiles.** The gene expression data set analysed here consists of microarray expression profiles for 208 RIAILs[12]. Worms were synchronized by hypochlorite treatment[23] and reared at 20 °C. Animals were collected as soon as the population started to lay embryos. RNA was extracted using the trizol method and hybridized to Agilent 4×44k microarrays using standard protocols as described previously[24]. Gene expression data, batch and dye information were obtained from the Gene Expression Omnibus (GEO) database. Microarray probe annotations were obtained from Wormbase release WS220.

**Data pre-processing.** Microarray probes matching multiple loci or pseudogenes according to WS220 were discarded. In addition, probes were discarded if they contained SNPs or insertions and deletions (indels) in CB4856 according to the sequencing data from the million mutation project[19], as were probes with more than 10 saturated values (value higher than 64,000) across the 208 RIAILs, or those not detected as expressed in at least 8 samples. The threshold for classifying a gene as expressed was defined using model cluster analysis (function Mclust in the R mclust library) on single channel log intensities. The lowest intensity cluster had a mean of ~3.65 and a standard deviation (s.d.) of ~0.03, so the threshold for expression was defined at 3.6 + 5 s.d. = 3.8 log intensity. No background subtraction was applied and intensities within arrays were normalized using LOESS (using the function normalizeWithinArrays in the limma library in R). Next, the reference intensities where normalized among arrays using quantile normalization, ensuring that the reference channel has the same empirical distribution across all the arrays (function normalizeBetweenArrays, limma library, R methods Gquantile and Rquantile). Log intensities of different probes for the same genes were averaged. Principle component analysis (PCA) on the uncentred data matrix highlighted two outlier samples that were excluded from all further analyses (Extended Data Fig. 1c).

**Developmental time course.** The developmental time course gene expression data set consists of gene expression profiles of pools of synchronized worms reared at 25 °C collected every 3 h for a total of 12 time points in three replicates, encompassing development from the middle of the L3 larval state to young adults[3]. Data were obtained from the SPELL database in $\log_2$ of fold change format, each of the 12 data points being the average across the three replicas.

**SNPs and indels.** SNPs and indels between the CB4856 and N2 genomes were obtained from the million mutation project[19]. In total, there are 173,898 SNPs and 5,515 indels between the two strains.

**Ranking samples by developmental progression: canonical correlation analysis.** Canonical correlation analysis was used to compare the gene expression profiles of the RIAILs to the expression profiles of the reference development time series. This analysis searches for uncorrelated trends in gene expression (called canonical variates) in one data set that best explain (that is, are maximally correlated with) gene expression trends in the other one. The cc function in the CCA library in R was used. RIAILs were sorted according to the angle formed by their loadings onto the first and second canonical variates transformed into a polar coordinate system. These projections correspond to the oogenesis and spermatogenesis gene expression trends, respectively. PCA performed on the RIAIL samples further supported this conclusion that germline development is a major source of expression variance (Extended Data Fig. 1d, e).

**Estimating the developmental stage of each sample: multivariate dynamic time warping.** After sorting the RIAILs according to their relative age, the absolute physiological age of each sampled was estimated by applying multivariate dynamic time warping[13] to the canonical variates extracted from the canonical correlation analysis. First, a smoothing spline was fitted to both the sorted RIAILs and to the reference time series projections on each of the first six canonical variates using the gam function in the gam R library and a smoothing parameter of 0.9 for the RIAILs canonical variates and and 0.25 for the reference ones. The difference is due to the fact that the RIAIL data are more variable and they need more smoothing to avoid over-fitting while the reference data points are already an average of three replicates. Next, the reference time series was interpolated with cubic splines using 700 points using spline function in R. Finally, multivariate dynamic time warping was applied using the projection of the RIAIL expression splines onto the first six gene expression trends extracted by the canonical correlation analysis and using the developmental time series as reference. The dtw function from the dtw library in R was used with the open start, open end and asymmetric step options.

***cis*-eQTLs.** A focused single-marker approach[25] was applied to detect *cis*-eQTLs. Standard linear regression was used to test the significance of the closest marker on gene expression. The significance of a model including only the marker effect was tested against a null model using an *F*-test. *P* values from the *F*-test were adjusted for multiple testing, controlling the FDR using the Benjamini–Hockberg method. This yielded 2,958 genes significant at FDR = 0.1. Next, a time term was included in the model as a covariate to model time-dependent gene expression changes. In this term, nonlinear expression dynamics were modelled using natural cubic splines basis functions (function ns of the splines library in R) using 4 degrees of freedom for all genes, chosen by evaluating the quality of the fit on the first 6 canonical variates of the canonical correlation analysis[26]. An *F*-test was used to test whether a model that included both the time term and the marker term was significantly better than a model that only includes the time term. This test yielded 4,246 significant eQTLs at FDR = 0.1. To define eQTLs including a significant interaction between genotype and time, the *F*-test was used to test whether a linear model that includes the age by maker interaction is significantly better than a model that only includes time and marker additive effects. This test yielded 905 eQTLs with significant time by marker interaction at FDR = 0.1. These results are robust to variation in time estimation (Extended Data Table 1a, b). eQTLs were classified as strictly additive if the raw *P* value for time by marker interaction was higher than 0.5 ($n = 1,247$). The 905 eQTLs with a significant time by marker interaction were further divided in 5 non-mutually exclusive classes by inspection of the time trends. *cis*-eQTLs that affected the magnitude of an expression response or with an increased amplitude of a periodic expression change were included in 'magnitude' class ($n = 174$). eQTLs that affect the timing of expression, for example, where the gene expression of the two alleles peaks at different times, were assigned to the 'timing' class. For genes with a monotonic expression dynamic, if the eQTLs show a different rate of expression change (that is, a faster or slower induction) they were classified as 'rate' ($n = 148$). eQTLs where Bristol and Hawaii alleles show very different temporal expression dynamics that could not be obtained by a trivial transformation were assigned to the 'shape' class ($n = 245$). We also considered cases where eQTLs are characterized by a change of sign of the effect during the time window; that is, when one allele shows higher expression at one time points but weaker induction at another time point. These eQTLs were labelled as 'inversions' ($n = 230$). Forty eQTLs where one allele has just one outlier point at the extreme of the time window were considered as false positives and were not further considered, together with 43 more eQTLs where the difference between the two alleles is very small and difficult to interpret and so also probably represent false positives. Classifications of all the *cis*-eQTLs are given in Supplementary Table 2.

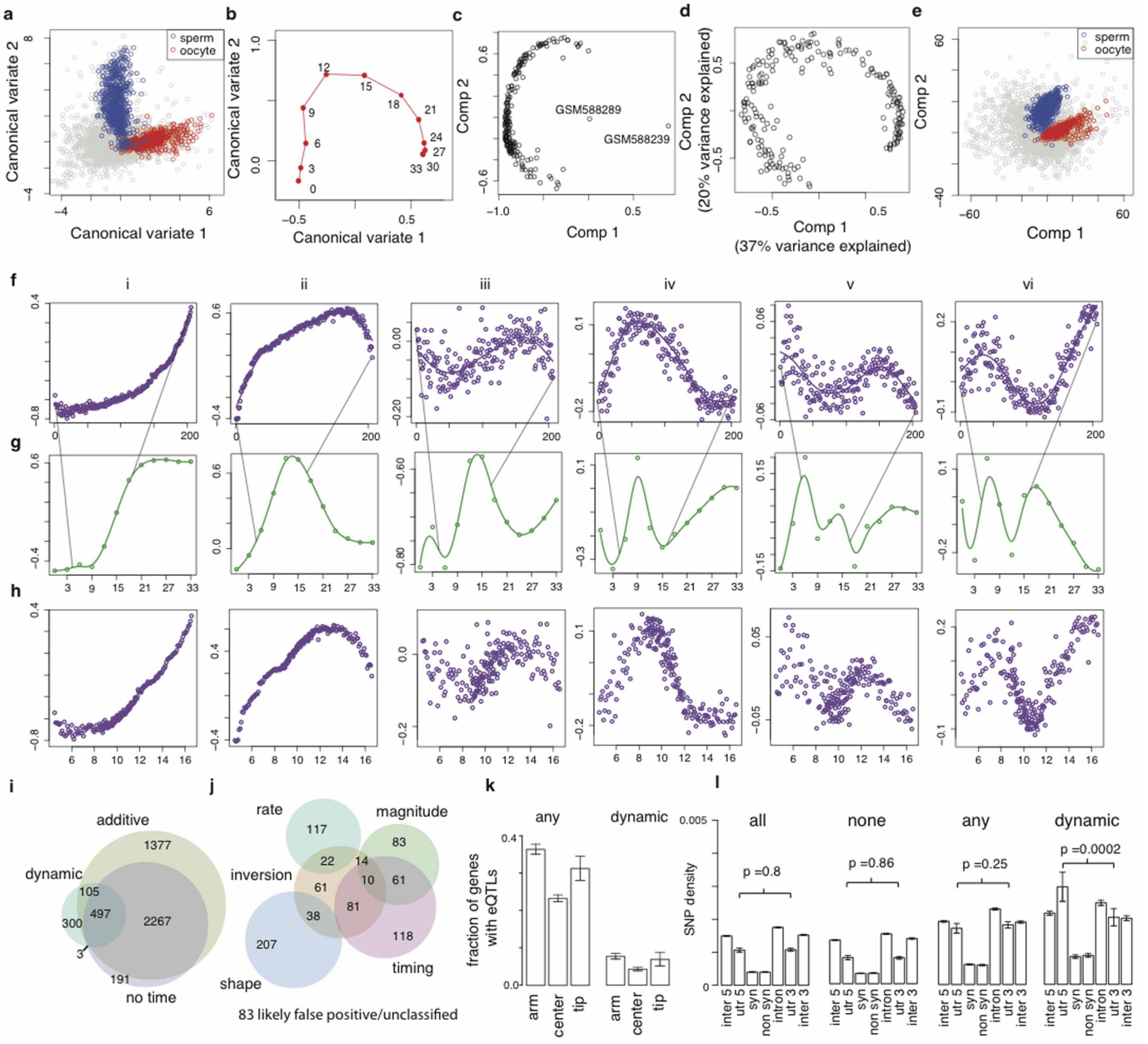
***trans*-eQTLs.** To identify *trans*-eQTLs, a genome-wide multivariate analysis was performed using random forest regression. Random forest is a powerful multivariate machine learning technique based on ensembles of regression trees. This algorithm can model nonlinear trends with age, can include a large number of correlated predictors, is robust to over-fitting, and provides a variable importance measure that implicitly takes into account interactions among predictors[18]. The random forest implementation provided in the randomForest library in R was used. After filtering out completely correlated markers that can degrade the performance of the random forest algorithm, 884 useful markers remain; that is, those that differ in at least one RIAIL. These markers, estimated age, batch and dye were included as predictors in the random forest models. The number of predictors to try at each bootstrap sample of the data (the 'mtry' parameter) was set to the maximum possible; that is, equal to the total number of predictors. This allows age and batch to always be included in the trees and it was empirically found to increase the performance compared to the default value of the mtry parameter. A forest of 1,000 trees was fitted on each gene and the mean decrease in prediction accuracy across the trees normalized by the standard deviation was used as a measure of variable importance. The random forest analysis was also run on 10 permuted data sets, each one obtained by permuting the sample labels to retain the correlation structure in gene expression. The null distribution of variable importance of each predictor from these randomizations was used to calculate empirical *P* values for importance. The matrix of empirical *P* values was then corrected for multiple testing controlling the FDR (function p.adjust in R, method = 'FDR'). This yields 5,803 eQTLs at 10% FDR. Extremely close eQTLs identified by random forests are likely to be linked to the same underlying causal variants. For this reason and to be conservative, hierarchical clustering of the genotype correlation matrix was used to group together highly correlated markers into 150 genomic intervals. For each gene, eQTLs mapping to the same interval were merged, keeping the eQTL with the highest variable importance score. This filtering leaves 4,976 eQTLs. The same 150 genomic intervals were used to define local and distant eQTLs. An eQTL that lies in the same interval as the affected gene plus or minus 500 kilobases was classified as *cis*. Using this definition leaves 3,164 *trans* eQTLs. *Trans* hotspots were defined using a Poisson test. According to this test, markers associated with nine or more eQTLs have a significantly higher number of eQTLs than expected by chance at FDR = 0.1. This yields 84 markers that are associated to 1,270 genes. Adjacent markers were merged into the same hotspot if they have a Pearson correlation coefficient of their genotypes across the RIAILs greater than 0.95. This leaves 61 *trans* hotspots. Two small hotspots map to chromosome I close

to the *zeel-1/peel-1* locus responsible for the genetic incompatibility between Bristol and Hawaii[27]. The two hotspots in this locus are supported by an extremely low number of Hawaii alleles and were not included in subsequent analyses. Genes in the *trans* hotspots were clustered using hierarchical clustering to highlight how changes in the gene expression dynamics associated to the hotspot. Close hotspots that show similar changes in expression dynamics are likely to share the same causal variants or combinations of causal variants and they were further merged into one. This is the case for two close hotspots on the left arm of chromosome I, for two close hotspots on the left arm of chromosome IV, two on the right arm of chromosome IV, three on the left arm of chromosome V, and for three hotspots on the left arm of chromosome X close to the *npr-1*. This leaves a total of 51 hotspots. The genes targeted by the top 10 *trans*-eQTL hotspots are listed in Supplementary Table 3. We also performed a single marker analysis (Extended Data Fig. 3) taking into account time by using a robust linear model fitting approach (function lmrob from robustbase package in R). The full model that includes time, marker and the interaction between marker and time was compared to a model that only includes time, using a likelihood ratio test (R function anova.lmrob with the test = 'Deviance' option). The analysis was repeated on ten permuted data sets and the number of genes showing at least one significant linkage in the permuted data sets was used to determine the FDR[4,28].

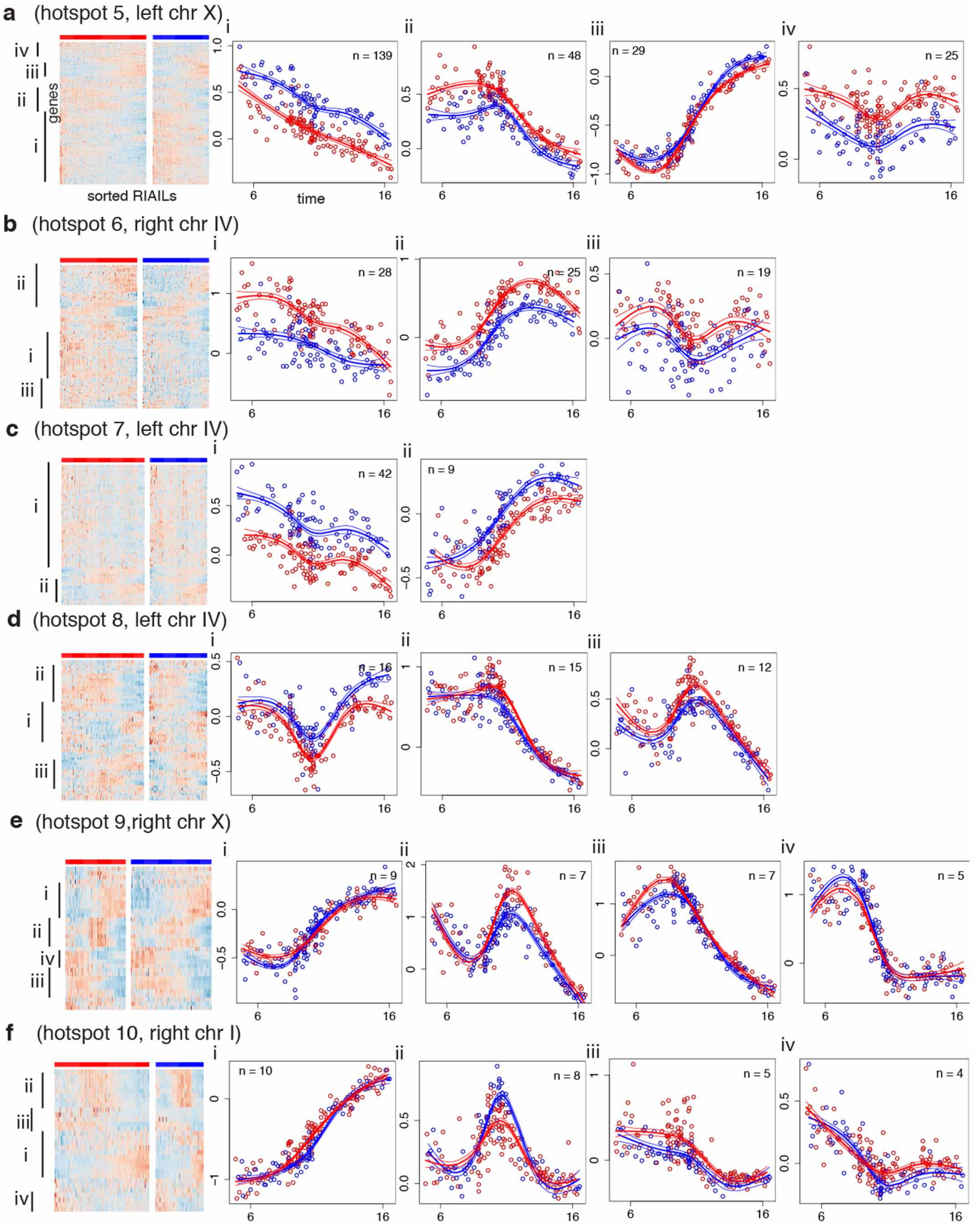
**Tissue specificity of gene expression.** Genes highly expressed in individuals tissues were defined using cell sorting or polyA-binding protein (PAB-1) immunoprecipitation, as reported previously[20]. Embryonic cells expressing tissue-specific fluorescent reporter constructs were sorted by FACS analysis to define embryo specific gene expression from the germline precursors, all neurons, BAG neurons, AVA neurons, AVE neurons, A-class motorneurons, dopaminergic and GABAergic neurons, body wall muscle, coelomocytes, hypodermis, intestine and pharyngeal muscle. A messenger-RNA-tagging strategy was used to isolate RNA from specific larval and adult cells: an epitope-tagged PAB-1 is expressed under control of cell-specific promoters and PAB-1–RNA complexes are immunoprecipitated. Larval-specific gene expression profiles from all neurons, A-class motor neurons, dopaminergic and glutamatergic neurons, PVD and OLL neurons, body wall muscle, coelomocytes, hypodermis, intestine, pharyngeal muscle, excretory cells and sheath cells. In all cases, extracted RNA was amplified and hybridized to Affymetrix tiling arrays[20]. Genes were defined to be tissue enriched if they show a twofold enrichment versus a reference sample at FDR $< 0.05$ (ref. 20). Tissue specificity of the *trans* bands and expression clusters was determined using Fisher's exact test, considering the genes that are annotated in at least one tissue. Among the 3,687 tissue-specific genes, 236 have a significant dynamic eQTL, which is moderately enriched compared to other genes (odds ratio $= 1.17$, $P = 0.04$). For each of these genes, we calculated the correlation coefficient between the spline fitted on the expression data for each of the two alleles and the centroids of the expression clusters in Fig. 2. We found that in 60% of cases both of the alleles belong to (that is, are maximally correlated with) the same cluster, and in another 15% of cases the two alleles belong to two distinct clusters but with the same tissue specificity. This suggests that, for at least 75% of these genes, the eQTLs cause changes in gene expression dynamics without changes in gene tissue specificity.

**Statistical analyses.** All statistical analyses were performed in R.

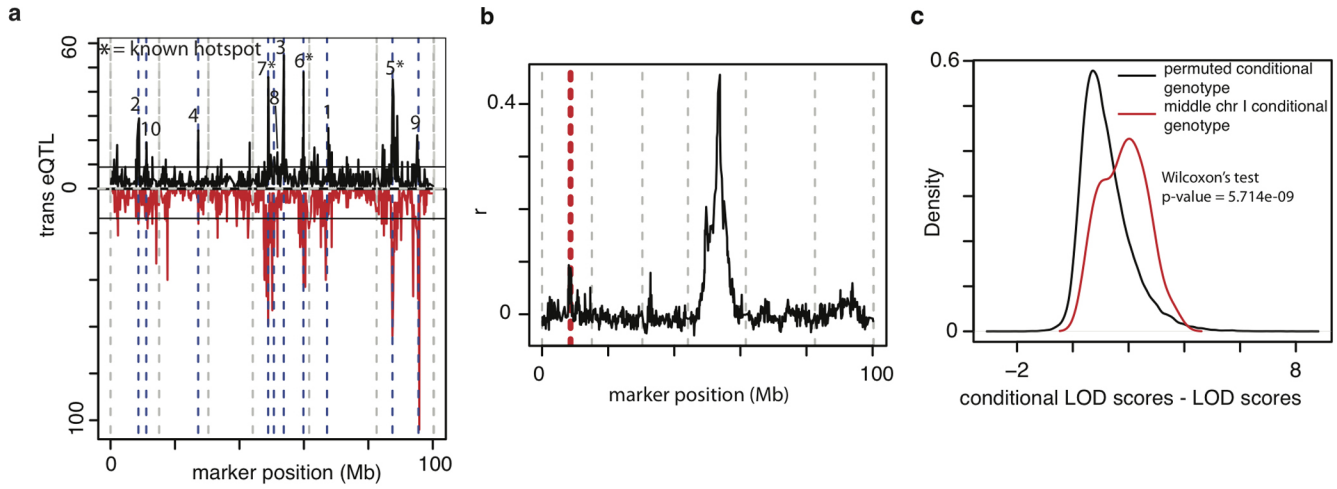

21. Rockman, M. V. & Kruglyak, L. Breeding designs for recombinant inbred advanced intercross lines. *Genetics* **179,** 1069–1078 (2008).
22. Rockman, M. V. & Kruglyak, L. Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet.* **5,** e1000419 (2009).
23. Kirienko, N. V. & Fay, D. S. Transcriptome profiling of the *C. elegans* Rb ortholog reveals diverse developmental roles. *Dev. Biol.* **305,** 674–684 (2007).
24. Capra, E. J., Skrovanek, S. M. & Kruglyak, L. Comparative developmental expression profiling of two *C. elegans* isolates. *PLoS ONE* **3,** e4055 (2008).
25. Ronald, J., Brem, R. B., Whittle, J. & Kruglyak, L. Local regulatory variation in *Saccharomyces cerevisiae*. *PLoS Genet.* **1,** e25 (2005).
26. Storey, J. D., Xiao, W., Leek, J. T., Tompkins, R. G. & Davis, R. W. Significance analysis of time course microarray experiments. *Proc. Natl Acad. Sci. USA* **102,** 12837–12842 (2005).
27. Seidel, H. S., Rockman, M. V. & Kruglyak, L. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science* **319,** 589–594 (2008).
28. Smith, E. N. & Kruglyak, L. Gene-environment interaction in yeast gene expression. *PLoS Biol.* **6,** e83 (2008).




**Extended Data Figure 1 | Inferring dynamic expression from individual expression profiles in each genotype and analysing the effect of local regulatory variation on the dynamics.** **a**, Genes related to oogenesis (red) and spermatogenesis (blue) have high scores on the first and second canonical variates, respectively. **b**, Canonical correlation analysis correctly sorts the samples in the reference expression time series. Reference data points are sorted by time along a trajectory that mirrors the one formed by RIAILs (Fig. 1) when projected onto the first two canonical variates. Numbers indicate the time in hours after the mid-L3 stage at which each reference sample was prepared. **c**, PCA on the uncentred RIAIL expression data reveals two outliers that are excluded from further analysis. **d**, PCA analysis of RIAIL gene expression profiles. Projection of the RIAILs onto first two forms a nonlinear trajectory. The first and second components (comp) explain 37% and 20% of the variance in the RIAILs expression data, respectively. **e**, Genes related to oogenesis (red) and spermatogenesis (blue) have high scores on the first and second principle components, respectively. **f**, Projections of the 206 RIAILs ranked by canonical correlation onto the first six canonical variates (i–vi). **g**, Projections of the reference time series onto the same canonical variates. Dynamic time warping of the RIAIL projections onto the reference projections is used to estimate the physiological age of each RIAIL. **h**, Projections of the 206 RIAILs onto the first six canonical variates (*y* axis) versus the final estimated age of each RIAIL (*x* axis). **i**, The number of eQTLs detected by a model that only includes local marker without considering time (no time), by a model that includes time and local marker additive effects (additive), and the number of 'dynamic' eQTLs; that is, the ones that are best explained by a model that also includes the interaction between local marker and time. **j**, Classification of the 'dynamic' eQTLs into non-mutually exclusive classes. **k**, Genes with eQTLs are in general biased towards the arms and tips of the chromosomes; the same is also true for genes with 'dynamic' eQTLs. Error bars represent 95% binomial confidence intervals. **l**, Density of polymorphisms in different regions of all genes, genes with no detected eQLTs, genes with any kind of eQTLs, and genes with 'dynamic' eQTLs. Intergenic regions are defined as regions between adjacent coding transcripts. 5′ UTR is defined as the region between the transcript start and the translation start site; 3′ UTR is the region between the stop codon and the transcript end. SNP density is defined as the number of SNPs divided by the length of the region of interest. Thus, non-synonymous and synonymous SNP densities are both defined as number of non-synonymous and synonymous SNPs divided by exon length. *P* values for enrichments were calculated using Fisher's exact test. Error bars represent 95% binomial confidence intervals.




**Extended Data Figure 2 | *Trans* hotspots regulate genes with diverse dynamics.** Left, clustered heat maps of the time-ordered expression of genes regulated by different *trans*-eQTL hotspots in RIAILs carrying the Bristol (red) or Hawaii (blue) alleles of the *trans* hotspot. Right, temporal expression profiles of the means of each of the indicated expression clusters when the relevant genomic region carries the Bristol (red) or Hawaii (blue) allele. The numbering of each hotspot follows that in Fig. 4B. Three *trans*-eQTL hotspots (numbers 5, 6 and 7) have been previously identified.




**Extended Data Figure 3 | Single maker analysis of *trans*-eQTLs.** **a**, The number of target genes per marker is shown for the random forest (black) and single marker analysis (red). *Trans* hotspots were defined using a Poisson test. According to this test, markers associated with 13 or more eQTLs have a significantly higher number of eQTLs than expected by chance at FDR = 0.1 in the single marker analysis, and so were defined as *trans* hotspots. Of the top 10 hotspots identified by random forest and presented in the manuscript, 9 are also identified by the single marker analysis. The missing hotspot is number 3 in the middle of chromosome IV, however 21 out of 55 genes (38%) regulated by this hotspot have at least one eQTL in the single marker analysis, and 11 of these are in the same region of chromosome IV as the hotspot, consistent with the random forest analysis. **b**, To investigate further why the random forest detects more genes as regulated by this hotspot, we calculated the genome-wide correlation between the empirical *P* values (from the random forest analysis) of the hotspot 3 marker and all other markers. The hotspot 2 marker (red dashed line) has the highest correlation aside from markers that are in linkage disequilibrium with the hotspot 3 marker. **c**, Marker log odds ratio (LOD) scores of genes in *trans* hotspot 3 are significantly higher when the genotype of the hotspot 2 marker is included in a linear model. This may contribute to why more genes are detected as regulated by hotspot 3 in the random forest—in which the influence of several markers is taken into account—than in the single marker analysis.

**Extended Data Table 1 | Robustness of the results to variation in time estimation**

**a**

| Canonical variates (CV) | mean | min | max |
|---|---|---|---|
| 1 and 2 | 8.98 | 2.31 | 15.72 |
| 1 to 3 | 9.96 | 2.69 | 16.14 |
| 1 to 4 | 10.46 | 4.48 | 16.1 |
| 1 to 5 | 10.49 | 4.39 | 16.66 |
| 1 to 6 | 10.54 | 4.58 | 16.71 |
| 1 to 7 | 10.51 | 4.58 | 16.62 |
| 1 to 8 | 10.51 | 4.58 | 16.62 |
| 1 to 9 | 10.55 | 4.58 | 16.43 |
| 1 to 10 | 10.57 | 4.58 | 16.43 |
| 1 to 11 | 10.61 | 4.72 | 16.43 |
| 1 to 12 | 10.61 | 4.72 | 16.43 |

**b**

| Age estimation | dynamic expression | Additive eQTL | Dynamic eQTL |
|---|---|---|---|
| relative age only | 14832(99.4%) | 4099 (96.6%) | 710 (78.5%) |
| 1 and 2 CV | 14867 (99.6%) | 4146 (97.6%) | 793 (87.7%) |
| 1 to 5 CV | 14892 (99.8%) | 4181 (98.2%) | 832 (92%) |
| 1 to 7 CV | 14894 (99.8%) | 4177 (98.1% | 839 (92.8%) |
| 1 to 12 CV | 14885 (99.7%) | 4147 (97.4%) | 824 (91.5%) |

**a**, Age estimation stability. Age estimation was performed using multivariate dynamic time warping and including an increasing number of canonical variates (CV) up to the maximum number of 12. The minimum, maximum and mean estimated age (in hours after mid-L3 stage) are shown. We considered the estimation to have stabilized after adding 6 canonical variates. **b**, Robustness of the *cis*-eQTL analysis to variation in age estimation. Shown are the number (and percentage) of genes that significantly change with age, with a significant additive *cis*-eQTL and significant dynamic *cis*-eQTL that are recovered from the original analysis (that uses 6 canonical variates to estimate age) when varying the age estimation. In the worst case, which occurs using only the relative age, we still recover almost 80% of the genes with dynamic eQTLs. Using one less or one more canonical variate in age estimation recovers well above 90% of the genes with dynamic eQTLs in the original analysis.

LETTER

doi:10.1038/nature12785

# HMGA2 functions as a competing endogenous RNA to promote lung cancer progression

Madhu S. Kumar[1], Elena Armenteros-Monterroso[1], Philip East[2], Probir Chakravorty[2], Nik Matthews[3], Monte M. Winslow[4] & Julian Downward[1,5]

**Non-small-cell lung cancer (NSCLC) is the most prevalent histological cancer subtype worldwide[1]. As the majority of patients present with invasive, metastatic disease[2], it is vital to understand the basis for lung cancer progression. *Hmga2* is highly expressed in metastatic lung adenocarcinoma, in which it contributes to cancer progression and metastasis[3–6]. Here we show that *Hmga2* promotes lung cancer progression in mouse and human cells by operating as a competing endogenous RNA (ceRNA)[7–11] for the let-7 microRNA (miRNA) family. Hmga2 can promote the transformation of lung cancer cells independent of protein-coding function but dependent upon the presence of let-7 sites; this occurs without changes in the levels of let-7 isoforms, suggesting that Hmga2 affects let-7 activity by altering miRNA targeting. These effects are also observed *in vivo*, where Hmga2 ceRNA activity drives lung cancer growth, invasion and dissemination. Integrated analysis of miRNA target prediction algorithms and metastatic lung cancer gene expression data reveals the TGF-β co-receptor Tgfbr3 (ref. 12) as a putative target of Hmga2 ceRNA function. Tgfbr3 expression is regulated by the Hmga2 ceRNA through differential recruitment to Argonaute 2 (Ago2), and TGF-β signalling driven by Tgfbr3 is important for Hmga2 to promote lung cancer progression. Finally, analysis of NSCLC-patient gene-expression data reveals that HMGA2 and TGFBR3 are coordinately regulated in NSCLC-patient material, a vital corollary to ceRNA function. Taken together, these results suggest that *Hmga2* promotes lung carcinogenesis both as a protein-coding gene and as a non-coding RNA; such dual-function regulation of gene-expression networks reflects a novel means by which oncogenes promote disease progression.**

The ceRNA hypothesis posits that specific RNAs can function as sinks for pools of active miRNAs, functionally liberating other transcripts targeted by that set of miRNAs[10]. Downregulation of the transcription factor Nkx2.1 promotes lung adenocarcinoma progression partially through derepression of Hmga2 (ref. 6), a non-histone chromosomal high-mobility group protein. Intriguingly, Hmga2 has been described as a prototypic let-7 target transcript, with seven conserved sites in its 3′ untranslated region (3′ UTR)[13]. Reduction of Hmga2 by RNA interference, which would deplete both Hmga2 protein and transcript, greatly reduces metastatic ability. Thus, it is possible that the transcript could operate independently of the protein in lung cancer progression.

To determine whether Hmga2 can operate as a ceRNA for the let-7 family, we generated an allelic series of Hmga2 expression constructs (Fig. 1a). In this series, we expressed the wild-type full-length *Hmga2* complementary DNA (WT); *Hmga2* with mutation of all seven predicted let-7 binding sites[13] (m7); *Hmga2* with mutation of the single in-frame start codon (ATG WT); or *Hmga2* with mutation of both the start codon and the let-7 binding sites (ATG m7). We then examined these constructs in two lung cancer cell lines generated from the $Kras^{LSL\text{-}G12D};Trp53^{flox/flox}$ mouse model: a cell line derived from a non-metastatic lung tumour that expresses very low levels of Hmga2 (368T1); and a cell line derived from a lymph node metastasis which expresses high levels of Hmga2 (482N1)[6]. Using two antibodies that recognize either the amino-terminus or the second AT-hook of the protein (M. Narita, personal communication), we found that the Hmga2 WT and m7 constructs efficiently express full-length Hmga2 protein (m7 overexpresses Hmga2 owing to loss of let-7-mediated suppression), whereas the Hmga2 ATG WT and ATG m7 constructs do not (Fig. 1b). Importantly, we observed similar levels of Hmga2 transcript expressed in the allelic series (in the case of the 482N1 cell line, the allelic series was mutated to abrogate binding to a short hairpin RNA (shRNA) against Hmga2) (Fig. 1c). Moreover, expression of the allelic series has no effect on the expression of various let-7 family members (Extended Data Fig. 1a). Taken together, this *Hmga2* allelic series allows us to compare specifically the roles of Hmga2 protein and transcript function on lung cell transformation.

We therefore compared the ability of the *Hmga2* allelic series to promote anchorage-independent growth of the lung cancer cell lines. We observed a striking promotion of soft-agar growth by both Hmga2 WT and ATG WT in the 368T1 and 482N1 cells (Figs. 1d, e); more modest growth was observed with Hmga2 m7, despite elevated protein expression relative to Hmga2 WT, and no growth was provided by Hmga2 ATG m7. This effect can be observed further in two additional human lung cancer cells (H1299 and SK-MES-1), as suppression of soft-agar growth by HMGA2 depletion can be rescued robustly by Hmga2 WT and ATG WT but more modestly by Hmga2 m7 (Extended Data Fig. 1b–e). Importantly, exogenous expression of let-7 reversed the ability of the *Hmga2* ceRNA to promote anchorage-independent growth, suggesting that let-7 regulates this effect (Extended Data Fig. 2b). To demonstrate that the effect of the Hmga2 ceRNA is driven by let-7 sites in the 3′ UTR, we expressed only the wild-type or let-7-site-mutated 3′ UTRs in 368T1 cells and examined the consequences on anchorage-independent growth. Notably, expression of the wild-type but not let-7-mutant 3′ UTR was sufficient to promote soft-agar growth in 368T1 cells (Extended Data Fig. 2c).

Beyond direct Hmga2 depletion, the Hmga2 WT and Hmga2 ATG WT constructs substantially rescued anchorage-independent growth in 482N1 cells stably overexpressing Nkx2.1, which we have previously shown to suppress lung cancer progression[6] (Extended Data Fig. 2d). Notably, this effect is not due to a general proliferative benefit of Hmga2 WT and ATG WT cells, as BrdU (bromodeoxyuridine; 5-bromo-2′-deoxyuridine) incorporation in adherent conditions was comparable across the allelic series (Extended Data Fig. 3a). However, when lung cancer cells are placed in suspension, Hmga2 depletion suppressed proliferation and this proliferation was rescued substantially by Hmga2 WT and ATG WT, and more marginally by Hmga2 m7 (Extended Data Fig. 3b); in contrast, the rate of apoptosis in the allelic series was

[1]Signal Transduction Laboratory, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3LY, UK. [2]Bioinformatics and Biostatistics Group, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3LY, UK. [3]Advanced Sequencing Facility, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3LY, UK. [4]Department of Genetics, Department of Pathology, the Stanford Cancer Institute, Stanford University School of Medicine, Stanford, California 94305, USA. [5]Lung Cancer Group, Division of Cancer Biology, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK.

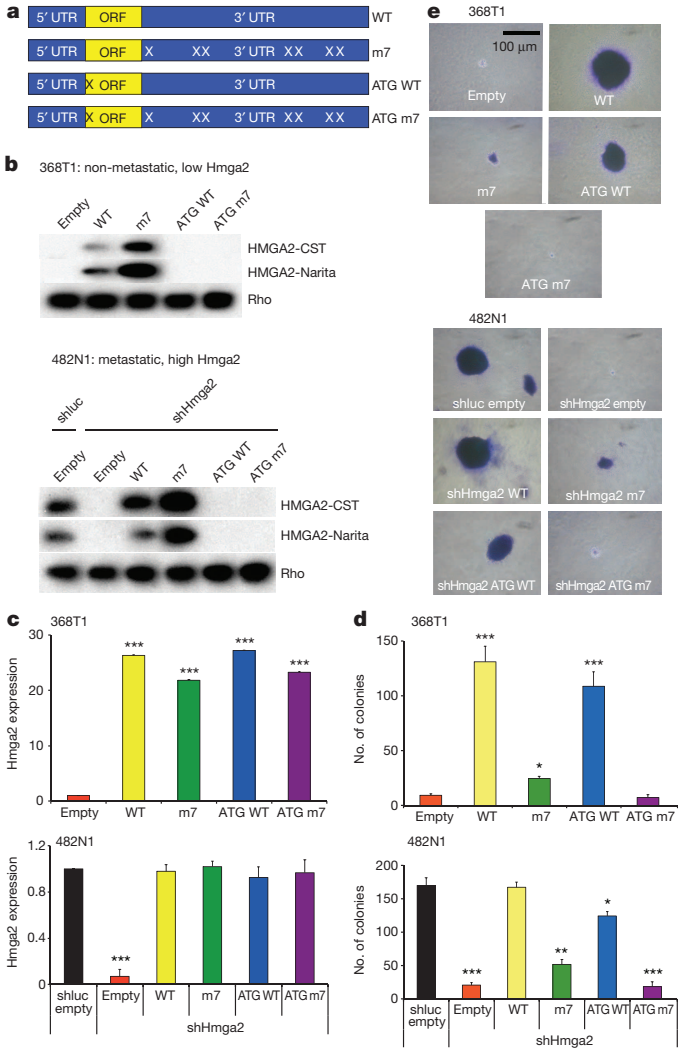



**Figure 1 | Hmga2 promotes lung cancer cell transformation in a protein-coding-independent but let-7-site-dependent manner.** **a**, Diagram of Hmga2 allelic series: expression constructs containing the entire Hmga2 cDNA ('WT'); the cDNA with all seven let-7 sites in the 3′ UTR mutated ('m7'); the cDNA with the start codon mutated ('ATG WT'); and the cDNA with both the start codon and let-7 sites mutated ('ATG m7'). ORF, open reading frame. Xs represent mutations of the let-7 sites in the Hmga2 3′ UTR. **b**, Hmga2 WT and m7 induce Hmga2 expression in non-metastatic lung cancer cells (368T1) and restore expression in metastatic lung cancer cells (482N1) depleted for endogenous Hmga2 (shHmga2). Two distinct HMGA2 antibodies are used: one recognizes the N-terminus of HMGA2 (HMGA2-CST) and the other recognizes the central AT-hook region of HMGA2 (HMGA2-Narita). **c**, Hmga2 RNA is comparably expressed by the WT, m7, ATG WT, and ATG m7 constructs in both 368T1 and 482N1 cells. Hmga2 expression is normalized to Gapdh. 368T1 values are normalized to empty and 482N1 values are normalized to shluc empty (which express a shRNA targeting luciferase and the empty expression vector). Values are technical triplicates, have been performed independently three times, and represent mean ± standard deviation (s.d.) with propagated error. **d**, Hmga2 WT and ATG WT promote substantial anchorage-independent growth in both 368T1 and 482N1 cells. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. **e**, Representative images of soft agar colonies. Magnification is ×10. $***P < 0.0005$; $**P < 0.005$; $*P < 0.05$.

not affected by growth in suspension (data not shown). Taken together, these results suggest that the Hmga2 transcript functions in a largely protein-coding-independent but let-7-site-dependent manner to promote lung cancer cell transformation *in vitro*.

To examine the effect of Hmga2 ceRNA activity on lung cancer cell dissemination *in vivo*, we intravenously transplanted 482N1 cells expressing either the control shRNA and expression construct or the shRNA targeting Hmga2 plus the *Hmga2* allelic series into syngeneic mice. As seen in Fig. 2a, micro-computed tomography (micro-CT) analysis revealed a substantial rescue of lung tumour formation with expression of either the Hmga2 WT or ATG WT constructs (with more modest effects with the Hmga2 m7 construct). Histopathological analysis and quantification of surface lesions confirmed the effects of Hmga2 ceRNA activity on *in vivo* lung tumorigenesis (Figs. 2b, c). Moreover, the control shRNA, the shHmga2 plus Hmga2 WT and the shHmga2 plus Hmga2 ATG WT transplants generate a highly metastatic disease, with lesions disseminating to both local and distant lymph nodes, kidney, and the abdominal and thoracic cavities (data not shown). We also examined the effect of Hmga2 ceRNA function on survival in the transplant system. We observed a dramatic reduction in survival in animals transplanted with the Hmga2 WT and ATG WT cells, similar to that seen with transplant of the control shRNA cells (Fig. 2d); we noted a more modest reduction in survival with transplant of the Hmga2 m7 cells. In total, these findings indicate that Hmga2 competing RNA activity dramatically promotes *in vivo* lung cancer formation.

To elucidate the mechanism of Hmga2 ceRNA function on lung carcinogenesis, we analysed the set of genes differentially expressed between metastatic and non-metastatic $Kras^{G12D};Trp53^{-/-}$ lung cancer cells[6] and compared them to the list of predicted let-7 target genes based upon the miRNA target prediction algorithm TargetScan[14] (Supplementary Table 5). Kras was not a candidate in this analysis, in spite of previous description of Kras as an important let-7 target[15]. Moreover, the *Hmga2* allelic series had no impact on either expression of K-Ras protein or activity of downstream K-Ras signalling pathways (Extended Data Fig. 3c). In contrast, we observed several known Hmga2 transcriptional targets, including components of the Igf2bp family[16], validating this approach.

To elucidate more broadly which transcripts are Hmga2 ceRNA targets, we initially examined whether let-7 sites are enriched among transcripts induced by the Hmga2 ceRNA through RNA-seq of the 482N1 allelic series combined with Sylamer analysis, which detects miRNA seed sites as nucleotide strings enriched within the 3′ UTRs of transcripts[17]. We first compared control to Hmga2-knockdown cells and observed a specific enrichment of let-7 sites lost with Hmga2 depletion (Extended Data Fig. 4a). We then determined whether this was specific to ceRNA activity by determining let-7-site enrichment upon re-expression of either the Hmga2 WT or ATG WT constructs in Hmga2-knockdown cells. In both conditions, let-7 sites were enriched among the upregulated transcripts (Extended Data Fig. 4b, c) Importantly, let-7 sites were not enriched with re-expression of either Hmga2 m7 or ATG m7 in the Hmga2-depleted background (Extended Data Fig. 4d, e). Moreover, analysis of fragments per kilobase of exon per million fragments mapped (FPKM) in the RNA sequencing (RNA-seq) results from control 482N1 cells showed Hmga2 was among the most highly expressed predicted let-7 target transcripts, suggesting that Hmga2 constitutes a physiologically germane fraction of the let-7 target milieu (Supplementary Table 6). Taken together, these results indicate that the Hmga2 ceRNA broadly regulates let-7 targets.

To assess Hmga2 ceRNA targets more specifically, we examined which transcripts were suppressed in response to Hmga2 depletion; 13 out of 34 predicted targets were suppressed by Hmga2 knockdown (Extended Data Fig. 5a). To delineate which of these were Hmga2 transcriptional targets versus ceRNA targets, we re-expressed either Hmga2 WT or ATG WT in knockdown cells. As seen in Extended Data Fig. 5b, 6 out of 13 transcripts were rescued by both Hmga2 WT and ATG WT, suggesting they are putative ceRNA targets; the remaining targets were rescued only by Hmga2 WT, suggesting they are targets of Hmga2 transcription factor function. These Hmga2 ceRNA targets were markedly enriched in let-7-regulated transcripts, as their repression by Hmga2 loss could be reversed with the use of a 'tough decoy' let-7 sponge transcript, designed to be an efficient and long-term

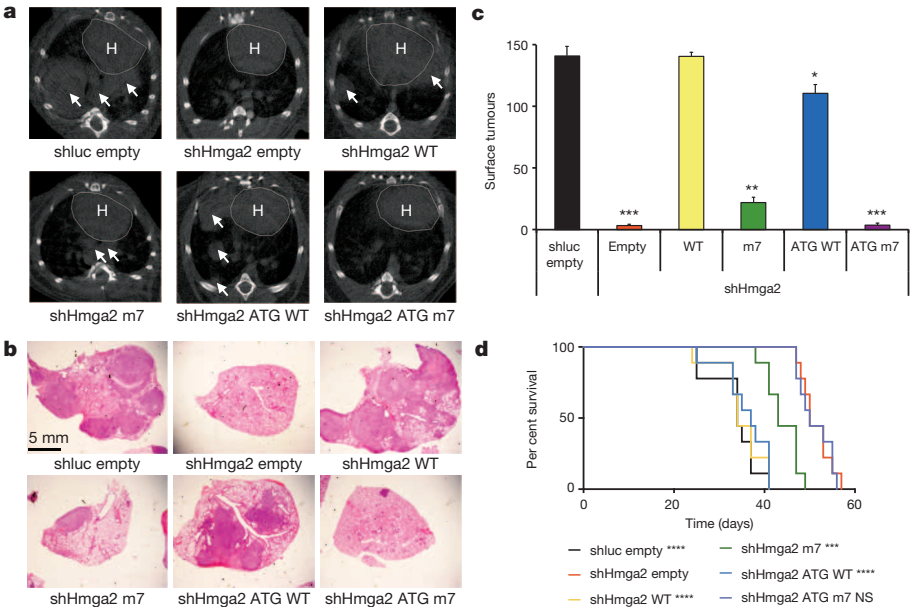



**Figure 2 | Hmga2 ceRNA activity enhances lung cancer progression *in vivo*.** **a**, Hmga2 WT and ATG WT restore lung tumour growth in response to endogenous Hmga2 knockdown. B6129SF1/Tac males were intravenously injected with 482N1 cells expressing either a control shRNA and empty vector (shluc empty) or shHmga2 with the *Hmga2* allelic series. Three weeks afterwards, animals were scanned by micro-CT and representative transverse images are shown. The heart is demarcated ('H') and white arrows identify lung tumours. **b**, Representative histological images of lungs transplanted with 482N1 cells from the series described in **a**. Magnification is ×1. **c**, Lung surface tumour counts were taken from animals transplanted with 482N1 cells from the series described in **a** ($n = 3$ animals per group). Values are technical triplicates and represent mean ± s.e.m. **d**, Hmga2 WT and ATG WT substantially reduce survival of animals transplanted with 482N1 cells expressing the shRNA targeting Hmga2. Animals were intravenously transplanted with cells as in **a**. Animals were subsequently aged for survival and a Kaplan-Meier analysis was performed ($n = 9$ animals per group). Median survival was 34 days for shluc empty and shHmga2 WT transplants; 37 days for shHmga2 ATG WT transplants; 43 days for shHmga2 m7 transplants; and 50 days for shHmga2 empty and ATG m7 transplants. Statistical significance was assessed by log-rank tests compared to shHmga2 empty. ****$P < 0.00005$; ***$P < 0.0005$; **$P < 0.005$; *$P < 0.05$; NS, not significant.

suppressor of miRNA function (Extended Data Fig. 5c)[18]. Conversely, overexpression of let-7 suppressed these transcripts, although Hmga2 transcriptional targets were also affected owing to depletion of Hmga2 (Extended Data Fig. 5d). Taken together, these studies outline a collection of putative target transcripts regulated by Hmga2 ceRNA function.

Among these six Hmga2 ceRNA target transcripts, we found the TGF-β co-receptor Tgfbr3 (ref. 12) to be both upregulated in metastatic lung cancer cells and a putative let-7 target. Furthermore, several Hmga2 ceRNA targets have been described as targets of TGF-β signalling[19]. Thus, we examined whether Hmga2 exerts ceRNA function through enhanced TGF-β signalling via Tgfbr3. Consistent with this, we found that in both 368T1 and 482N1 cells, Hmga2 WT and ATG WT promote the expression of Tgfbr3 protein (Fig. 3a). This Tgfbr3 upregulation also occurs to a lesser degree at the messenger RNA level, as has been described previously for miRNA targets[20] (Fig. 3b). Moreover, exogenous expression of let-7 reversed the ability of the Hmga2 ceRNA to upregulate Tgfbr3, suggesting that this effect is controlled by let-7 (Extended Data Fig. 2a). An important consideration in ceRNA–target analysis is the absolute levels of Hmga2, Tgfbr3 and let-7 transcripts in cells, so we determined the copies per cell of these factors (Extended Data Fig. 4f). We observed that Hmga2 and Tgfbr3 had similar levels of transcript, as might be expected for two factors that can titrate expression of one another; similar results were observed in FPKM analysis of these transcripts in control 482N1 cells by RNA-seq (Supplementary Table 6). Furthermore, total let-7 family expression was within an order of magnitude of Hmga2 and Tgfbr3. As this pool of let-7 regulates the entire target set, it is possible that miRNA occupancy to be a limiting factor, allowing for the contribution of a ceRNA-like Hmga2. Taken together, these results suggest that Hmga2 may regulate Tgfbr3 expression as a let-7 ceRNA.

In line with these observations of Hmga2 promoting Tgfbr3 expression, Hmga2 WT and ATG WT activated TGF-β signalling through phosphorylation of Smad2 (Fig. 3a). This effect was let-7-dependent, as exogenous let-7 suppressed Smad2 phosphorylation (Extended Data Fig. 2a). It is likely that the TGF-β pathway is active in the absence of exogenous ligand owing to low but detectable levels of TGF-β in serum during cell culture[21]. We further examined whether Hmga2 ceRNA function affects TGF-β pathway activation by two methods. First, we found that a TGF-β reporter (CAGA12)[22] was potently induced by Hmga2 WT and ATG WT (Extended Data Fig. 5e). Second, analysis of TGF-β target transcript levels revealed specific expression of these genes with the Hmga2 WT and ATG WT constructs (Extended Data Fig. 5f). Notably, we observed little activation of the TGF-β pathway by Hmga2 m7, despite previous reports of Hmga2 functioning as a co-activator for Smad2, Smad3 and Smad4 in the epithelial–mesenchymal transition (EMT)[23]; this is likely to be due to the lack of upstream activation of the pathway. Consistent with this, the Hmga2 ceRNA does not induce an EMT in either 368T1 or 482N1 cells (Extended Data Fig. 6a). Overall, these results indicate that the Hmga2 ceRNA induces expression of Tgfbr3 and potentiates TGF-β signalling.

To determine whether the effects of Hmga2 on Tgfbr3 occur through let-7-mediated derepression, we first examined the effect of Hmga2 ceRNA function on a reporter containing the *Tgfbr3* 3′ UTR. In both 368T1 and 482N1 cells, Hmga2 WT and ATG WT induced expression of luciferase under the control of the wild-type *Tgfbr3* 3′ UTR, but not if the let-7 site was mutated (Fig. 3c). Furthermore, we found that the effect of the Hmga2 ceRNA was broadly miRNA-dependent, as Hmga2 WT and ATG WT induced the reporter expression in *Dicer1*-intact sarcoma cells, but not in a *Dicer1*-null derivative cell line[24] (Extended Data Fig. 6b). To assess directly whether Hmga2 induces Tgfbr3 through competition away from Ago2, we performed RNA immunoprecipitation (RIP) on Ago2 in lung cancer cells expressing the *Hmga2* allelic series. As shown in Fig. 3d, we found that Hmga2 WT and ATG WT were recruited to Ago2 at levels comparable to the

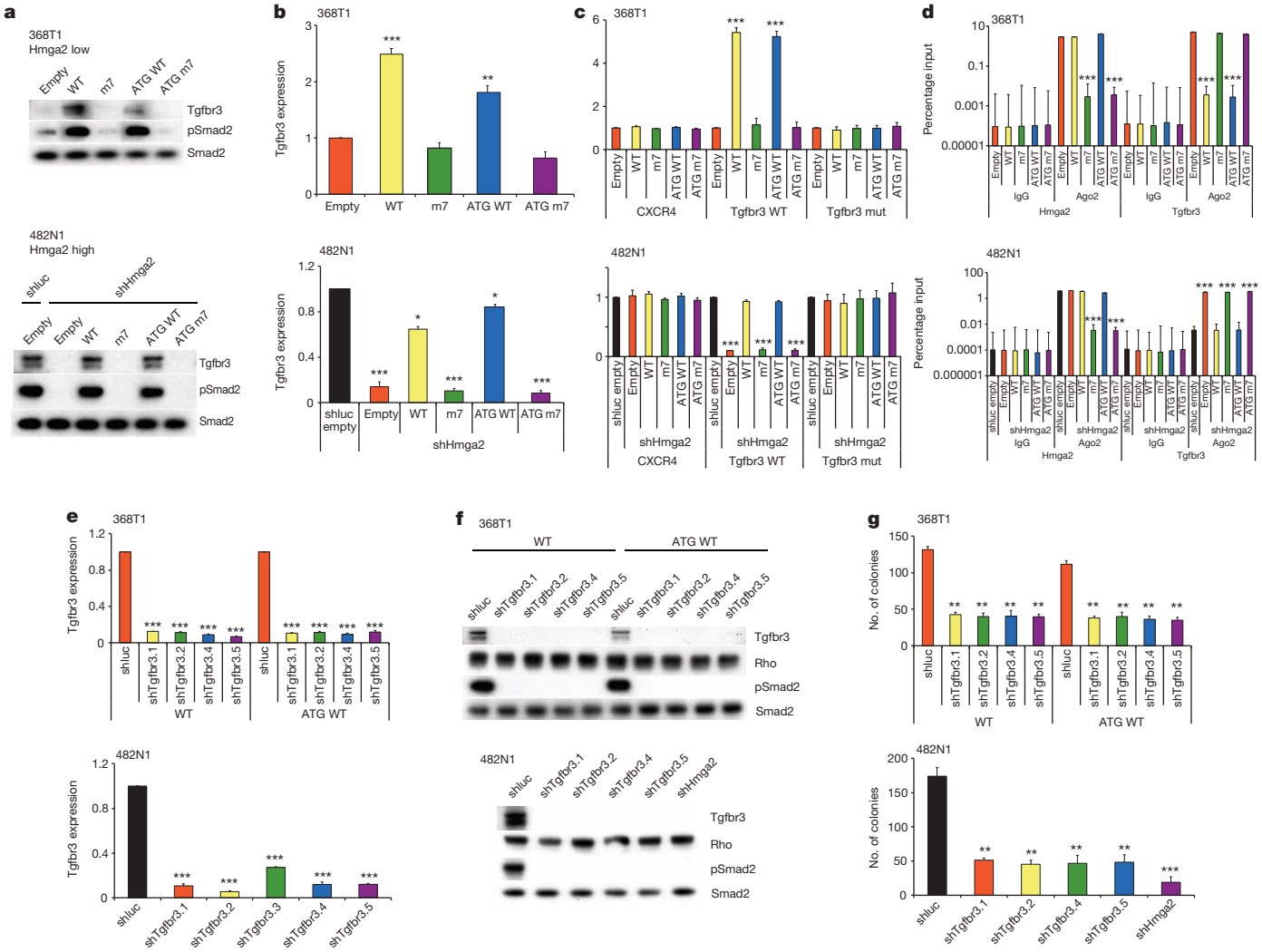



**Figure 3 | Hmga2 ceRNA activity enhances TGF-β signalling through overexpression of Tgfbr3.** **a**, Hmga2 WT and ATG WT substantially induce both Tgfbr3 protein expression and phosphorylation of Smad2 (pSmad2) in both 368T1 and 482N1 cells. **b**, Hmga2 WT and ATG WT significantly promote expression of Tgfbr3 mRNA in both 368T1 and 482N1 cells. Tgfbr3 expression is normalized to Gapdh. 368T1 values are normalized to empty and 482N1 values are normalized to shluc empty. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. with propagated error. **c**, Hmga2 WT and ATG WT specifically induce expression of a luciferase *Tgfbr3* 3′ UTR reporter in a *let-7*-site-dependent manner in both 368T1 and 482N1 cells. Cells were transfected with *Renilla* constructs of the control siCXCR4 multimer[30] and either the *Tgfbr3* wild-type or let-7-mutant 3′ UTR reporter. Values are normalized to co-transfected pGL3 plasmid. 368T1 values are normalized to empty and 482N1 values are normalized to shluc empty. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. with propagated error. **d**, Hmga2 WT and ATG WT displace Tgfbr3 from Ago2-based RNA-induced silencing complexes. Lysates from 368T1 and 482N1 cells of the *Hmga2* allelic series underwent either control immunoprecipitation (IgG) or immunoprecipitation for Ago2. RNA was purified and qRT–PCR was performed for Hmga2 and Tgfbr3 on both the immunoprecipitated and input RNAs. Values are depicted as the percentage of input RNA, are technical triplicates, have been performed independently twice, and represent mean ± s.d. **e**, Multiple shRNAs elicit substantial knockdown of Tgfbr3 mRNA in both 368T1 and 482N1 cells. 482N1 cells were infected with control shRNA (shluc) or a set of shRNAs targeting Tgfbr3 (shTgfbr3.1–3.5), and 368T1 WT and ATG WT cells were infected with shluc or shTgfbr3.1, 3.2, 3.4 and 3.5. RNA was purified and qRT–PCR was performed. Tgfbr3 expression is normalized to Gapdh and 368T1 WT and ATG WT and 482N1 values are normalized to shluc. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. with propagated error. **f**, Multiple shRNAs induce knockdown of Tgfbr3 and suppress TGF-β pathway activity in 368T1 and 482N1 cells. Cells were infected with shRNAs as in **e** and Western blot analysis was performed for Tgfbr3, pSmad2 and total Smad2 (Smad2). **g**, Tgfbr3 depletion reduces anchorage-independent growth of 368T1 WT and ATG WT and 482N1 cells. Cells were infected with the listed shRNAs and plated for anchorage-independent growth and colonies were counted as above. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. $***P < 0.0005$; $**P < 0.005$; $*P < 0.05$.

recruitment of endogenous Hmga2, whereas Hmga2 m7 and ATG m7 were not. Moreover, Hmga2 WT and ATG WT cells had a substantial decrease in Tgfbr3 recruitment to Ago2. Of note, these effects on Ago2 occupancy by Tgfbr3 are not caused by a change in let-7 activity, as various let-7 family members were comparably loaded on Ago2 across the *Hmga2* allelic series (Extended Data Fig. 7). Thus, these results demonstrate Hmga2, through its let-7-binding sites, displaces Tgfbr3 from miRNA-mediated repression by Ago2. In total, these results suggest that the Hmga2 ceRNA directly functions by blocking recruitment of Tgfbr3 to the Ago2-based miRNA repression complex.

To examine whether Hmga2 ceRNA activity through Tgfbr3 is functionally relevant, we used shRNAs to deplete Tgfbr3 in 482N1 cells and 368T1 cells expressing either Hmga2 WT or ATG WT. At both the mRNA and protein level, multiple shRNAs reduced Tgfbr3 expression (Figs. 3e, f). Moreover, knockdown of Tgfbr3 led to substantial suppression of TGF-β signalling, as evidenced by loss of Smad2

phosphorylation, CAGA12 reporter activity and expression of TGF-β target genes (Fig. 3f and Extended Data Fig. 8a, b). We then assessed the functional effect of Tgfbr3 loss on Hmga2 ceRNA-driven soft-agar colony formation. In all cells, Tgfbr3 knockdown notably suppressed anchorage-independent growth, but not to the same extent as Hmga2 depletion in 482N1 cells (Fig. 3g). This occurred without generally affecting proliferation, as measured by BrdU incorporation (Extended Data Fig. 8c). We further functionally analysed the broader set of six Hmga2 ceRNA targets by individually depleting them by short interfering RNA (siRNA) (Extended Data Fig. 8d). When we compared their effects on anchorage-independent growth, both Hmga2 and Tgfbr3 loss strongly suppressed growth, Hmga1 depletion modestly reduced colony formation, whereas the remaining targets had little effect (Extended Data Fig. 8e). It should be noted that these other targets include extracellular factors like Angptl2 and Col1a2 that might promote lung cancer progression *in vivo* in a non-cell-autonomous manner, and could thus still be relevant to Hmga2 ceRNA activity in lung cancer progression. This considered, these results still indicate that Tgfbr3, although certainly not the only relevant Hmga2 ceRNA target, is an important effector of Hmga2 ceRNA function in lung cancer cell transformation.

To determine whether this effect of Tgfbr3 is driven through potentiation of TGF-β signalling, we inhibited the TGF-β pathway with the TGF-β-receptor-kinase inhibitor SB-431542 (SB)[25]. In 368T1 and 482N1 cells, SB treatment led to a substantial inhibition of Smad2 phosphorylation (Extended Data Fig. 9a). In addition, SB treatment of 482N1 and 368T1 Hmga2 WT and ATG WT cells markedly suppressed CAGA12 reporter activity and expression of TGF-β target genes (Extended Data Fig. 9b, c). We then examined whether SB could inhibit Hmga2 ceRNA-driven soft-agar colony formation and observed a striking reduction in anchorage-independent growth (Extended Data Fig. 9d). This impaired colony formation was not due to general proliferative arrest, as SB treated cells had a similar rate of BrdU incorporation (Extended Data Fig. 9e). Notably, many Hmga2 ceRNA targets are in fact TGF-β target genes, as their expression is suppressed by SB treatment and induced by exogenous addition of TGF-β (Extended Data Fig. 9f, g). Thus, it is possible that Hmga2 could function in a feed-forward loop in which it regulates TGF-β target genes directly through ceRNA function and indirectly through TGF-β signalling via Tgfbr3. In summary, these results indicate that TGF-β signalling through Tgfbr3 is an important pathway downstream of Hmga2 ceRNA function.

Based on the above findings, we wanted to examine whether HMGA2 functions as a ceRNA for TGFBR3 (the human orthologues of Hmga2 and Tgfbr3, respectively) in NSCLC patients. An important corollary of the ceRNA hypothesis is the coordinate regulation of a competing RNA and its targets, such that upregulation of the ceRNA should lead to higher expression of the target RNA and vice versa[10]. To assess this, we used NSCLC gene-expression data generated by the Cancer Genome Atlas (TCGA) and sorted the patient cohort into the top and bottom quartiles of HMGA2 expression (HMGA2 high and low, respectively) (Fig. 4a). As seen in Fig. 4b, we observed significantly higher levels of TGFBR3 transcript in HMGA2 high versus low patient samples. To address the converse relationship, we sorted the TCGA data set into top and bottom quartiles of TGFBR3 expression (TGFBR3 high and low, respectively) (Fig. 4c). When we compared HMGA2 transcript levels between the groups, we found *HMGA2* to be significantly overexpressed in TGFBR3 high versus low patient samples (Fig. 4d). To extend and validate these findings, we carried out similar gene-expression analyses of HMGA2 and TGFBR3 in an independent lung-adenocarcinoma-patient gene-expression cohort, the Director's Challenge data set[26]. Similar to the findings with the TCGA cohort, we observed HMGA2 and TGFBR3 to be coordinately expressed in the Director's Challenge data set (Extended Data Fig. 10a–d). Although we focused specifically on high and low expressors of HMGA2 and TGFBR3, for which ceRNA activity is more likely to occur[10], HMGA2 and TGBR3 expression was broadly correlated across both data sets (Extended Data

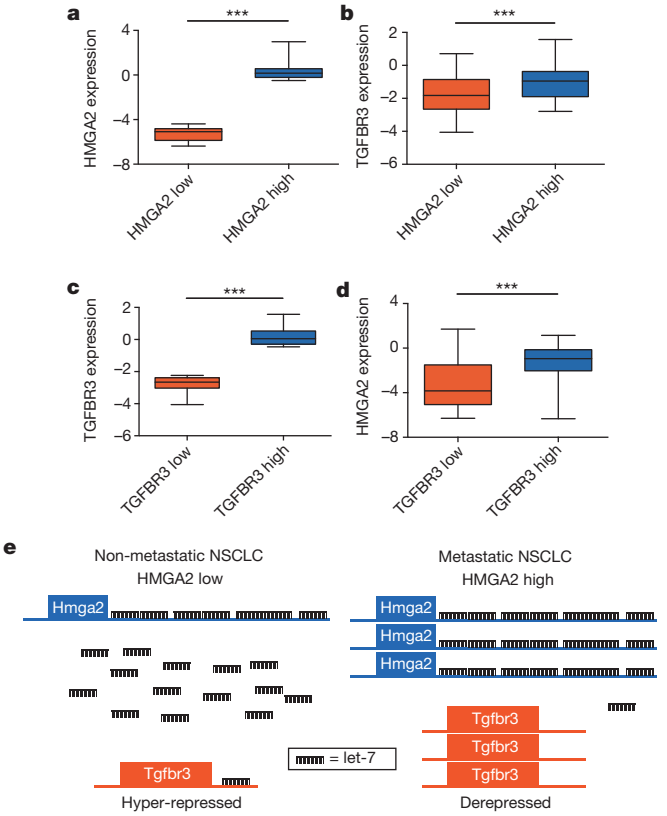



**Figure 4 | HMGA2 and TGFBR3 are reciprocally and coordinately upregulated in NSCLC patients.** **a**, The Cancer Genome Atlas (TCGA) NSCLC gene-expression data set was sorted according to HMGA2 expression. The top and bottom quartiles (HMGA2 low and high, respectively) were selected (45 patients per group) and HMGA2 expression was compared using box-and-whisker plots. **b**, The TCGA data set was sorted into top and bottom quartiles of HMGA2 expression as in **a**, and TGFBR3 expression was compared using box-and-whisker plots. **c**, The TCGA data set was sorted into top and bottom quartiles of TGFBR3 expression (TGFBR3 low and high, respectively) as in **a**, and TGFBR3 expression was compared using box-and-whisker plots. **d**, The TCGA data set was sorted into top and bottom quartiles of TGFBR3 expression as in c, and HMGA2 expression was compared using box-and-whisker plots. In all box-and-whisker plots, values are presented on a $\log_2$ scale. Significance was assessed by the Mann–Whitney test with a Bonferroni correction for multiple hypothesis testing. ***$P < 0.0005$. **e**, Model for Hmga2 acting as a competing endogenous RNA for Tgfbr3. In non-metastatic NSCLC, Hmga2 expression is low, leading to suppressed Tgfbr3 expression by let-7. In metastatic NSCLC, Hmga2 expression is elevated, titrating away let-7 from Tgfbr3 and allowing for its overexpression. This titration occurs without changes in let-7 expression, reflecting competition for microRNA occupancy by target transcripts.

Fig. 10e, f). As these two data sets constitute two of the largest collections of NSCLC gene-expression data available, we believe that these findings are consistent with HMGA2 functioning as a ceRNA for TGFBR3 in NSCLC patients. It is possible that the coexpression of HMGA2 and TGFBR3 could correspond to additional tumour characteristics that these data sets do not include; future studies in independent data sets would be needed to assess this issue. In total, our results suggest a model in which Hmga2 promotes lung cancer progression by competing for let-7 occupancy with other targets, including Tgfbr3, leading to the upregulation of those targets (Fig. 4e). Importantly, this occurs without changes in the levels of let-7 family microRNAs, reflecting specific competition for microRNA binding among targets.

Here we have outlined a novel gene-expression pathway in which a protein-coding gene, *Hmga2*, operates largely independently of its protein-coding function to promote lung cancer progression as a competing endogenous RNA. Although much of this ceRNA activity is

driven by overexpression of TGF-β signalling through Tgfbr3, there are likely to be additional Hmga2 ceRNA targets to be found in future studies. Moreover, *HMGA2* is overexpressed in many other cancer types[27], so it is possible that HMGA2 functions as a ceRNA in cancer sites beyond lung. Taken more broadly, these findings raise the possibility that many protein-coding genes differentially expressed in cancer might contribute to tumorigenesis through this distinct mode of regulatory gene expression. Moreover, these results raise issues with the validation of candidates in RNA interference screens. The 'gold standard' assay for validating an siRNA target is expression of an siRNA-resistant form of the coding sequence[28]; however, such an approach overlooks the possibility that depletion of both the full-length RNA and protein might contribute to a given phenotype, requiring complementation by the full-length transcript. Such dual-function ceRNA and protein activities necessitate a deeper exploration of the coding genome in biological systems.

## METHODS SUMMARY

**Soft-agar assays.** Soft-agar assays were performed essentially as described previously[6]. Assays were carried out in triplicate and quantified by microscope.

**Intravenous transplantation.** Intravenous injection was performed on 12-week-old B6129SF1/Tac male mice (Taconic), essentially as described previously[6]. In short, $10^5$ cells in 50 μl PBS were injected into 12 animals per group. Three weeks post injection, animals were scanned using the SkyScan 1176 micro-CT scanner as described previously[29]. Three mice per group were then euthanized at random for surface tumour and histopathological analysis. Surface tumours were quantified by counting all the visible tumours on the lung pleura; quantification was carried out blind to the expression construct. The remaining nine mice were aged for survival analysis. All procedures were performed under an approved project license as per UK Home Office regulations.

**RNA immunoprecipitation.** RNA immunoprecipitation was carried out on 482N1 and 368T1 cells with control antibodies (immunoglobulin G (IgG)) or antibody targeting Ago2 as per manufacturers' instructions (Millipore). Total RNA was used for either qRT–PCR (quantitative polymerase chain reaction with reverse transcription) of mRNAs or miRNA-specific qRT-PCR as above.

**Public gene-expression array analysis.** NSCLC gene-expression data sets (from both The Cancer Genome Atlas (TCGA) and the Director's Challenge[26]) were downloaded and processed using standard methods. Patient expression profiles were sorted by HMGA2- or TGFBR3-expression status, and the top and bottom quartiles of both groups were selected. Target gene expression was then analysed and represented as box-and-whisker plots. Statistical significance was assessed using Mann–Whitney tests with correction for multiple hypothesis testing.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 5 February; accepted 16 October 2013.**
**Published online 4 December 2013.**

1. Ferlay, J. *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* **127,** 2893–2917 (2010).
2. Herbst, R. S., Heymach, J. V. & Lippman, S. M. Lung cancer. *N. Engl. J. Med.* **359,** 1367–1380 (2008).
3. Di Cello, F. *et al.* HMGA2 participates in transformation in human lung cancer. *Mol. Cancer Res.* **6,** 743–750 (2008).
4. Meyer, B. *et al.* HMGA2 overexpression in non-small cell lung cancer. *Mol. Carcinog.* **46,** 503–511 (2007).
5. Sarhadi, V. K. *et al.* Increased expression of high mobility group A proteins in lung cancer. *J. Pathol.* **209,** 206–212 (2006).
6. Winslow, M. M. *et al.* Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature* **473,** 101–104 (2011).
7. Ebert, M. S., Neilson, J. R. & Sharp, P. A. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature Methods* **4,** 721–726 (2007).
8. Karreth, F. A. *et al.* In vivo identification of tumor- suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell* **147,** 382–395 (2011).
9. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465,** 1033–1038 (2010).
10. Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146,** 353–358 (2011).
11. Tay, Y. *et al.* Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* **147,** 344–357 (2011).
12. Wang, X. F. *et al.* Expression cloning and characterization of the TGF-β type III receptor. *Cell* **67,** 797–805 (1991).
13. Mayr, C., Hemann, M. T. & Bartel, D. P. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science* **315,** 1576–1579 (2007).
14. Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19,** 92–105 (2009).
15. Johnson, S. M. *et al.* RAS is regulated by the let-7 microRNA family. *Cell* **120,** 635–647 (2005).
16. Brants, J. R. *et al.* Differential regulation of the insulin-like growth factor II mRNA-binding protein genes by architectural transcription factor HMGA2. *FEBS Lett.* **569,** 277–283 (2004).
17. van Dongen, S., Abreu-Goodger, C. & Enright, A. J. Detecting microRNA binding and siRNA off-target effects from expression data. *Nature Methods* **5,** 1023–1025 (2008).
18. Xie, J. *et al.* Long-term, efficient inhibition of microRNA function in mice using rAAV vectors. *Nature Methods* **9,** 403–409 (2012).
19. Massagué, J. TGFβ signalling in context. *Nature Rev. Mol. Cell Biol.* **13,** 616–630 (2012).
20. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466,** 835–840 (2010).
21. Oida, T. & Weiner, H. L. Depletion of TGF-β from fetal bovine serum. *J. Immunol. Methods* **362,** 195–198 (2010).
22. Dennler, S. *et al.* Direct binding of Smad3 and Smad4 to critical TGF β-inducible elements in the promoter of human plasminogen activator inhibitor-type 1 gene. *EMBO J.* **17,** 3091–3100 (1998).
23. Thuault, S. *et al.* HMGA2 and Smads co-regulate SNAIL1 expression during induction of epithelial-to-mesenchymal transition. *J. Biol. Chem.* **283,** 33437–33446 (2008).
24. Ravi, A. *et al.* Proliferation and tumorigenesis of a murine sarcoma cell line in the absence of DICER1. *Cancer Cell* **21,** 848–855 (2012).
25. Inman, G. J. *et al.* SB-431542 is a potent and specific inhibitor of transforming growth factor-beta superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. *Mol. Pharmacol.* **62,** 65–74 (2002).
26. Shedden, K. *et al.* Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nature Med.* **14,** 822–827 (2008).
27. Park, S. M. *et al.* Let-7 prevents early cancer progression by suppressing expression of the embryonic gene HMGA2. *Cell Cycle* **6,** 2585–2590 (2007).
28. Kaelin, W. G., Jr. Molecular biology. Use and abuse of RNAi to study mammalian gene function. *Science* **337,** 421–422 (2012).
29. Kumar, M. S. *et al.* The GATA2 transcriptional network is requisite for RAS oncogene-driven non-small cell lung cancer. *Cell* **149,** 642–655 (2012).
30. Doench, J. G., Petersen, C. P. & Sharp, P. A. siRNAs can function as miRNAs. *Genes Dev.* **17,** 438–442 (2003).

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

**Acknowledgements** We thank the laboratory of T. Jacks for providing the 482N1 and 368T1 lung cancer cell lines and the laboratories of P.A. Sharp and P. Chambon for providing the KPD sarcoma cell lines. We thank the LRI FACS Facility for the BrdU analyses and the LRI Biological Resources Unit for assistance with the animal studies. We thank M.S. Ebert and E. De Bruin for critical review of the manuscript. M.S.K. is a Long-Term Fellow of the Human Frontier Science Program and the European Molecular Biology Organization. This work was funded by Cancer Research UK and by the European Commission's Seventh Framework Programme (FP7/2007-2013) under the grant agreement Lungtarget (project no. 259770).

**Author Contributions** M.S.K. and J.D. designed the study. M.S.K. and E.A.-M. performed the experiments described. M.S.K., P.E. and P.C. conducted bioinformatics analyses. N.M. performed the next-generation sequencing studies. M.M.W. provided necessary reagents and conceptual advice. M.S.K. and J.D. wrote the manuscript.

**Author Information** Sequence data have been uploaded to the Gene Expression Omnibus under accession number GSE50932. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D. (julian.downward@cancer.org.uk).

LETTER

doi:10.1038/nature12799

# Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids

C. J. Cambier[1], Kevin K. Takaki[2], Ryan P. Larson[1,3], Rafael E. Hernandez[4], David M. Tobin[2], Kevin B. Urdahl[1,3,4], Christine L. Cosma[2] & Lalita Ramakrishnan[1,2,5]

**The evolutionary survival of *Mycobacterium tuberculosis*, the cause of human tuberculosis, depends on its ability to invade the host, replicate, and transmit infection. At its initial peripheral infection site in the distal lung airways, *M. tuberculosis* infects macrophages, which transport it to deeper tissues[1]. How mycobacteria survive in these broadly microbicidal cells is an important question. Here we show in mice and zebrafish that *M. tuberculosis*, and its close pathogenic relative *Mycobacterium marinum*, preferentially recruit and infect permissive macrophages while evading microbicidal ones. This immune evasion is accomplished by using cell-surface-associated phthiocerol dimycoceroserate (PDIM) lipids[2] to mask underlying pathogen-associated molecular patterns (PAMPs). In the absence of PDIM, these PAMPs signal a Toll-like receptor (TLR)-dependent recruitment of macrophages that produce microbicidal reactive nitrogen species. Concordantly, the related phenolic glycolipids (PGLs)[2] promote the recruitment of permissive macrophages through a host chemokine receptor 2 (CCR2)-mediated pathway. Thus, we have identified coordinated roles for PDIM, known to be essential for mycobacterial virulence[3], and PGL, which (along with CCR2) is known to be associated with human tuberculosis[4,5]. Our findings also suggest an explanation for the longstanding observation that *M. tuberculosis* initiates infection in the relatively sterile environment of the lower respiratory tract, rather than in the upper respiratory tract, where resident microflora and inhaled environmental microbes may continually recruit microbicidal macrophages through TLR-dependent signalling.**

Pattern recognition receptors (PRRs) such as the TLRs enable host recognition of diverse microbes through their PAMPs[6]. Macrophages recruited through TLR signalling pathways can eradicate organisms invading the oropharyngeal mucosa, for example, *Streptococcus pneumoniae*[7]. In contrast, pathogenic mycobacteria appear to use macrophages and myeloid dendritic cells for transport across epithelial barriers to their infection niche[1,8]. Mycobacteria are replete with TLR PAMPs—such as lipoproteins and bacterial cell wall peptidoglycan—that have been shown to activate cytokine responses in cultured macrophages[1]. Yet *in vivo* studies find TLR signalling to be dispensable in the early stages of infection[9], suggesting that mycobacteria have evolved mechanisms to circumvent the bactericidal consequences of TLR signalling.

To explore these mechanisms, we used zebrafish larvae infected with *M. marinum*, a close genetic relative of *M. tuberculosis* and the causative agent of tuberculosis in ectotherms. This model has yielded important insights into the pathogenesis and genetics of human tuberculosis[10]. In humans, the earliest interactions between mycobacteria and phagocytes occur at the lung epithelial surface. Such interactions can be modelled in the larva by injection of bacteria or other chemical stimuli into the hindbrain ventricle (HBV), a neuroepithelium-lined cavity to which phagocytes are recruited[8] (Fig. 1a). We used morpholino knockdown to create zebrafish deficient in MyD88, a common downstream adaptor molecule for TLR signalling pathways[6]. As expected, MyD88 morphants had decreased macrophage recruitment to *Staphylococcus aureus* and *Pseudomonas aeruginosa*, mucosal bacteria that can be commensal or pathogenic[11–13] (Fig. 1b). Similarly, macrophage recruitment to the nonpathogenic *Mycobacterium smegmatis* was MyD88 dependent. In contrast, macrophage recruitment to *M. marinum* was MyD88 independent (Fig. 1c). This finding suggested that pathogenic mycobacteria have the ability to mask PAMPs that would otherwise induce TLR signalling during the initial infection phase. We proposed that such a factor would be a cell-surface-associated virulence determinant. In this light, PDIM seemed a likely candidate, particularly because it is present only in pathogenic mycobacteria, including *M. tuberculosis* and *M. marinum*, but absent in non-pathogenic *M. smegmatis*[2]. We created a *M. marinum* mutant that lacks PDIM on its surface by knocking out the PDIM transporter, encoded by the *mmpL7* gene, and confirmed that it was attenuated in zebrafish larvae (Fig. 1d and Extended Data Fig. 2). If PDIM is masking PAMPs, then macrophage recruitment to Δ*mmpL7* bacteria should be MyD88 dependent, and this was the case (Fig. 1e). In contrast, macrophage migration remained MyD88 independent in response to *M. marinum* deficient in another cell-surface-associated virulence determinant, Erp (Δ*erp*) (Fig. 1d, e and Extended Data Fig. 2)[14]. This result was consistent with *M. smegmatis* possessing a functional *erp*[14], and suggested further that the evasion of MyD88-dependent immune detection was mediated specifically by PDIM.

Our model posits that pathogenic mycobacteria use PDIM to evade recruitment of MyD88-dependent macrophage populations detrimental to their survival. Therefore, we predicted that wild-type mycobacteria would be unaffected in MyD88 morphants, whereas the attenuation of Δ*mmpL7* should be reversed. We found both to be the case (Fig. 1f). For these assays, approximately 80 *M. marinum* were injected into the HBV. However, MyD88 morphants were previously reported to be susceptible to higher *M. marinum* inocula delivered intravenously[15]. We confirmed these findings, showing that MyD88 deficiency increased susceptibility at later time points after intravenous administration of >300 colony forming units (c.f.u.) (Extended Data Fig. 3). It is likely that MyD88 exerts its protective responses at these later stages through mechanisms distinct from the ones we have uncovered, such as through interleukin (IL)-1-mediated responses[9]. Indeed, IL-1 expression was undetectable 3 h after infection, when we observed MyD88-dependent macrophage recruitment (data not shown), suggesting an IL-1-independent role for MyD88 in mediating recruitment towards PDIM-deficient mycobacteria.

Further characterization of wild-type versus PDIM-deficient bacteria revealed that both strains recruited cells expressing the macrophage-specific marker Mpeg1 (ref. 8) (Extended Data Fig. 4a and Supplementary Videos 1, 2). We next asked whether these macrophages possessed differential microbicidal potential. We examined the expression of inducible nitric oxide synthase (iNOS) in these recruited cells because: (1) it is induced in macrophages upon TLR signalling[6], and can be expressed by zebrafish[16], mouse[17] and human[18] macrophages after mycobacterial infection; and (2) mycobacteria are known to be susceptible to reactive

[1]Department of Immunology, University of Washington, Seattle, Washington 98195, USA. [2]Department of Microbiology, University of Washington, Seattle, Washington 98195, USA. [3]Seattle Biomedical Research Institute, Seattle, Washington 98109, USA. [4]Department of Pediatrics, University of Washington, Seattle, Washington 98195, USA. [5]Department of Medicine, University of Washington, Seattle, Washington 98195, USA.

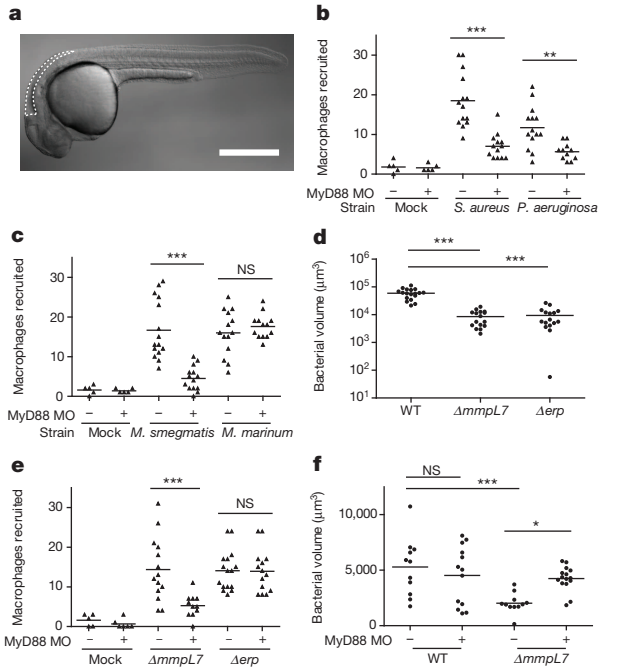



**Figure 1 | PDIM-mediated evasion of MyD88-dependent macrophage recruitment. a**, Schematic of a 2 days post-fertilization (dpf) zebrafish showing the HBV injection site outlined with dashed white line. Scale bar, 500 μm. **b**, **c**, Mean macrophage recruitment at 3 h post-infection (hpi) into the HBV of wild-type or MyD88-morphant (MO) fish after infection with 150 *S. aureus*, 200 *P. aeruginosa* (**b**), 80 *M. marinum* or 85 *M. smegmatis* (**c**). Representative of three separate experiments. **d**, Mean bacterial burdens at 3 dpi after HBV infection of wild-type (WT) fish with 80 wild-type, Δ*mmpL7* or Δ*erp* *M.marinum*. Representative of three separate experiments. **e**, Mean macrophage recruitment at 3 hpi into the HBV of wild-type or MyD88-morphant fish after infection with Δ*mmpL7* or Δ*erp* *M. marinum*. Representative of four separate experiments. **f**, Mean bacterial burdens of wild-type or MyD88-morphant fish at 3 dpi after HBV infection with wild-type or Δ*mmpL7* *M. marinum*. Representative of three separate experiments. Significance testing for all panels done using one-way analysis of variance (ANOVA), with Bonferroni's post-test for comparisons shown. $*P < 0.05$, $***P < 0.001$. NS, not significant.

nitrogen species (RNS) in both murine[17] and human[18] macrophages. We found very few iNOS-positive macrophages arriving in response to wild-type *M. marinum*, whereas the majority of those arriving in response to Δ*mmpL7* bacteria were iNOS positive (Fig. 2a–c and Extended Data Fig. 4b). Δ*erp* bacteria elicited very few iNOS-expressing macrophages (Fig. 2c and Extended Data Fig. 4b), further showing that this early manipulation of macrophage recruitment and/or activation is a specific characteristic of PDIM. We confirmed that RNS were the major mediators of MyD88-dependent macrophage microbicidal activity by showing that the nitric oxide scavenger carboxy-2-phenyl-4,4,5,5-tetramethylimidazolinone-3-oxide-1-oxyl (CPTIO) and $N^G$-nitro-L-arginine methyl ester hydro-chloride (L-NAME ) reversed growth attenuation of the Δ*mmpL7* mutant (Fig. 2d and Extended Data Fig. 4c).

Together, our findings suggested that PDIM mediates an immune evasion strategy, whereby mycobacteria evade detection by TLRs so as to avoid recruitment of iNOS-expressing, microbicidal macrophages. To test this idea, we co-infected red fluorescent wild-type bacteria with green fluorescent wild-type or Δ*mmpL7* bacteria. We found that wild-type bacteria were attenuated in the presence of Δ*mmpL7* bacteria, and that this attenuation transfer was specifically caused by co-infection with Δ*mmpL7* and not with wild-type or Δ*erp* bacteria (Fig. 2e and Extended Data Fig. 5a, b). Furthermore, this transfer of attenuation from Δ*mmpL7* to wild-type bacteria was dependent on macrophages; no attenuation was observed when macrophages were depleted before infection using a morpholino against the myeloid transcription factor PU.1 (Fig. 2f)[16]. Attenuation transfer was similarly dependent on MyD88 signalling, as well as on RNS production (Fig. 2g, h and Extended Data Fig. 5c).

As PDIM is not the only substrate for the MmpL7 transporter, we confirmed that the effects we observed were due to the lack of PDIM *per se* by using a PDIM synthesis mutant, Δ*mas*, showing it to both recruit macrophages in a MyD88-dependent fashion and to transfer attenuation to wild-type bacteria (Extended Data Fig. 6). Finally, to rule out the possibility that the PDIM-deficient mutants simply had increased expression of the culpable PAMP(s), we co-injected heat-killed, crushed wild-type bacteria together with live wild-type bacteria. If the culpable PAMP(s) are expressed by wild-type bacteria, then these PAMPs should become exposed by crushing the bacteria and cause attenuation of the live bacteria. We found this to be the case (Fig. 2i). Altogether, these results suggest that PDIM physically masks underlying mycobacterial PAMPs, thereby preventing mycobacterial delivery into microbicidal macrophages.

To corroborate our findings in a second model, we infected mice through the aerosol route with wild-type *M. tuberculosis* (H37Rv) or with an isogenic strain (Δ*drrA*) defective for proper PDIM surface localization and virulence in mice[3]. At 21 days post-infection (dpi), we found substantially greater proportions of iNOS-producing cells among the $CD11b^+Ly6C^{hi}$ inflammatory monocyte population in the lungs of mice infected with the Δ*drrA* mutant compared to mice infected with the wild-type strain (Fig. 3 and Extended Data Fig. 7). Thus, PDIM-mediated evasion of TLR-dependent immune recognition is shared by *M. tuberculosis* in the context of the mammalian lung, consistent with its central role in avoidance of TLR-dependent antimicrobial mechanisms such as iNOS and antimicrobial peptides[6].

We next sought to understand the mechanism by which mycobacteria recruit the permissive macrophages that are essential for their transport into host tissues. Given our previous finding that *M. marinum* recruits only macrophages (and not neutrophils) to the HBV[8], we considered macrophage-specific chemokines as candidates for mediating this recruitment. We investigated CCR2, which has been implicated in macrophage migration to bacterial pathogens in mice[19], including macrophages that are permissive to *M. tuberculosis* replication after aerosol infection[20]. We identified the functional zebrafish CCR2 orthologue (see Methods) and confirmed that its knockdown resulted in reduced macrophage migration in response to recombinant human chemokine ligand 2 (CCL2) and not to the closely related human macrophage chemokines CCL4 and CCL5 (Extended Data Fig. 8a). The specificity of CCL2-mediated macrophage migration was revealed by the following findings: (1) human and mouse CCL2 induced macrophage but not neutrophil migration (Extended Data Fig. 8b, c); (2) recombinant human IL-8, a neutrophil chemokine, induced neutrophil but not macrophage migration (Extended Data Fig. 8b, c); (3) human leukotriene B4 (LTB4) induced recruitment of both neutrophils and macrophages (Extended Data Fig. 8b, c), as expected[16]; and (4) MyD88 knockdown did not diminish CCL2-mediated macrophage migration, ruling out TLR-mediated migration in response to any endotoxin that might be contaminating the chemokine preparations (Extended Data Fig. 8b).

CCR2 morphants had reduced macrophage migration in response to wild-type *M. marinum*, confirming the role of this pathway in recruitment (Fig. 4a). Recruitment to PDIM-deficient *M. marinum* was unaffected, showing that TLR PAMPs trigger recruitment through a CCR2-independent pathway (Fig. 4a). Accordingly, we found that *M. marinum* infection induced CCL2, and that CCL2 morphants also had reduced macrophage recruitment in response to infection (Fig. 4b and Extended Data Fig. 9).

Turning to the question of which bacterial determinant induced the CCR2 pathway, we considered PGL, a molecule closely related to PDIM in both *M. marinum* and *M. tuberculosis*[2]. Although many clinical *M. tuberculosis* isolates have lost PGL, its presence has been linked to increased virulence[5]. Moreover, among *M. tuberculosis* clinical isolates, PGL expression was linked to *Ccl2* expression in a mouse lung infection model[21]. Similarly, we found that PGL was required for *ccl2* induction in the zebrafish larva; deletion of the *M. marinum pks15* locus

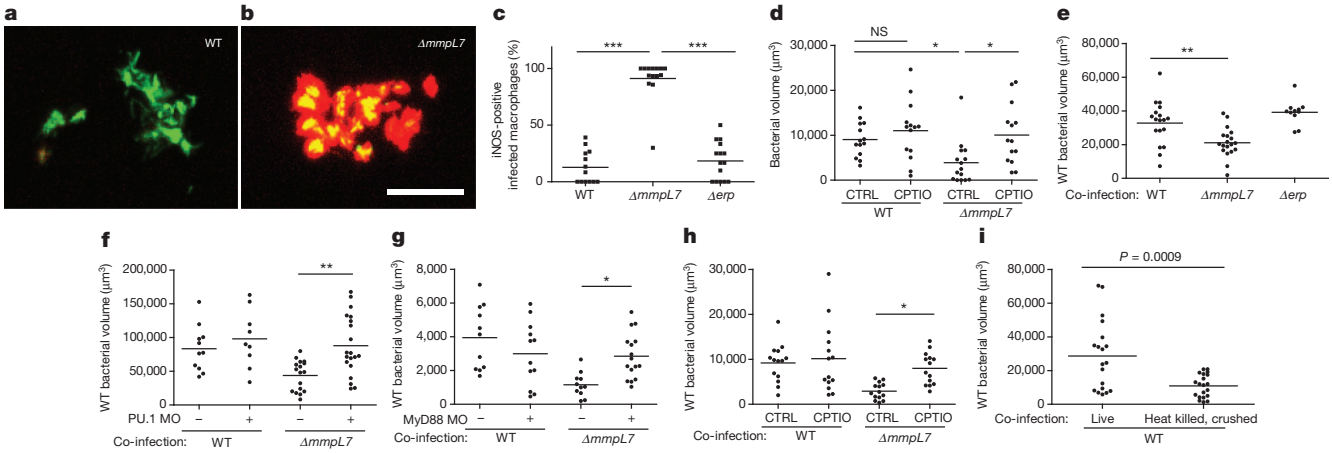



**Figure 2 | Increased iNOS-dependent microbicidal activity of macrophages recruited to PDIM-deficient mycobacteria.** **a**, **b**, Representative images of wild-type (WT) (**a**) and Δ*mmpL7* (**b**) *M. marinum*-infected fish from **c**. $N = 13$ (wild-type) and 14 (Δ*mmpL7*) larvae per group. Scale bar, 50 μm. **c**, Percentage of infected macrophages that were iNOS positive in the HBV at 3 dpi with 80 wild-type, Δ*mmpL7* or Δ*erp M. marinum*. Representative of three separate experiments. **d**, Mean bacterial burdens of 2 dpf control (CTRL)- or RNS scavenger (CPTIO)-treated fish after HBV infection with 80 wild-type or Δ*mmpL7 M. marinum*. Representative of two separate experiments. NS, not significant. **e**–**h**, Mean bacterial volume of red fluorescent wild-type *M. marinum* (infection inoculum 30–40) when co-infected with 30–40 green fluorescent wild-type, Δ*mmpL7* or Δ*erp M. marinum* at 3 dpi in wild-type (**e**), PU.1-morphant (MO) (**f**), MyD88-morphant (**g**), or CPTIO-treated (**h**) larvae. **e**, **g**, Co-infection of wild-type and Δ*mmpL7 M. marinum* in wild-type or MyD88-morphant fish is representative of at least three separate experiments, and co-infection with Δ*erp* is representative of two separate experiments. **f**, **h**, Representative of two separate experiments. **a**–**h**, Significance testing done using one-way ANOVA, with Bonferroni's post-test for comparisons shown. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. **i**, Mean bacterial volume of red fluorescent wild-type *M. marinum* at 3 dpi (infecting inoculum 30–40) when co-infected with the volume equivalent of 30–40 heat-killed, crushed wild-type *M. marinum*. Representative of two separate experiments. Student's unpaired *t*-test.

specifically abrogates PGL, but not PDIM, production (data not shown) and resulted in loss of *ccl2* induction. Δ*mmpL7* bacteria, which lack surface expression of both PGL and PDIM, similarly failed to induce *ccl2*, highlighting that this chemokine is not induced through TLR interactions, but rather is specifically induced through PGL-mediated interactions (Fig. 4b). Furthermore, Δ*pks15* bacteria recruited fewer macrophages upon infection of wild-type larvae, and this reduction was similar to that seen in CCR2 morphants infected with wild-type bacteria (Fig. 4c). There was no additional reduction in recruitment when CCR2 morphants were infected with PGL-deficient bacteria, suggesting that PGL recruits macrophages solely through the CCR2 pathway (Fig. 4c).

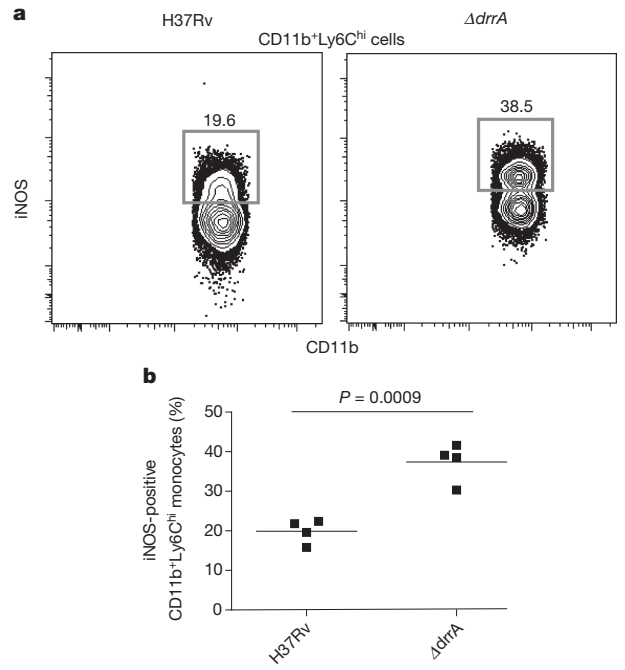



**Figure 3 | Elevated frequencies of iNOS-expressing inflammatory monocytes in mice infected with PDIM-deficient *M. tuberculosis*.** **a**, **b**, C57BL/6 mice were infected through the aerosol route with H37Rv or an isogenic PDIM-deficient mutant (Δ*drrA*). Lung tissue was harvested at 21 dpi and iNOS protein expression was measured using flow cytometry. Representative fluorescence-activated cell sorting (FACS) plots (**a**) and graphical depiction (**b**) of frequencies of iNOS-expressing cells within the $CD11b^+Ly6C^{hi}$ inflammatory monocyte population. Representative of two separate experiments. Student's unpaired *t*-test.

Our findings implicate PGL in bacterial virulence and, correspondingly, the CCR2 pathway in host susceptibility. Globally, a large proportion of *M. tuberculosis* isolates are PGL deficient due to a frameshift in *pks15* (ref. 2). However, the importance of PGL in mediating virulence and/or transmission is underscored by its presence in many of the W-Beijing strains, which are becoming rapidly enriched among *M. tuberculosis* isolates globally[5], and have predominated in outbreaks in North America, where tuberculosis is not prevalent[5]. Infectivity is a key requirement for transmission, and our data suggested that PGL may enhance infectivity through CCR2-mediated recruitment of permissive macrophages at the earliest stages of infection. This enhancement may be particularly relevant in the context of human infections, in which the infectious dose is thought to be as low as 1–3 bacteria[22,23]. To test the hypothesis that PGL enhances infectivity at low doses, we compared the ability of wild-type and PGL-deficient strains to establish infection. Confocal microscopy was used 5 hours after HBV injection to select those animals that had received 1–3 bacteria (Extended Data Fig. 10), and then again at 5 dpi to identify which animals were still infected. We found that 89% of the wild-type but only 18% of the Δ*pks15* infections were successful (Fig. 4d). Concurrent administration of recombinant CCL2 restored the infectivity of Δ*pks15* bacteria, provided the CCR2 pathway was intact (Fig. 4d). Correspondingly, we found that wild-type bacteria had a lower infectivity rate in CCR2 morphants (Fig. 4d). Consistent with our finding that PGL recruits macrophages solely through CCR2, there was no further decrease in infectivity in CCR2 morphants infected with the PGL mutant (Fig. 4d). Finally, the infectivity of wild-type bacteria in MyD88 morphants was undiminished (90% for wild type versus 83% for morphants), consistent with our finding that TLR signalling is not involved in macrophage recruitment to wild-type bacteria.

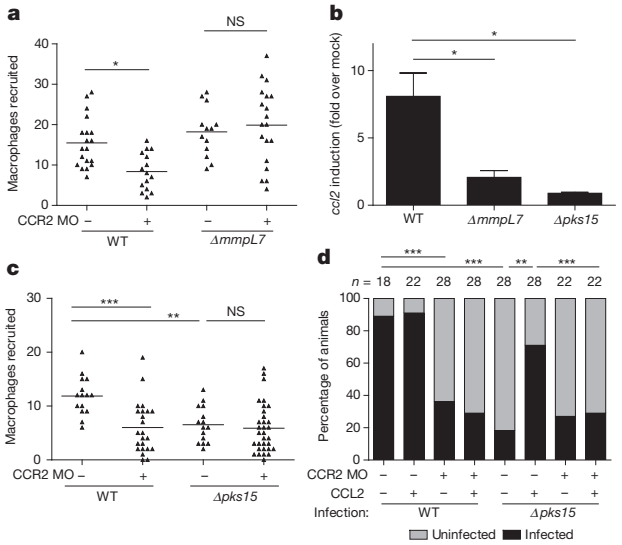



**Figure 4 | Macrophage recruitment and subsequent infectivity is mediated by mycobacterial PGL and host CCR2.** **a**, Mean macrophage recruitment at 3 hpi into the HBV of wild-type or CCR2-morphant (MO) fish after infection with 80 wild-type (WT) or Δ*mmpL7 M. marinum*. Representative of three independent experiments. One-way ANOVA, with Bonferroni's post-test for comparisons shown. $*P < 0.05$. **b**, *ccl2* messenger RNA levels (mean ± standard error of the mean (s.e.m.) of four biological replicates) induced at 3 h after caudal vein infection of 2 dpf larvae with 250–300 wild-type, Δ*pks15* or Δ*mmpL7 M. marinum*. One-way ANOVA with Tukey's post-test. $*P < 0.05$. **c**, Mean macrophage recruitment at 3 hpi into the HBV after infection with 80 wild-type or Δ*pks15 M. marinum*. Representative of three separate experiments. One-way ANOVA with Bonferroni's post-test for comparisons shown. $**P < 0.01$, $***P < 0.001$. **d**, Wild-type and CCR2-morphant fish, with or without the addition of 5 μg ml$^{-1}$ CCL2, were infected in the HBV with 1–3 wild-type or Δ*pks15 M. marinum*. Graph shows the percentage of fish that were infected (black) or uninfected (grey) after 5 days. *n* = number of larvae per group. Representative of two separate experiments. Significance was evaluated using Fisher's exact test for each comparison. $**P < 0.01$, $***P < 0.001$. NS, not significant.

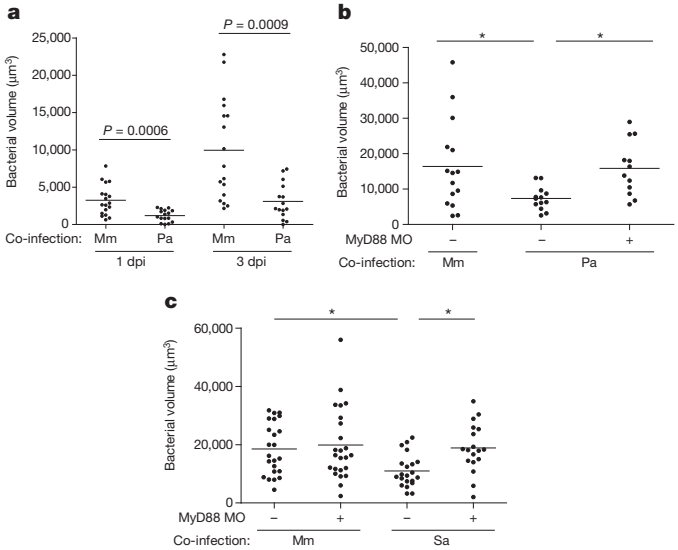



**Figure 5 | MyD88-dependent macrophage recruitment elicited by other bacterial pathogens and commensals attenuates pathogenic mycobacteria.** **a**, Mean bacterial volume of red fluorescent *M. marinum* (infecting inoculum 30–40) after co-infection with either 30–40 green fluorescent *M. marinum* (Mm) or 300 *P. aeruginosa* (Pa) at 1 and 3 dpi. Representative of three separate experiments. Significance assessed using Student's *t*-test. **b**, **c**, Mean bacterial volume of 30–40 red fluorescent wild-type *M. marinum* (infecting inoculum 30–40) after co-infection with either 30–40 green fluorescent wild-type *M. marinum* or 300 *P. aeruginosa* (**b**) or 300 *S. aureus* (Sa) (**c**) at 3 dpi, in wild-type or MyD88-morphant (MO) larvae. Significance tested by one-way ANOVA with Bonferroni's post-test for comparisons shown. $*P < 0.05$.

These findings highlight the interdependency between bacterial PGL and host CCR2 signalling in driving bacterial infectivity under the low inoculum conditions relevant to human infection. Previous investigations into the role of PGL and CCR2 may have failed to reveal these mechanisms because those studies used higher inocula[24] and, in the study of CCR2 signalling, a PGL-deficient strain[25]. Indeed, our finding that CCR2 signalling is a host susceptibility factor is reinforced by human studies showing an association between the high expression of CCL2 and tuberculosis susceptibility[4]. Furthermore, the association appears to be stronger in east Asian populations[26], where clinical isolates are enriched for the predominantly PGL-expressing W-Beijing strains[27]. In light of our findings, we propose that the enrichment of PGL expression among these strains is influencing this association, as the CCR2 pathway would be most relevant in the context of bacterial PGL stimulation.

Finally, our data suggested an explanation for why *M. tuberculosis* must reach the alveolar surfaces of the distal lung in order to initiate infection[23] (Extended Data Fig. 1). It is well established that tuberculosis results from inhalation of small aerosol droplets containing ~1–3 bacteria, which are capable of reaching the alveolar surfaces of the distal lung; in contrast, large droplets harbouring ~$10^4$ bacteria are trapped in the upper bronchial passages and are far less successful at establishing infection[22,23]. These observations have led to the idea that the alveolar surfaces of the distal lung offer a more favourable environment for mycobacterial proliferation. We wondered if commensal microbes from the oropharyngeal surfaces, as well as inhaled environmental organisms, might lead to continual TLR signalling in the upper respiratory tract that would then override the mycobacterial PDIM-dependent immune evasion strategies we identified. In contrast, the lower respiratory tract, which is relatively sterile[28], would favour recruitment of *Mycobacterium*-permissive macrophages. To test this hypothesis, we co-infected animals with *M. marinum* together with bacterial colonizers of the pharynx that induce TLR signalling—either *S. aureus*, a common Gram-positive colonizer of the nasopharynx in both adults and children[12], or the Gram-negative bacterium *P. aeruginosa,* also reported to colonize the pharynx of asymptomatic adults and children[11]. Co-infection with *P. aeruginosa* resulted in the attenuation of wild-type mycobacteria by 1 dpi, and continued into 3 dpi (Fig. 5a). Mycobacterial growth was attenuated despite rapid clearing of *P. aeruginosa*: 56% of the animals had cleared the co-infected *P. aeruginosa* by 1 dpi, and 76% by 3 dpi, with only a few residual bacteria in the remaining animals. Thus, it was not the physical presence of, but rather the detrimental immunological milieu induced by *P. aeruginosa* that was responsible for the attenuation of *M. marinum*. Consistent with our hypothesis, we found that the detrimental effect of *P. aeruginosa* on mycobacterial survival was MyD88 dependent (Fig. 5b). *S. aureus* co-infection also had a MyD88-dependent detrimental effect on *M. marinum* survival (Fig. 5c).

Our previous work identified strategies by which intracellular mycobacteria manipulate host pathways after having traversed epithelial barriers; these involve a bacterial protein secretion system that expands the bacterial niche through macrophage recruitment to the nascent granuloma[10]. We now describe what may be the first contact between mycobacteria and their hosts, and the manner in which mycobacteria manipulate recruitment, and potentially influence the differentiation or activation state of the first responding macrophages so as to gain access to their preferred niche (Extended Data Fig. 1). The choreographed entry involves two related mycobacterial lipids acting in concert to avoid one host pathway while inducing another. Our findings link PDIM, recognized as an absolutely essential mycobacterial virulence factor, to the evasion of TLR detection and thus explain the dispensability of TLR-mediated immunity in protection against *M. tuberculosis* infection in both human and animal studies[9,29]. In contrast, PGL is dispensable for virulence, being variably present among clinical isolates. Yet its presence in the ancestral *M. cannetti* strains as well as in *M. marinum*,

the closest genetic relative of the *M. tuberculosis* complex[2], suggests its integral role in the evolution of mycobacterial pathogenicity. Tuberculosis is an ancient disease, and the enhanced infectivity conferred by PGL may have been essential for most of its history before human crowding, with its greatly increased opportunities for transmission, made it dispensable[30].

Our findings suggest a central role for commensal flora in choreographing mycobacterial entry. Not only must pathogenic mycobacteria possess a physical barrier to prevent host TLR-mediated detection, but they must also evade TLR signalling initiated by other organisms, by entering through the distal lung (Extended Data Fig. 1). Our work may also explain the paradox that smaller *M. tuberculosis* droplets are more infectious than larger ones. However, the requirement placed on mycobacteria to gain entry through the distal lung makes tuberculosis less contagious than most other respiratory infections, thus assigning a protective role to the commensal flora. Conversely, the persistence of human tuberculosis for over 70,000 years[30] attests to the effectiveness of the mycobacterial evolutionary survival kit (masking lipid, recruiting lipid and small infection droplets) to simultaneously evade and manipulate the host and its commensal flora.

## METHODS SUMMARY

*M. marinum*, *S. aureus* and *P. aeruginosa* constitutively expressed fluorescent proteins GFP, Wasabi or tdTomato to allow visualization. Zebrafish larvae (of undetermined sex given the early developmental stages used) were infected at 36–48 h post-fertilization (hpf) via caudal vein or HBV injection. Larvae were randomly allotted to the different experimental conditions. Fluorescence images were captured and quantitative fluorescence was used as a surrogate for bacterial burdens. For the macrophage recruitment assays, macrophages and neutrophils in the HBV were enumerated using differential interference contrast microscopy 3 h after HBV infection. For determination of infection burdens in the HBV, 1 and 3 dpi larvae were mounted in 1.5% agarose (low melting point) and confocal z-stacks of 2 μm were obtained. For the infectivity assay, 2 dpf larvae were injected in the HBV with a concentration of mycobacteria that resulted in an average of 0.8 bacteria per injection. Larvae harbouring 1–3 bacteria were identified at 5 hpi using confocal microscopy, and were re-imaged at 5 dpi and scored as infected or uninfected. Antibody staining for iNOS was performed as described[8] by confocal microscopy. CPTIO (Sigma) was used at a final concentration of 500 μM in 0.1% dimethylsulphoxide in fish water. Larvae were incubated immediately after infection and fresh CPTIO was added every 24 h for the duration of the experiment. For quantitative real-time PCR, complementary DNA was synthesized from pools of 20–40 larvae as previously described[8]. *ccl2* RNA levels were determined using SYBR green and the primers 5′-GTCTGGTGCTCTTCGCTTTC-3′ and 5′-TGCAGAG AAGATGCGTCGTA-3′. Ten-week-old female C57BL/6 mice were infected through the aerosol route with *M. tuberculosis* strains. For iNOS staining, lungs were harvested and processed at 21 dpi. Statistical analyses were formed using Graphpad Prism software. Zebrafish and mouse husbandry and all experiments performed on them were in compliance with Institutional Animal Care and Use Committee approved protocols.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 21 May; accepted 18 October 2013.**
**Published online 15 December 2013.**

1. Philips, J. A. & Ernst, J. D. Tuberculosis pathogenesis and immunity. *Annu. Rev. Pathol.* **7,** 353–384 (2012).
2. Onwueme, K. C., Vos, C. J., Zurita, J., Ferreras, J. A. & Quadri, L. E. N. The dimycocerosate ester polyketide virulence factors of mycobacteria. *Prog. Lipid Res.* **44,** 259–302 (2005).
3. Murry, J. P., Pandey, A. K., Sassetti, C. M. & Rubin, E. J. Phthiocerol dimycocerosate transport is required for resisting interferon-γ-independent immunity. *J. Infect. Dis.* **200,** 774–782 (2009).
4. Flores-Villanueva, P. O. *et al.* A functional promoter polymorphism in monocyte chemoattractant protein-1 is associated with increased susceptibility to pulmonary tuberculosis. *J. Exp. Med.* **202,** 1649–1658 (2005).
5. Reed, M. B. *et al.* A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* **431,** 84–87 (2004).
6. Medzhitov, R. Recognition of microorganisms and activation of the immune response. *Nature* **449,** 819–826 (2007).
7. Weiser, J. N. The pneumococcus: why a commensal misbehaves. *J. Mol. Med.* **88,** 97–102 (2010).
8. Yang, C.-T. *et al.* Neutrophils exert protection in the early tuberculous granuloma by oxidative killing of mycobacteria phagocytosed from infected macrophages. *Cell Host Microbe* **12,** 301–312 (2012).
9. Mayer-Barber, K. D. *et al.* Cutting edge: caspase-1 independent IL-1 production is critical for host resistance to *Mycobacterium tuberculosis* and does not require TLR signaling *in vivo*. *J. Immunol.* **184,** 3326–3330 (2010).
10. Ramakrishnan, L. Revisiting the role of the granuloma in tuberculosis. *Nature Rev. Immunol.* **12,** 352–366 (2012).
11. Rosenthal, S. & Tager, I. B. Prevalence of Gram-negative rods in the normal pharyngeal flora. *Ann. Intern. Med.* **83,** 355–357 (1975).
12. Wertheim, H. F. L. *et al.* The role of nasal carriage in *Staphylococcus aureus* infections. *Lancet Infect. Dis.* **5,** 751–762 (2005).
13. Eddens, T. & Kolls, J. K. Host defenses against bacterial lower respiratory tract infection. *Curr. Opin. Immunol.* **24,** 424–430 (2012).
14. de Mendonça-Lima, L. *et al.* The allele encoding the mycobacterial Erp protein affects lung disease in mice. *Cell. Microbiol.* **5,** 65–73 (2003).
15. van der Vaart, M., van Soest, J. J., Spaink, H. P. & Meijer, A. H. Functional analysis of a zebrafish *myd88* mutant identifies key transcriptional components of the innate immune system. *Dis. Model. Mech.* **6,** 841–854 (2013).
16. Tobin, D. M. *et al.* The *lta4h* locus modulates susceptibility to mycobacterial infection in zebrafish and humans. *Cell* **140,** 717–730 (2010).
17. Chan, J., Xing, Y., Magliozzo, R. S. & Bloom, B. R. Killing of virulent *Mycobacterium tuberculosis* by reactive nitrogen intermediates produced by activated murine macrophages. *J. Exp. Med.* **175,** 1111–1122 (1992).
18. Kröncke, K. D., Fehsel, K. & Kolb-Bachofen, V. Inducible nitric oxide synthase in human diseases. *Clin. Exp. Immunol.* **113,** 147–156 (1998).
19. Serbina, N. V., Jia, T., Hohl, T. M. & Pamer, E. G. Monocyte-mediated defense against microbialpathogens. *Annu. Rev. Immunol.* **26,** 421–452 (2008).
20. Antonelli, L. R. V. *et al.* Intranasal Poly-IC treatment exacerbates tuberculosis in mice through the pulmonary recruitment of a pathogen-permissive monocyte/macrophage population. *J. Clin. Invest.* **120,** 1674–1682 (2010).
21. Ordway, D. *et al.* The hypervirulent *Mycobacterium tuberculosis* strain HN878 induces a potent TH1 response followed by rapid down-regulation. *J. Immunol.* **179,** 522–531 (2007).
22. Bates, J. H., Potts, W. E. & Lewis, M. Epidemiology of primary tuberculosis in an industrial school. *N. Engl. J. Med.* **272,** 714–717 (1965).
23. Wells, W. F., Ratcliffe, H. L. & Grumb, C. On the mechanics of droplet nuclei infection; quantitative experimental air-borne tuberculosis in rabbits. *Am. J. Hyg.* **47,** 11–28 (1948).
24. Sinsimer, D. *et al.* The phenolic glycolipid of *Mycobacterium tuberculosis* differentially modulates the early host cytokine response but does not in itself confer hypervirulence. *Infect. Immun.* **76,** 3027–3036 (2008).
25. Scott, H. M. & Flynn, J. L. *Mycobacterium tuberculosis* in chemokine receptor 2-deficient mice: influence of dose on disease progression. *Infect. Immun.* **70,** 5946–5954 (2002).
26. Feng, W. X. *et al.* CCL2−2518 (A/G) polymorphisms and tuberculosis susceptibility: a meta-analysis. *Int. J. Tuberc. Lung Dis.* **16,** 150–156 (2012).
27. Gagneux, S. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103,** 2869–2873 (2006).
28. Charlson, E. S. *et al.* Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am. J. Respir. Crit. Care Med.* **184,** 957–963 (2011).
29. von Bernuth, H., Picard, C., Puel, A. & Casanova, J.-L. Experimental and natural infections in MyD88- and IRAK-4-deficient mice and humans. *Eur. J. Immunol.* **42,** 3126–3135 (2012).
30. Comas, I. *et al.* Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nature Genet.* **45,** 1176–1182 (2013).

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

**Acknowledgements** We thank S. Falkow and P. Edelstein for sharing their knowledge and insights, P. Donald for discussions about human infectivity in tuberculosis, B. Cormack for manuscript review and editing, J. Bubeck-Wardenberg for the fluorescent *S. aureus* strain, K. Hicks for initial MyD88 experiments, T.-Y. Chen, B. Moody, P. Manzanillo and J. Cox for help with lipid analyses, and J. Cameron for fish facility management. Supported by a National Science Foundation predoctoral fellowship to C.J.C., a Senior Research Training Fellowship from the American Lung Association and the National Institutes of Health (NIH) Training Grant "Training Clinical and Basic Immunologists" to R.P.L., a NIH "Academic Pediatric Infectious Disease" Training Grant award to R.E.H., an American Cancer Society Postdoctoral Fellowship and NIH Bacterial Pathogenesis Training Grant Award to D.M.T., and NIH grants to K.B.U. and L.R. D.M.T. is a recipient of the NIH Director's New Innovator Award and L.R. is a recipient of the NIH Director's Pioneer Award.

**Author Contributions** C.J.C., C.L.C., K.K.T., D.M.T., R.E.H. and L.R. conceived and designed *M. marinum*/zebrafish experiments and analysed data; C.J.C., C.L.C., K.K.T., D.M.T. and R.E.H. performed these experiments; R.P.L. and K.B.U. designed the *M. tuberculosis*/mouse experiments and analysed the data; R.P.L. performed the mouse experiments; C.J.C., C.L.C. and L.R. wrote the paper; C.J.C., R.P.L. and K.K.T. prepared the figures; and all authors edited 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 the paper. Correspondence and requests for materials should be addressed to L.R. (lalitar@uw.edu).

## METHODS

**Bacterial strains and methods.** *M. marinum* strain M (ATCC BAA-535) and the Δ*erp* mutant have been described[31]. The Δ*mmpL7* and Δ*pks15* mutants were generated as described in the following section. Fluorescently labelled bacterial strains were generated by transformation with the pTEC15 or pTEC27 plasmids (deposited with Addgene, plasmids 30174 and 30182, respectively), resulting in *msp12*-driven expression of the Wasabi or tdTomato fluorescent proteins, respectively. Mycobacteria were grown at 33 °C in Middlebrook 7H9 broth or on 7H10 agar (both by Difco) supplemented with 0.5% bovine serum albumin, 0.005% oleic acid, 0.2% glucose, 0.2% glycerol, 0.085% sodium chloride and 0.05% Tween-80 (broth culture only). 50 μg ml$^{-1}$ hygromycin was added as appropriate. For sucrose counter-selection, 7H10 agar was supplemented with 10% sucrose. Single-cell suspensions of bacteria were prepared as described[32]. To prepare heat-killed crushed *M. marinum*, bacteria were incubated at 80 °C for 20 min and then homogenized in a Biospec Bead Beater together with 0.1 mm silica spheres for 1 min. The *P. aeruginosa* PAO1 fluorescent strain used in this study has been described[33]. The *S. aureus* Newman strain expressing pOS1-SdrC-mCherry #391 was a gift from J. Bubeck Wardenburg.

**Targeted deletion of *mmpL7* and *pks15*.** A 2,638 bp PstI fragment containing part of the *mmpL7* (MMAR_1764) open reading frame (ORF) was cloned into a pBluescript-derived vector, pBSXKpn.2 (C.L.C., unpublished observations). A 1,124 bp KpnI fragment internal to *mmpL7* was then excised and replaced with the *aph* cassette, conferring kanamycin resistance. The sucrose counter-selectable marker, *sacB*, and an additional marker, *hygA*, conferring hygromycin resistance, were then added to create pJENK7.1::Hyg. This construct was transformed into the wild-type reference strain, M, and kanamycin-resistant colonies were selected. Subsequent screening for sucrose sensitivity identified merodiploids that were verified by Southern blotting. One such merodiploid was then grown in liquid culture for 10 days and plated on sucrose-containing medium. Sucrose- and kanamycin-resistant hygromycin-sensitive colonies were then verified by Southern blotting to identify the *ΔmmpL7* mutant, KT15. This strain was verified to be deficient in surface localization of PDIM (data not shown), and exhibited colony morphology defects previously reported for *M. marinum* PDIM mutants[34]. The *pks15* (MMAR_1762) locus was deleted as follows. Flanking regions upstream and downstream of *pks15* were amplified by PCR using primers 5′pks15F (5′-CCGCTCG AGGGTGCGATGCGTGGTATC-3′), 5′pks15R.2 and (5′-CGACTAGTTCAGT TGCTCCTGTTCATG-3′), 3′pks15F, and (5′-GGAGCAACTGAACTAGTACC ATCCGACACCGACTG-3′) and 3′pks15R.2 (5′-CCGTCTAGAGTGGTGCTG TTCGGCGTC-3′), respectively. These fragments were sequentially inserted, directly adjacent to each other, into pBluescriptSK+::SacBHyg.1 (C.L.C., unpublished observations), a pBluescript derivative which contains *sacB* and *hygA* external to the multiple cloning site. The resulting construct, pPKS15KO, bears an unmarked deletion of the *pks15* ORF, and was used to transform strain M, and hygromycin-resistant colonies were selected. Putative merodiploids were verified by Southern blotting and then counter-selected on sucrose as described earlier, to produce the sucrose-resistant hygromycin-sensitive isolate (KT21), which was then verified by Southern blotting. Additional verification by thin-layer chromatography determined that PGL was absent, whereas PDIM production was retained (data not shown), consistent with deletion of *pks15* in *M. tuberculosis*[5].

**Zebrafish husbandry and infections.** Wild-type AB zebrafish were maintained as described[35]. Larvae (of undetermined sex given the early developmental stages used) were infected at 36–48 h post-fertilization (hpf) via caudal vein or HBV injection using thawed single-cell suspensions of known titre[32,35]. The number of animals to be used for each experiment was guided by past results with other bacterial mutants and/or zebrafish morphants. Larvae were randomly allotted to the different experimental conditions. Zebrafish husbandry and all experiments performed on them were in compliance with Institutional Animal Care and Use Committee approved protocols.

**Microscopy and image-based quantification of infection level.** Wide-field microscopy was performed using a Nikon Eclipse Ti-E equipped with a C-HGFIE 130W mercury light source, Chroma FITC (41001) filter, and ×2/0.10 Plan Apochromat objective. Fluorescence images were captured with a CoolSNAP HQ2 Monochrome Camera (Photometrics) using NIS-Elements (version 3.22). Quantification of fluorescent *M. marinum* infection using images of individual embryos using Fluorescent Pixel Count (FPC) was performed as previously described[32]. For confocal imaging, larvae were imbedded in 1.5% agarose (low melting point)[36]. A series of z-stack images with a 2 μm step size was generated through the infected HBV, using the galvo scanner (laser scanner) of the Nikon A1 confocal microscope with a ×20 Plan Apo 0.75 NA objective. Bacterial burdens were determined by using the three-dimensional surface-rendering feature of Imaris (Bitplane Scientific Software)[8].

**Hindbrain assays.** Macrophage recruitment assays were performed as previously described[35]. For determination of HBV infection burdens, 1 and 3 dpi larvae were mounted in 1.5% agarose and confocal z-stacks of 2 μm were obtained.

**iNOS staining.** Antibody staining of larvae was performed as described[8,37]. Larvae were then imaged using confocal microscopy and the number of infected macrophages that were positive for iNOS staining was determined for each larva.

**iNOS scavengers.** Fish were treated as previously described[38]. CPTIO or L-NAME (Sigma) were used at a final concentration of 500 μM and 1 mM, respectively, in 0.1% dimethylsulphoxide in fish water. Fish were incubated immediately following infection and fresh inhibitor was added every 24 h until bacterial burden was determined.

**Morpholinos.** The morpholinos described in Supplementary Table 1 were injected at the 1–4-cell stage as previously described[16].

**Reverse transcription PCR to verify efficacy of MyD88 morpholino.** RNA was extracted from pools of 15–40 embryos using TRIzol reagent (Life Technologies), treated with Turbo DNA-Free Kit (Life Technologies) and cDNA synthesized with PrimeScript (Takara). Primers used for PCR were as follows: actin, forward, 5′-ACCTGACAGACTACCTGATG, reverse, 5′-TGAAGGTGGTCTCATGGATAC; *myd88*, forward, 5′-ATGGCATCAAAGTTAAGTATAGACC, reverse, 5′-AGG GCAGTGAGAGTGCTTTG.

**Identification of candidate CCR2 orthologue in zebrafish.** Basic Local Alignment Search Tool (BLAST) searches of the zebrafish genome (http://www.ensembl.org) identified two closely related *CCR*-like genes on chromosome 16: ENSDARG00000079829 and ENSDARG00000062999. In BLAST comparisons to the human genome, ENSDARG00000079829 was found to have the highest homology to human CCR2 (*E* value $8.8 \times 10^{-112}$), whereas ENSDARG00000062999 was most highly homologous to human CCR4 (*E* value $2 \times 10^{-90}$). In addition, annotation of the zebrafish genome from NCBI annotates ENSDARG00000079829 as a *CCR2*-like gene. We confirmed expression of the mRNA and identified the short 5′ upstream exon ATGTCGGCGACACAAAACAGTA using 5′ rapid amplification of cDNA ends (RACE)[39].

**Identification of zebrafish CCL2 orthologue.** Protein sequences of human and mouse CCL2 were used to interrogate the zebrafish genome by BLAST. Expression levels of the four most closely related zebrafish proteins were then examined at 3 hpi to identify the likely functional orthologue (Extended Data Fig. 9a). Of the four candidates, only ENSDARG00000041835 was significantly induced at 3 hpi. Knockdown of ENSDARG00000041835 resulted in a decrease in macrophage recruitment into the HBV at 3 hpi (Extended Data Fig. 9b).

**Quantitative real-time PCR.** cDNA was synthesized from pools of 20–40 larvae as previously described[37]. Quantification of *ccl2* RNA levels was determined using SYBR green and the following primer pair: 5′-GTCTGGTGCTCTTCGCTTTC-3′ and 5′-TGCAGAGAAGATGCGTCGTA-3′.

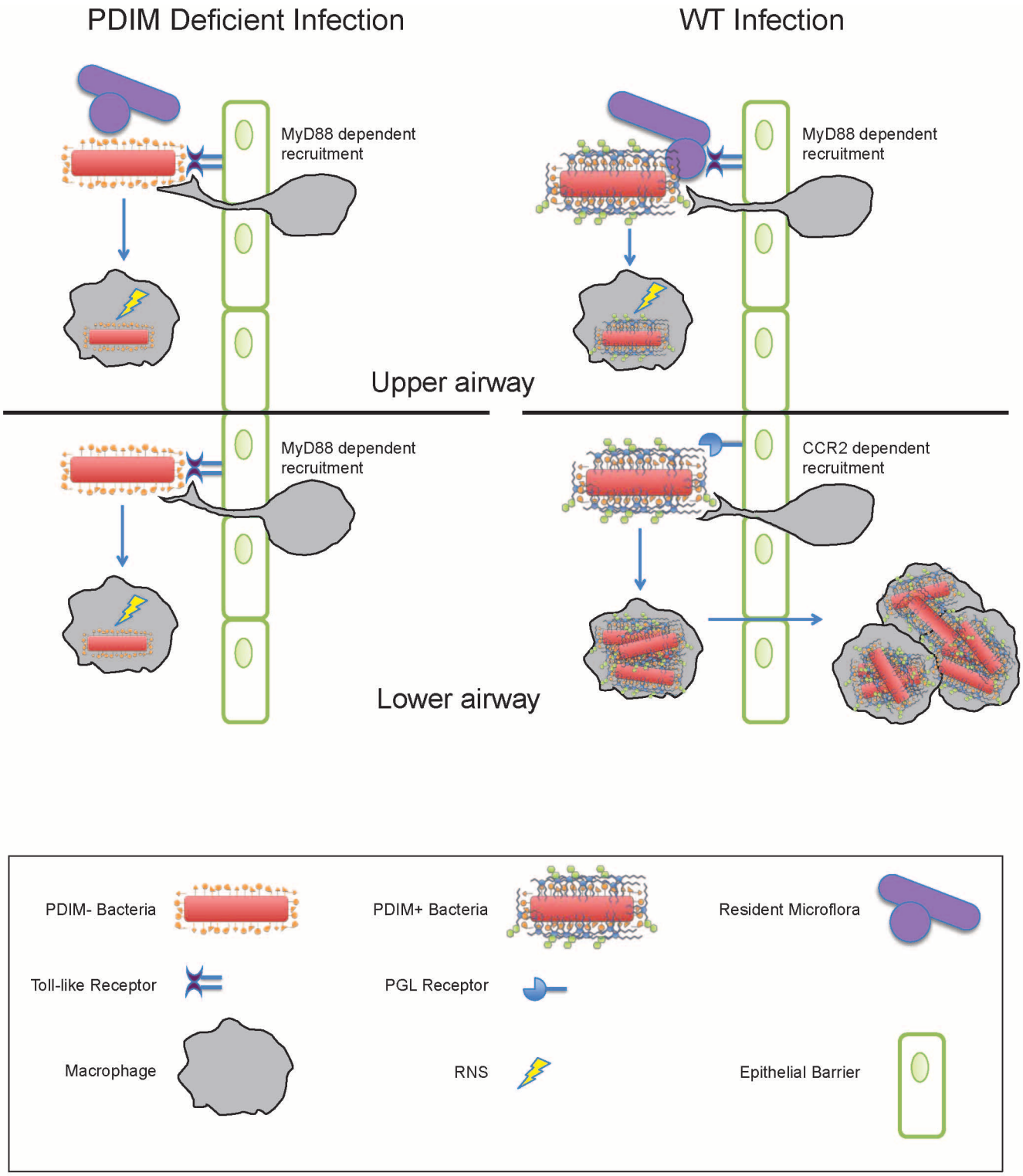
**Infectivity assay.** 2 dpf larvae were infected via the HBV[35] with an average of 0.8 bacteria per injection. Fish harbouring 1–3 bacteria were then identified at 5 hpi by confocal microscopy. These infected fish were then evaluated at 5 dpi and were scored as infected or uninfected, based on the presence or absence of fluorescent bacteria.

**Mice, aerosol infections, and flow cytometry.** C57BL/6 mice were purchased from Jackson Laboratories. All mice were housed under specific pathogen-free conditions at the Seattle Biomedical Research Institute, and all experiments were performed in compliance with the respective Institutional Animal Care and Use Committee approved protocols. Ten-week-old female mice were randomized to the different experimental groups. The number of mice to be used to adequately power the experiment was guided by the results of the corresponding zebrafish experiments. A stock of *M. tuberculosis* strain H37Rv or the isogenic PDIM-deficient *ΔdrrA* strain was sonicated before use and mice were infected in an aerosol infection chamber (Glas-Col) with approximately 200 c.f.u. of H37Rv or 1,000 c.f.u. of *ΔdrrA* to achieve similar bacterial burdens at 21 dpi. The infectious dose in each experiment was determined by plating lung tissue of two mice from each group. Colonies on 7H10 agar plates were counted after 21 days of incubation at 37 °C. Lung tissue was perfused with 5 ml of PBS administered through the right ventricle of the heart, finely chopped using a gentleMACS Octo Dissociator (Miltenyi Biotec) and incubated at 37 °C for 30 min in HEPES buffer containing Liberase Blendzyme 3 (Roche Applied Science). After digestion, single-cell suspensions were prepared by passing tissue through a cell strainer. Single-cell suspensions were then stained for flow cytometric analysis. Lung single-cell suspensions were surface stained at 4 °C for 20 min in the presence of Fc block (24G2) with the following antibodies from eBioscience: PE-Cy7-labelled anti-CD4 (GK1.5, eBioscience), anti-CD8α (53-6.7, eBioscience), anti-CD11c (N418), and FITC-labelled anti-Ly6G (1A8) to exclude T cells, dendritic cells and neutrophils. Alveolar macrophages were excluded based on their high CD11c expression and autofluorescence. PerCPCy5.5-labelled anti-Ly6C (HK1.4) and APC-efluor-780-labelled anti-CD11b (M1/70) were used to identify CD11b$^{+}$Ly6C$^{hi}$ monocytes. Intracellular staining was done after fixation and permeabilization, following the manufacturer's recommendations (eBioscience). Cells were fixed and permeabilized using eBioscience's Fix/Perm buffer for 1 h at 4 °C, followed by staining for iNOS with anti-NOS2 Alexa Fluor 405 (C-11, Santa Cruz Biotechnology) or mouse IgG1 isotype control

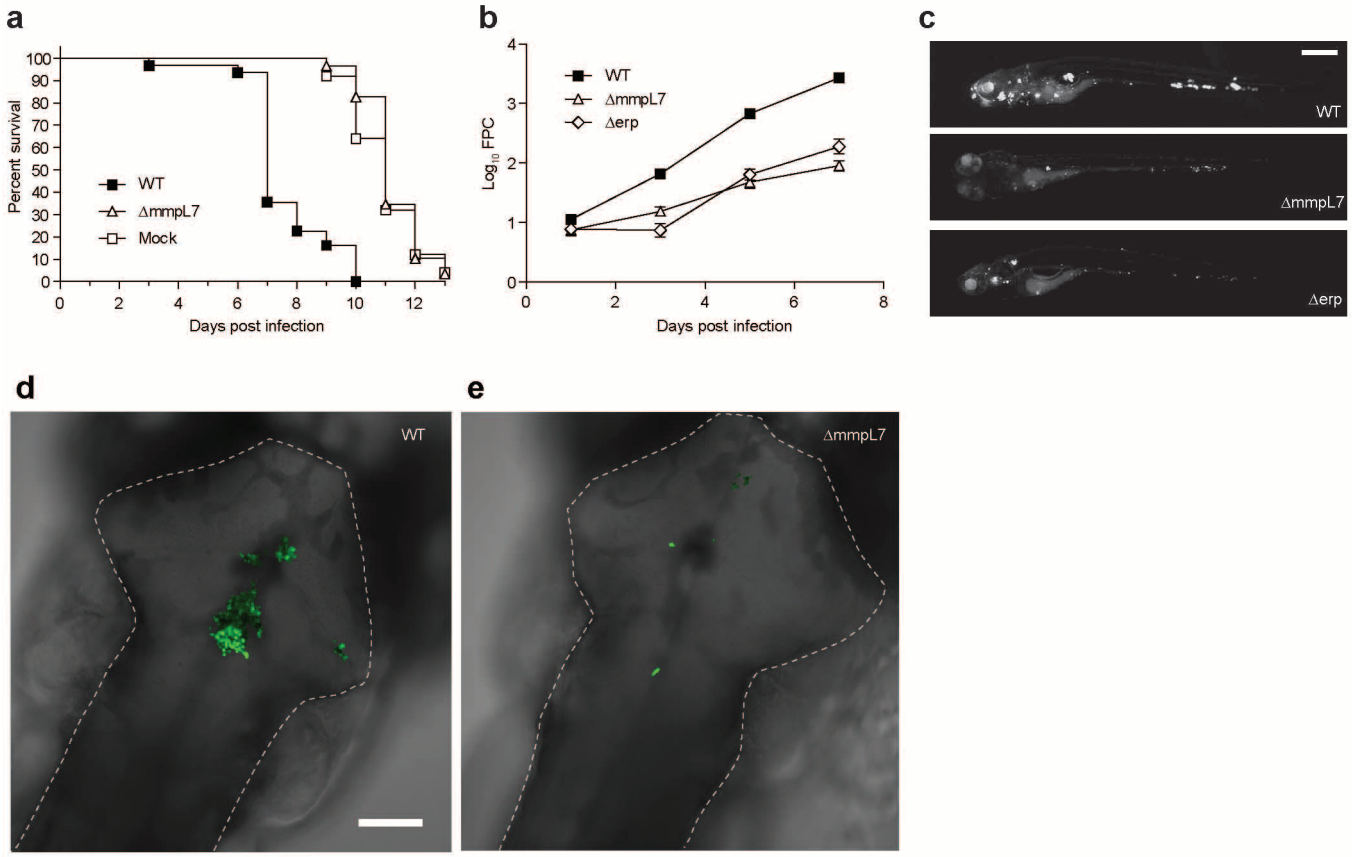
for 30 min at 4 °C. Samples were analysed on an LSR-II (BD Biosciences) and FlowJo Software (Treestar).

**Statistics.** Statistical analyses were performed using Prism 5.01 (GraphPad). For data sets requiring $\log_{10}$ transformation before ANOVA, embryos with no detectable fluorescence above background were assigned a value of 0.9, with 1 being the limit of detection, before $\log_{10}$ transformation. Post-test $P$ values are as follows: $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

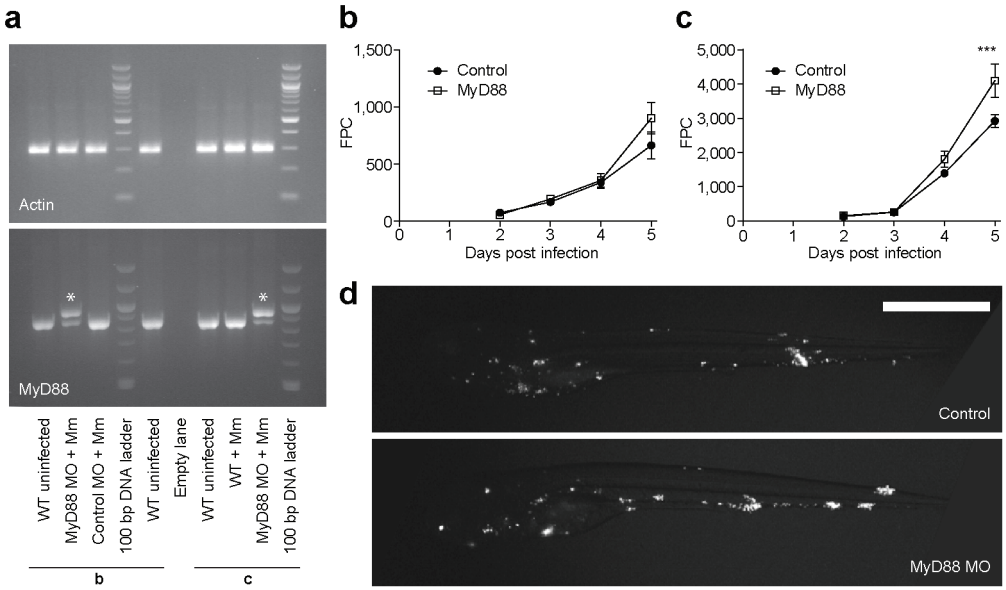

31. Cosma, C. L., Klein, K., Kim, R., Beery, D. & Ramakrishnan, L. *Mycobacterium marinum* Erp is a virulence determinant required for cell wall integrity and ntracellular survival. *Infect. Immun.* **74,** 3125–3133 (2006).
32. Takaki, K., Cosma, C. L., Troll, M. A. & Ramakrishnan, L. An *in vivo* platform for rapid high-throughput antitubercular drug discovery. *Cell Rep.* **2,** 175–184 (2012).
33. Brannon, M. K. *et al. Pseudomonas aeruginosa* Type III secretion system interacts with phagocytes to modulate systemic infection of zebrafish embryos. *Cell. Microbiol.* **11,** 755–768 (2009).
34. Yu, J. *et al.* Both phthiocerol dimycocerosates and phenolic glycolipids are required for virulence of *Mycobacterium marinum*. *Infect. Immun.* **80,** 1381–1389 (2012).
35. Takaki, K., Davis, J. M., Winglee, K. & Ramakrishnan, L. Evaluation of the pathogenesis and treatment of *Mycobacterium marinum* infection in zebrafish. *Nature Protocols* **8,** 1114–1124 (2013).
36. Davis, J. M. & Ramakrishnan, L. The role of the granuloma in expansion and dissemination of early tuberculous infection. *Cell* **136,** 37–49 (2009).
37. Clay, H. *et al.* Dichotomous role of the macrophage in early *Mycobacterium marinum* infection of the zebrafish. *Cell Host Microbe* **2,** 29–39 (2007).
38. Lepiller, S. *et al.* Imaging of nitric oxide in a living vertebrate using a diaminofluorescein probe. *Free Radic. Biol. Med.* **43,** 619–627 (2007).
39. Maruyama, I. N., Rakow, T. L. & Maruyama, H. I. cRACE: a simple method for identification of the 5′ end of mRNAs. *Nucleic Acids Res.* **23,** 3796–3797 (1995).



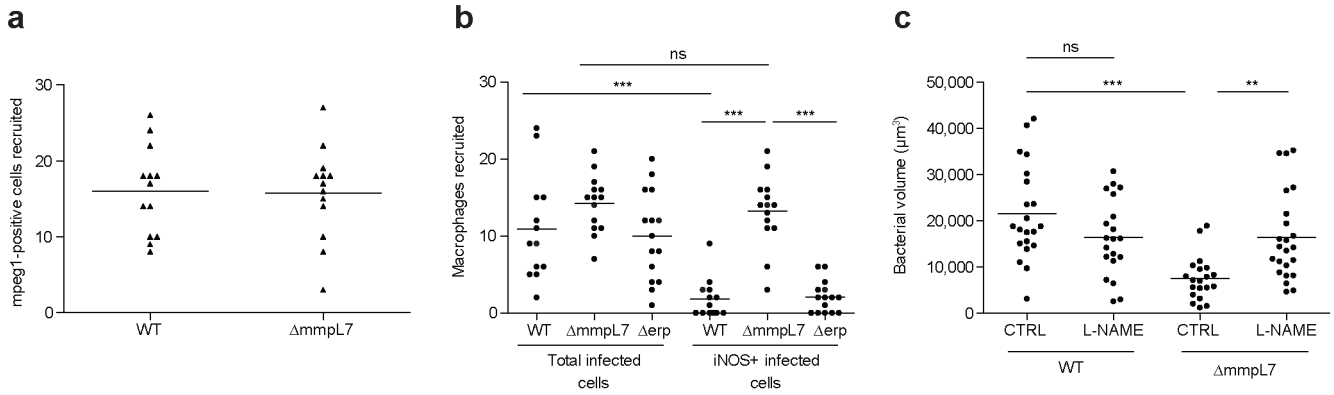

**Extended Data Figure 1 | Coordinate use of PDIM-mediated immune evasion and PGL-mediated recruitment by pathogenic mycobacteria.** Models for infection with wild-type (WT) and PDIM-deficient mycobacteria are shown in the context of the relatively sterile lower airway versus the upper airway, with its higher levels of resident microflora and inhaled environmental organisms.



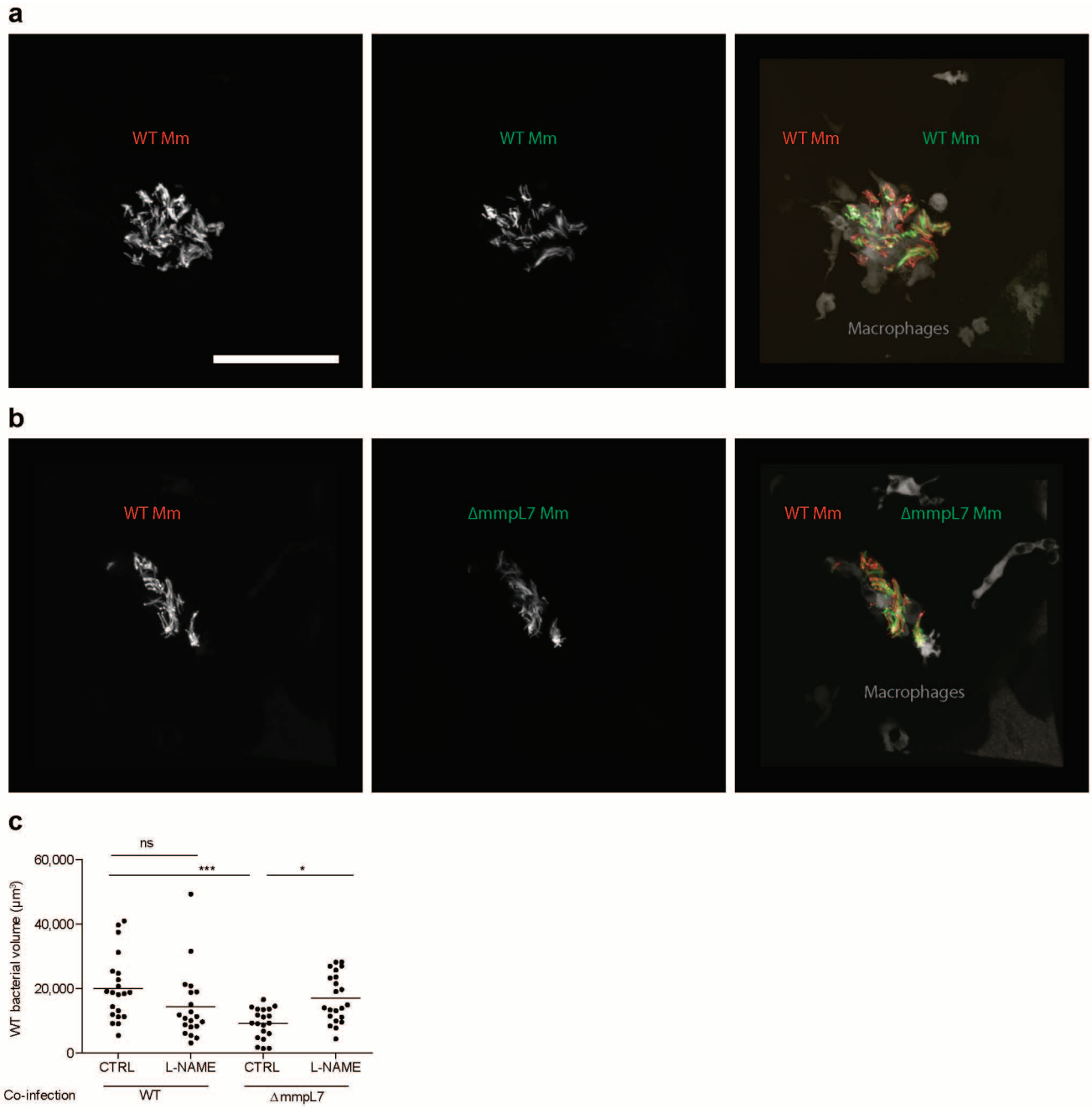

**Extended Data Figure 2 | *ΔmmpL7* bacteria are attenuated in zebrafish larvae.** **a**, Kaplan–Meier graph showing daily survival of larvae infected via caudal vein injection with medium (mock), 29 wild-type or 70 Δ*mmpL7* *M. marinum*. $N = 25$ (mock), 31 (wild-type), or 29 (Δ*mmpL7*) larvae per group. Mean time to death (days): mock (11), wild type (7.6) and Δ*mmpL7* (11.2). Survival was compared by log-rank test: wild type versus mock and wild type versus Δ*mmpL7*, $P < 0.0001$; mock versus Δ*mmpL7*, $P = 0.5601$. **b**, **c**, Larvae were infected via caudal vein injection 1 dpf with 550 wild-type, 650 *ΔmmpL7*, or 700*Δerp*, fluorescent *M. marinum*. **b**, Infection burdens were measured by Fluorescent Pixel Count (FPC; mean ± s.e.m.). **c**, Representative images at 7 dpi. $N = 29$ (wild-type and Δ*mmpL7*) or 30 (Δ*erp*) larvae per group. Scale bar, 500 μm. At 3, 5 and 7 dpi, $\log_{10}$ FPC was compared by ANOVA, with Dunnett's post-test. ***$P < 0.001$. **d**, **e**, Representative images from wild-type (**d**) and Δ*mmpL7* (**e**) *M. marinum* HBV infections quantified in Fig. 1d. $N = 18$ (wild-type) or 16 (Δ*mmpL7*) larvae per group. HBVs are outlined with a dashed white line. Scale bar, 100 μm.



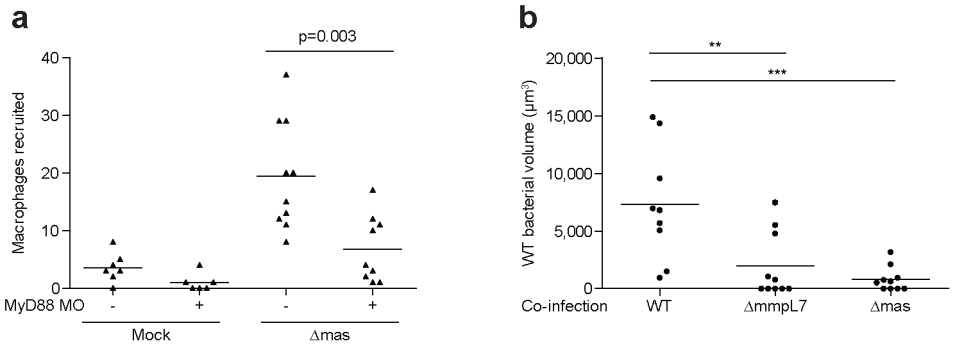

**Extended Data Figure 3 | Knockdown of MyD88 results in a late, dose-dependent hypersusceptibility to *M. marinum* systemic infection.** **a**, RT–PCR for actin (top) and *myd88* (bottom), demonstrating that that the majority of *myd88* transcripts at 7 dpf are abnormal in MyD88 morphants. Lanes marked '**b**' and '**c**' correspond to morphants from the same experiments depicted in panels **b** and **c**, respectively. The abnormal larger transcript (indicated by an asterisk) results from the inclusion of intron 2 in the final transcript, incorporating a premature stop codon that truncates the protein before the TIR (Toll/interleukin receptor) domain. **b**, **c**, Caudal vein infection of MyD88 morphants with 141 (**b**) or 325 (**c**) c.f.u. *M. marinum*/larva. Bacterial burden was assessed by FPC, values plotted represent the mean ± s.e.m. Time points were compared by one-way ANOVA and Bonferroni's post-tests. *** $P < 0.001$. **d**, Representative images of larvae at 5 dpi from experiment in **c**, $N = 30$ control, 15 MyD88 morphant. Scale bar, 500 μm.



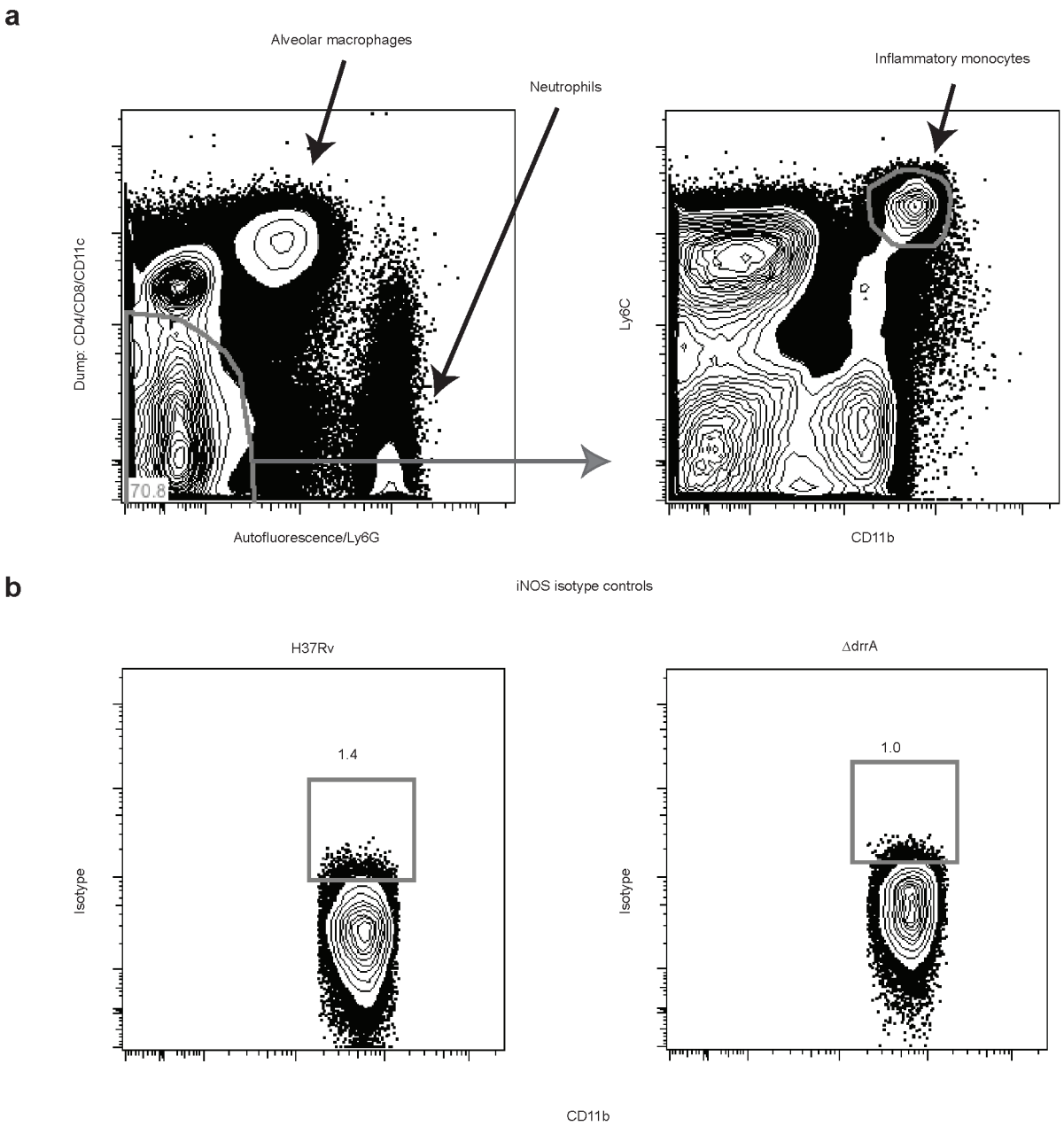

**Extended Data Figure 4 | Characteristics of macrophages recruited to wild-type and PDIM-deficient bacteria.** **a**, Mean Mpeg1-positive macrophages recruited at 3 hpi into the HBV of wild-type fish after infection with 80 wild-type or Δ*mmpL7 M. marinum*. **b**, Data from Fig. 2c expressed as mean numbers of total infected macrophages and iNOS-expressing infected macrophages after HBV infection with 80 wild-type, Δ*mmpL7*, or Δ*erp M. marinum*. **c**, Bacterial burdens after L-NAME treatment. Mean bacterial burdens of 2 dpf control (CTRL)- or iNOS inhibitor (L-NAME)-treated fish after HBV infection with 80 wild-type or Δ*mmpL7 M. marinum*. NS, not significant.



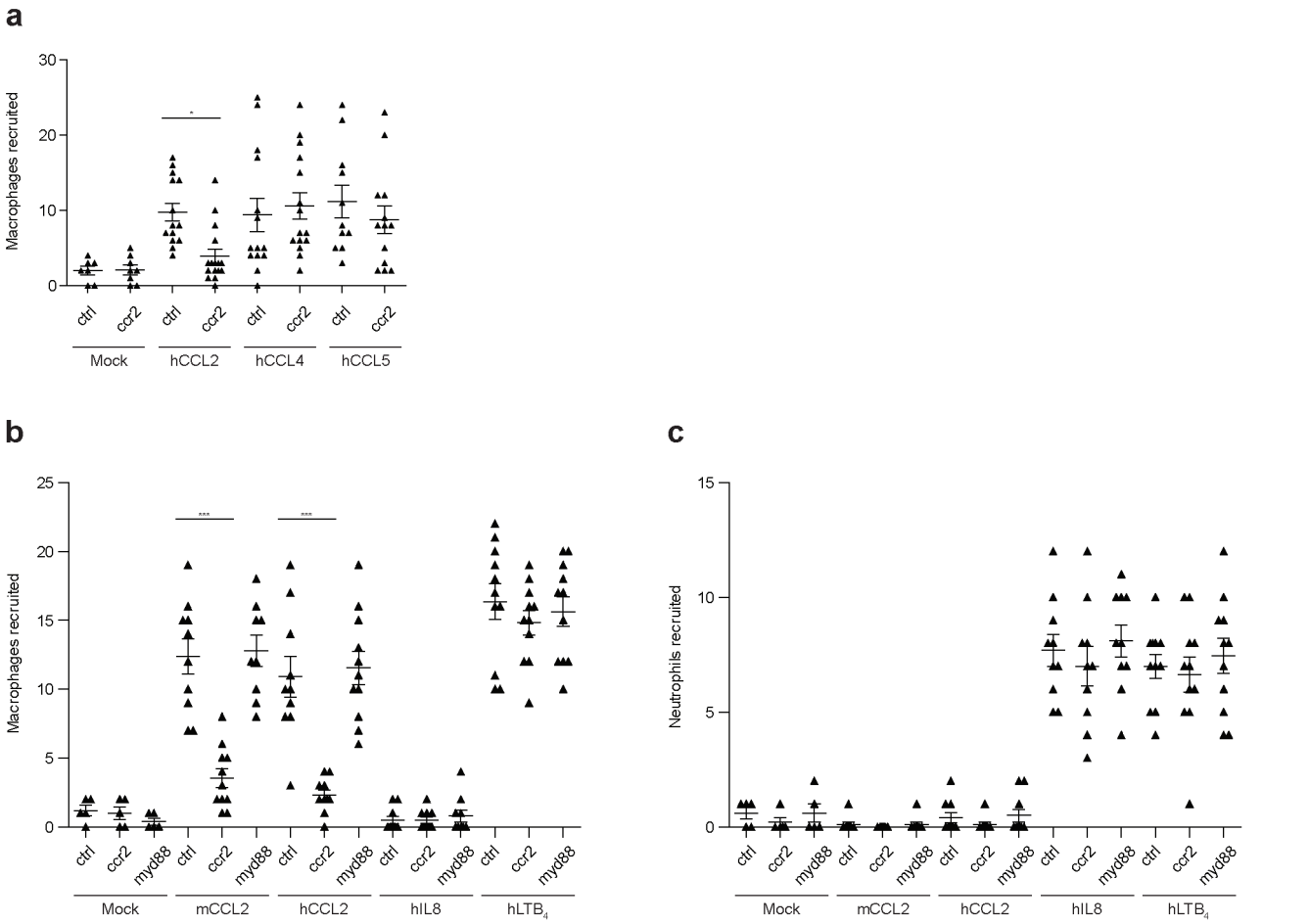

**Extended Data Figure 5 | Wild-type bacterial burdens after co-infection with wild-type or Δ*mmpL7* bacteria.** Representative images from the HBV co-infections quantified in Fig. 2e. **a**, **b**, Red fluorescent wild-type (WT) *M. marinum* co-infected with green fluorescent wild-type (**a**) or Δ*mmpL7* (**b**) *M. marinum*. $N = 18$ (wild-type) and 19 (Δ*mmpL7*) larvae per group. Scale bar, 50 µm. **c**, Wild-type bacterial burdens after co-infection with wild-type or Δ*mmpL7 M. marinum* with and without L-NAME treatment. Significance tested by one-way ANOVA with Bonferroni's post-test for comparisons shown.



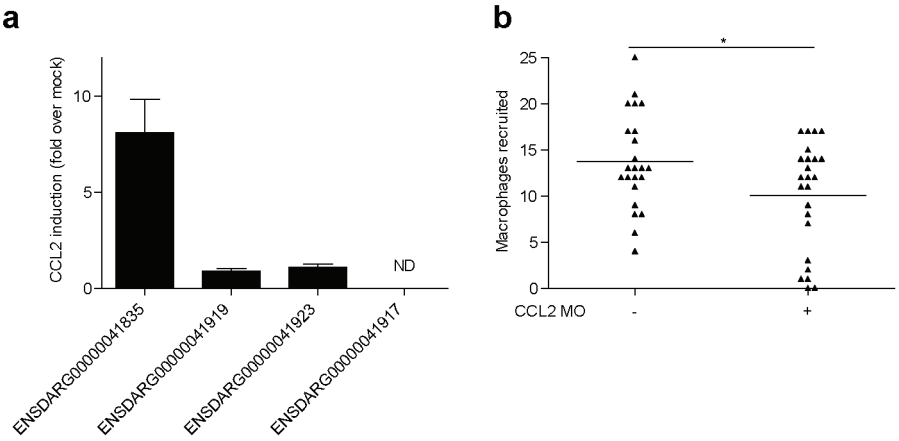

**Extended Data Figure 6 | MyD88-dependent macrophage recruitment occurs in response to PDIM deficiency rather than being due to loss of another MmpL7-exported product. a**, Mean macrophage recruitment at 3 hpi into the HBV of wild-type or MyD88-morphant (MO) larvae after infection with 80 Δ*mas M. marinum*. Student's unpaired *t*-test. **b**, Mean surviving bacterial volume of red fluorescent wild-type *M. marinum* (initial infection dose of 30–40 c.f.u.) when co-infected with 30–40 green fluorescent wild-type, Δ*mmpL7* or Δ*mas M. marinum* at 3 dpi. Representative of two separate experiments. Significance tested by one-way ANOVA with Tukey's post-test.



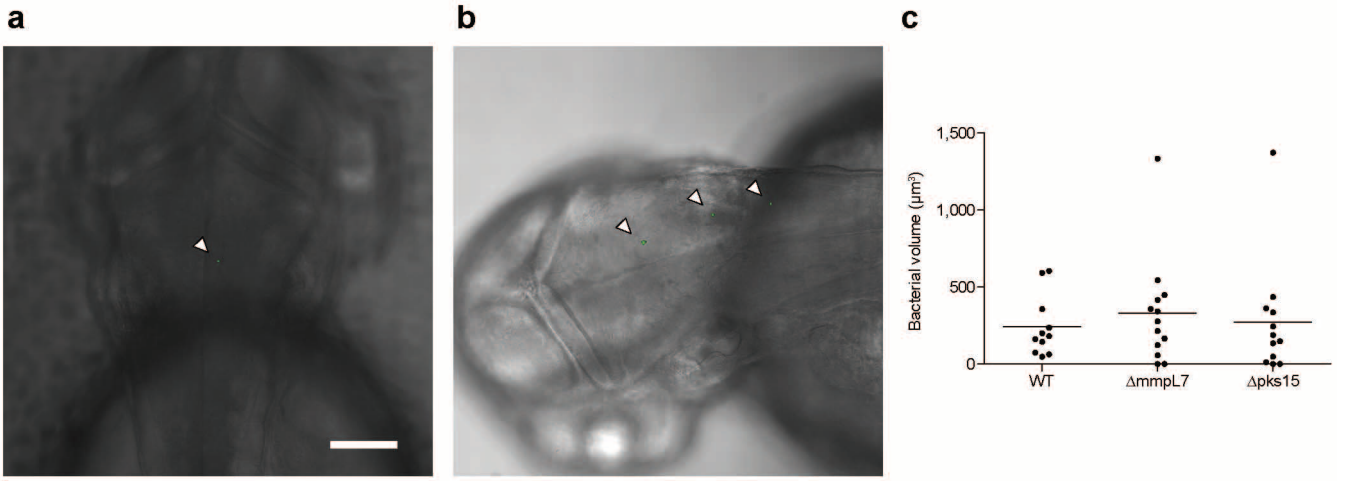

**Extended Data Figure 7 | Gating strategy and isotype controls for iNOS staining of mouse lung.** **a**, Representative gating strategy for isolation of inflammatory monocytes. A dump channel containing anti-CD4, CD8 and CD11c was plotted against a channel exhibiting autofluorescence and also containing anti-Ly6G. Using these markers, T cell, dendritic cell, alveolar macrophage, and neutrophil cell populations were excluded from the double-negative gate. Inflammatory monocytes were identified within the double-negative population by their co-expression of Ly6C and CD11b. These cells were then evaluated for intracellular iNOS expression. **a**, $N = 4$ per group (Fig. 3a, b) or **b**, with isotype control antibodies, $N = 4$ per group.




**Extended Data Figure 8 | Specificity of CCL2-mediated macrophage recruitment in wild-type and CCR2-morphant larvae** **a**, Mean macrophage recruitment at 3 hpi into the HBV of control (ctrl), or CCR2-morphant (CCR2) larvae after injection of vehicle control ('mock'; 0.1% BSA in PBS), human CCL2 (hCCL2), human CCL4 (hCCL4), or human CCL5 (hCCL5). **b**, **c**, Mean macrophage (**b**) and neutrophil (**c**) recruitment at 3 hpi into the HBV of control (CTRL), CCR2-morphant (CCR2), or MyD88-morphant (MyD88) larvae after injection of vehicle control (mock), murine CCL2 (mCCL2), human CCL2 (hCCL2), human IL-8 (hIL-8), or human LTB4 (hLTB4). Representative of three separate experiments. Significance assessed by one-way ANOVA with Bonferroni's post-test for the comparisons shown. $*P < 0.05$; $***P < 0.001$.




**Extended Data Figure 9 | Identification of zebrafish CCL2 orthologue.** **a**, mRNA levels of potential CCL2 orthologues (mean ± s.e.m. of four biological replicates) induced at 3 h after caudal vein infection of 2 dpf larvae with 250–300 wild-type *M. marinum*. These assays were performed on the same cDNA pools as the data presented in Fig. 4b. **b**, Mean macrophage recruitment at 3 hpi into the HBV of wild-type or CCL2 morphant (MO) fish following infection with 80 *M. marinum*. Representative of two separate experiments.




**Extended Data Figure 10 | Infectivity assay.** **a**, **b**, Representative 5 hpi images from Fig. 4d following HBV infection with one (**a**) or three (**b**) *M. marinum*. Scale bar, 100 µm. *N* values for fish represented in **a** and **b** (that is, those found to be infected with 1–3 bacteria) are presented in Fig. 4d (18, 22, 28, 28, 28, 28, 22, 22 for the respective conditions as specified in the figure). **c**, Mean bacterial burdens 5 h after HBV infection with 1–3 wild-type (WT), Δ*mmpL7* or Δ*pks15 M. marinum*.

LETTER

doi:10.1038/nature12808

# Transcranial amelioration of inflammation and cell death after brain injury

Theodore L. Roth[1], Debasis Nayak[1], Tatjana Atanasijevic[1], Alan P. Koretsky[1], Lawrence L. Latour[1] & Dorian B. McGavern[1]

**Traumatic brain injury (TBI) is increasingly appreciated to be highly prevalent and deleterious to neurological function[1,2]. At present, no effective treatment options are available, and little is known about the complex cellular response to TBI during its acute phase. To gain insights into TBI pathogenesis, we developed a novel murine closed-skull brain injury model that mirrors some pathological features associated with mild TBI in humans and used long-term intravital microscopy to study the dynamics of the injury response from its inception. Here we demonstrate that acute brain injury induces vascular damage, meningeal cell death, and the generation of reactive oxygen species (ROS) that ultimately breach the glial limitans and promote spread of the injury into the parenchyma. In response, the brain elicits a neuroprotective, purinergic-receptor-dependent inflammatory response characterized by meningeal neutrophil swarming and microglial reconstitution of the damaged glial limitans. We also show that the skull bone is permeable to small-molecular-weight compounds, and use this delivery route to modulate inflammation and therapeutically ameliorate brain injury through transcranial administration of the ROS scavenger, glutathione. Our results shed light on the acute cellular response to TBI and provide a means to locally deliver therapeutic compounds to the site of injury.**

TBI encompasses injuries that range from mild to severe[1,3], and occurs when the brain is exposed to external forces that induce focal and/or diffuse pathologies, including vascular damage, oedema, axonal shearing and neuronal cell death[4–6]. TBI is usually divided into two phases: the primary insult and the ensuing secondary reaction. It is postulated that primary cell death cannot be prevented without avoiding the injury itself, but that secondary damage is amenable to therapeutic intervention because it is driven by pathogenic parameters such as ROS[7,8], calcium release[9], glutamate toxicity[10,11], mitochondrial dysfunction[12], inflammation[6], and so on. Animal models of TBI have been developed that reflect mild, moderate and severe forms of injury[5], but therapeutic research in these models has not yet translated successfully into the clinic[4,13]. Thus, there is an increasing need to develop additional TBI models, temporally map the dynamics of brain injury responses, and devise therapeutic interventions.

In humans, primary injury to the meninges and vasculature can be observed in the absence of conspicuous brain damage after minor head trauma. As part of an ongoing study of mild TBI, we evaluated research magnetic resonance imaging (MRI) with contrast from patients presenting to the emergency room within 48 h of a minor head injury. Over a period of 30 months, 142 patients were enrolled with a baseline Glasgow Coma Scale of 15, reporting loss of consciousness or post-traumatic amnesia, and a clinical computed tomography (CT) scan without evidence of injury to the parenchyma. Meningeal haemorrhage was seen on CT in 18 patients (12.7%), including subarachnoid blood in 13 (9.1%) and subdural blood in 7 (4.9%). Focal enhancement of the meninges was observed on post-contrast fluid attenuated inversion recovery (FLAIR) MRI imaging (Fig. 1a) in 69 (48.6%) patients, and without concomitant meningeal haemorrhage in 53 (36.9%) patients. Enhancement is the result of extravasation of gadolinium contrast into space containing free fluid with a T1 relaxation time constant equivalent to that of cerebrospinal fluid (CSF)[14].

To understand better the immunopathogenesis of focal brain injury, we developed a novel closed-skull model of mild TBI amenable to intravital imaging studies. Thinning the murine skull bone to ~30 μm allows the underlying meninges and brain parenchyma to be imaged by two-photon laser scanning microscopy (TPM) without overt brain injury or inflammation[15]. Thinning the skull bone beyond 30 μm causes increased pliability and concavity, which compresses the meningeal space (referred to as a compression injury) (Extended Data Fig. 1). Sequential thinning of the skull bone from 50 to 10 μm induced increasing amounts of meningeal cell death (Fig. 1b). Cell death and inflammation associated with over-thinning was reproducibly generated by quickly thinning the skull bone to ~20–30 μm and then manually promoting concavity with minimal downward pressure (Extended Data Fig. 1). We used this model to define the dynamics of inflammation and the mechanisms that cause cell death after focal TBI.

Using TPM we first mapped the kinetics and severity of brain pathology, starting 5 min after compression injury. Immediately after injury quantum dots injected intravenously leaked from vessels into the subarachnoid and perivascular spaces (Fig. 1c and Supplementary Video 1). Within 30 min, ROS were detected in the meninges (Fig. 1d) and holes appeared in the glial limitans due to astrocyte cell death (Fig. 1e and Supplementary Video 1). Transcranially administered SR101, a 600 molecular weight (MW) dye, leaked through the glial limitans into the parenchyma after compression injury, but remained largely within the meningeal space after standard skull thinning (Fig. 1f). Compression also induced cell death in the meninges that increased steadily over time, but was not observed in the parenchyma until 9–12 h after injury (Fig. 1g, h). Parenchymal cell death at 12 h was indiscriminate, as neurons, astrocytes, oligodendrocytes and microglia were all lost in the lesion site (Extended Data Fig. 2).

We next sought insights into the dynamics of the innate inflammatory response. Meningeal macrophages (long, rod-like cells) died within 30 min of compression injury (Fig. 2a and Supplementary Video 1). In response to meningeal cell death, microglia extended processes through the compromised glial limitans into the meninges (Extended Data Fig. 3a and Supplementary Video 1). We also observed a coordinated microglial response to compression injury. Most microglia within 50 μm of the meninges retracted all processes except for ~2–3 that extended towards the glial limitans, forming a stable contiguous network resembling a 'honeycomb' structure (Fig. 2b and Supplementary Video 2). Long-term TPM revealed that the honeycomb network formed within an hour of injury and could be maintained for up to 12 h (Supplementary Video 2). Honeycomb microglia surrounded surviving astrocytes in the glial limitans and aligned with the junctions between individual cells (Extended Data Fig. 3b and Supplementary Video 2).

In response to astrocyte death (Fig. 1d and Supplementary Video 1), a morphologically distinct microglial reaction was observed; microglia retracted their ramified processes and extended a single, non-branching process towards the glial limitans that resembled a jellyfish (Fig. 2c, Extended Data Fig. 3b and Supplementary Video 3). 'Jellyfish' microglia

[1]National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA.

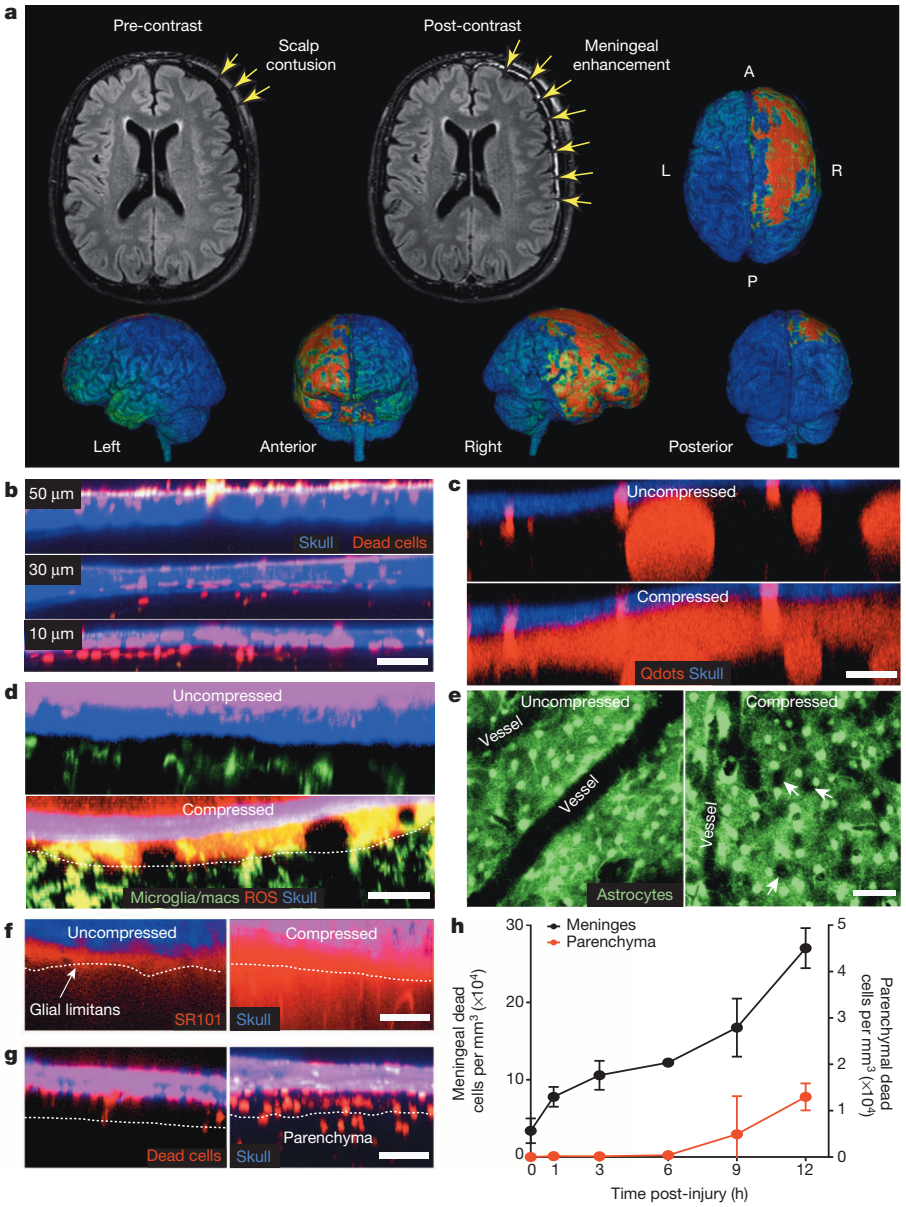



**Figure 1 | Pathology associated with compression injury.** **a**, MRI of a patient's brain 19 h after a fall from 1.8 m, with reported loss of consciousness, post-traumatic amnesia, a Glasgow Coma Scale of 15 on arrival, and a negative CT scan. After administration of gadolinium diethylenetriaminepentacetate (Gd-DTPA) contrast agent, FLAIR MRI (greyscale images) revealed focal enhancement in the meninges along the convexity underlying the area of blunt trauma, better visualized on surface-rendered three-dimensional FLAIR (pseudo-coloured three-dimensional images), involving the frontal and temporal lobes as well as anterior aspects of the cerebral falx. **b**–**d**, **f**, **g**, Maximum projections (5 µm wide) are shown in the *xz* plane of two-photon z-stacks captured through a thinned murine skull. **b**, Images of skull bone (blue) and underlying meninges show sequential skull thinning from 50 to 10 µm. Dead cells (red) were labelled by transcranial propidium iodide (PI) administration. Scale bar, 50 µm. **c**, Intravenously injected Q-dots (red) leak from blood vessels into the meninges 15 min after compression injury, indicative of vascular damage. Scale bar, 50 µm. **d**, ROS (red) labelled with Amplex Red appear in the meninges 30 min after compression injury of CX3CR1$^{gfp/+}$ (green) mice relative to uncompressed controls. White dotted line indicates the glial limitans. macs, macrophages. Scale bar, 30 µm. **e**, *xy* maximal projections (25 µm in depth) captured in GFAP–GFP mice show holes (white arrows) in the glial limitans as a result of astrocyte (green) death, which starts to occur 5 min after compression injury. Scale bar, 50 µm. **f**, Transcranially applied SR101 (red) diffuses into the brain parenchyma 30 min after injury, but is largely excluded from the parenchyma in an uncompressed control mouse. Scale bar, 50 µm. **g**, Cell death (PI$^+$ cells; red) becomes apparent in the brain parenchyma 12 h after compression injury and is not observed in uncompressed controls. Scale bar, 50 µm. **h**, Quantification of cell death (mean + standard deviation (s.d.)) in the meninges and parenchyma after compression injury. All data in the figure are representative of three mice per group (four mice, **g**) and at least three independent experiments.

started to form almost immediately after compression injury (Supplementary Videos 3, 4), and some were motile, whereas others remained stationary (Supplementary Videos 3–5). We commonly observed honeycomb networks of microglia interspersed with clusters of jellyfish microglia (Supplementary Videos 4, 5), which probably reflects variation in lesion severity along the glial limitans. Honeycomb microglia could even transform within 5 min into jellyfish microglia, presumably after astrocyte cell death (Supplementary Video 5), whereas naive microglia required ~30 min to acquire a jellyfish morphology. Jellyfish projections were usually linked to cell bodies via thin processes (Supplementary Video 6) and often formed a continuous phagocytic layer at the glial limitans (Fig. 2c, Extended Data Fig. 3b and Supplementary Videos 3, 6, 7). Over time, microglia residing in the glial limitans died, particularly after tissue swelling (or oedema) was observed (Supplementary Video 7). Peripherally derived myelomonocytic cells (neutrophils and monocytes) also responded to brain damage. Within an hour of compression injury,

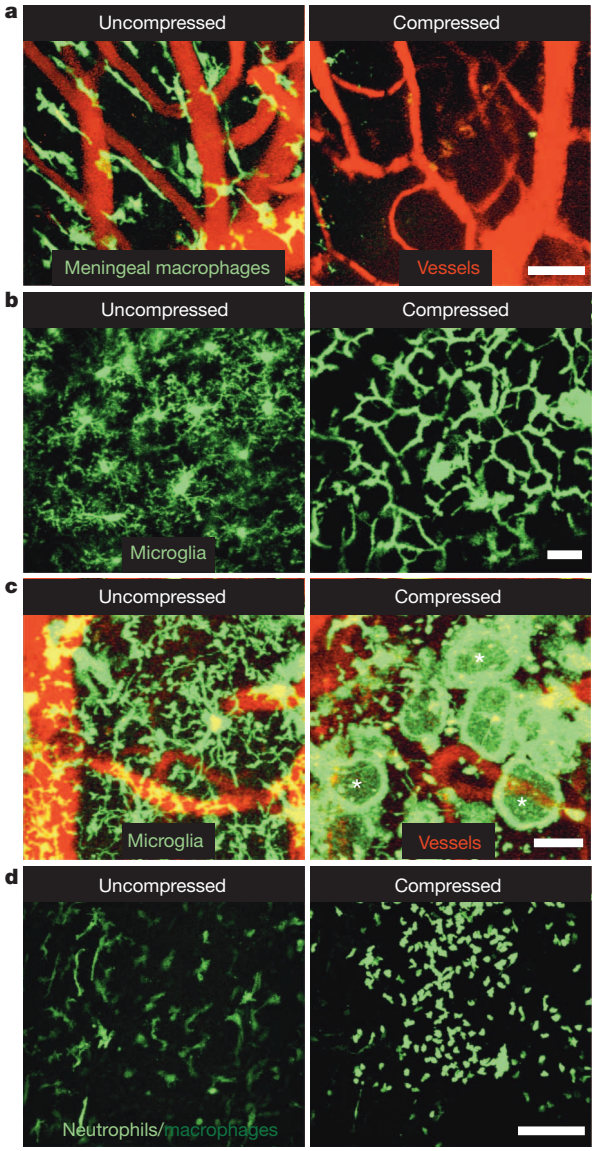



**Figure 2 | Innate immune response to a compression injury.** **a**–**d**, Twenty-five micrometre *xy* maximum projections from CX3CR1$^{gfp/+}$ mice (**a**–**c**) or LysM$^{gfp/+}$ mice (**d**) captured at 30 min (**a**), 1 h (**b**), 2 h (**c**) or 6 h (**d**) in a normal thinned skull preparation (uncompressed) or after a compression injury. **a**, Meningeal macrophages (green) visualized in CX3CR1$^{gfp/+}$ mice burst and die within 30 min of compression injury relative to uncompressed controls. Blood vessels are red. Scale bar, 50 µm. **b**, Microglia (green) retract their ramified processes and form a highly connected honeycomb network at the glial limitans after compression injury. Scale bar, 25 µm. **c**, A subset of microglia (green) after compression injury retract ramified processes and generate a single, flat, motile, phagocytotic process at the base of the glial limitans, resembling a jellyfish (examples denoted with white asterisks). Blood vessels are red. Scale bar, 25 µm. **d**, LysM$^{gfp/+}$ neutrophils (green) are recruited to the site of injury, but not to an uncompressed thinned skull window. Green cells residing in the uncompressed window represent meningeal macrophages. Scale bar, 100 µm. Data are representative of three mice per group and at least three independent experiments.

myelomonocytic cells (probably neutrophils) localized exclusively to the meninges, were highly motile, and interacted with dead cells during the 12 h observation period (Fig. 2d and Supplementary Video 8).

To modulate TBI lesions locally, we applied compounds to the intact skull bone. We discovered that SR101, when applied to an intact (non-thinned) skull bone passed directly into the meninges within 10 min (referred to as a 'transcranial application') (Fig. 3a). We next tested a range of differently sized fluorescent dextrans (3,000–70,000 MW). Dextrans of 40,000 MW and below were able to pass through the intact skull into the meninges (Fig. 3b), although larger dextrans required longer diffusion times (Fig. 3c). A 70,000 MW dextran was unable to pass transcranially in 30 min. In addition, a variety of fluorescent small molecules and macromolecules passed through the murine skull bone and achieved measurable steady-state concentrations in the meninges dependent on molecular weight (Fig. 3d and Extended Data Fig. 4). Passage through an intact skull yielded a meningeal concentration approximately one half that achieved by thinned skull application (Extended Data Fig. 4d, e). We assessed the feasibility of passing compounds through thicker skulls by applying the contrast agent manganese chloride to an

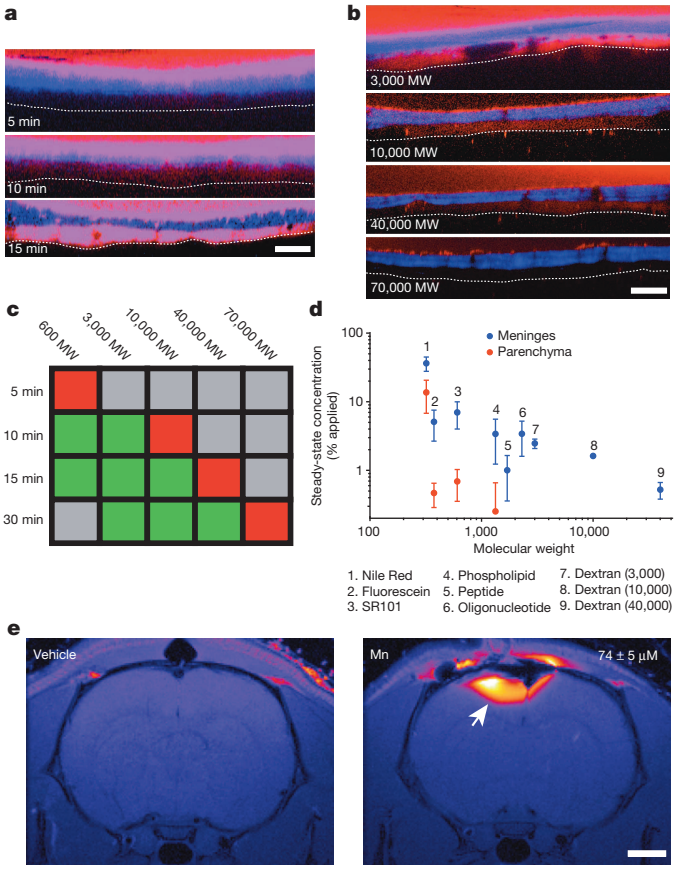



**Figure 3 | Metrics of transcranial diffusion through the skull bone.** **a**, A 600 MW fluorescent dye, SR101 (red), was applied to an intact mouse skull for the indicated time and then the skull (blue) was quickly thinned and imaged. Five micrometre *xz* maximum projections show that SR101 is detectable in the meninges beginning 10 min after application and fully saturates the space within 15 min. White dotted line indicates glial limitans. Scale bar, 50 µm. **b**, The size dependence of diffusion through an intact skull bone was evaluated 30 min after continuous transcranial application of the indicated molecular weight dextrans (red). Dextrans that passed successfully through the skull generated fluorescence in the meninges. The skull bone is shown in blue, and the glial limitans is denoted with a white dotted line. Scale bar, 50 µm. **c**, A colour-coded table summarizing the imaging results shown in panels **a** and **b** denotes the presence (green) or absence (red) of fluorescent dye in the meninges at the indicated molecular weight and time. Grey, not tested. **d**, Fluorescent compounds of increasing molecular weights were passed transcranially through a thinned skull window during imaging. Steady-state concentrations (mean + s.d.) of the fluorescent compounds in the meninges and parenchyma were quantified from normalized fluorescence intensities. See also Extended Data Fig. 4. **e**, Manganese chloride (Mn; 500 mM solution) applied transcranially to an intact rat skull (~1 mm thick) is visible by MRI in the brain parenchyma 2 h after application (white arrow). The mean parenchymal manganese concentration ± s.d. is provided. Scale bar, 1 mm. All data in the figure are representative of three mice (or rats) per group and at least three independent experiments.

intact rat skull bone (~1 mm thick) and imaging transcranial passage by MRI (Fig. 3e). Transcranially applied manganese chloride was clearly visible in the rat brain parenchyma 2 h after application.

We next defined the mechanisms underlying compression injury-induced inflammation. Trancranial application of purinergic receptor[16] (P2RY12 or P2RX4) inhibitors before compression injury prevented both honeycomb and jellyfish morphologies, whereas P2RY6 antagonism only blocked the jellyfish response (Fig. 4a, b, Extended Data Fig. 5a and Supplementary Video 9). In contrast, P2RX7 antagonism had no effect on microglia, but almost entirely eliminated neutrophil recruitment (Fig. 4a–c, Extended Data Fig. 5b and Supplementary Video 9). Astrocytes are known to amplify purinergic receptor signalling through ATP-induced ATP release via connexin hemichannels, which can be blocked with carbenoxolone (CBX)[17]. Transcranial application of CBX, but not a specific pannexin inhibitor (probenecid), before compression injury caused microglia to remain ramified, extending only small, ill-defined circular processes at the glial limitans, similar to what was observed after P2RY6 antagonism (Fig. 4d, e, Extended Data Fig. 5c and Supplementary Video 9). CBX inhibited the formation of honeycomb and jellyfish microglia, whereas pannexin inhibition slowed the onset and magnitude of neutrophil recruitment (Fig. 4f). Pre-treatment with CBX also significantly increased SR101 leakage through the glial limitans, suggesting that ATP release by astrocytes and purinergic signalling in microglia help maintain barrier integrity between the meninges and parenchyma (Fig. 4g and Extended Data Fig. 5d).

To identify the primary mediator of cell death after compression injury, we focused on the role of ROS, which appeared in the meninges shortly after compression injury (Fig. 1d). Transcranial administration of the ROS scavenger glutathione (GSH) resulted in near complete survival of meningeal macrophages after a more severe injury (that is, skull fracture) (Fig. 5a and Supplementary Video 10) as well as glial limitans preservation (Fig. 5b, f). In addition, microglia beneath this layer remained in a non-reactive, ramified state (Fig. 5a, e and Supplementary Video 10). Preservation of the glial limitans after GSH treatment resulted in refilling of the subarachnoid space beneath the compression injury, which pushed the thinned skull bone upward (Supplementary Video 10). GSH administration also eliminated the recruitment of myelomonocytic cells (Fig. 5c, g and Supplementary Video 10).

Cell death was first observed in the meninges and later spread to the parenchyma after compression injury (Fig. 1h). Transcranial pre-treatment with GSH resulted in a 50% reduction in meningeal death, but administration after injury had no effect (Fig. 5h), indicating that half the initial meningeal cell death following compression injury is due to ROS. GSH, when applied continuously starting at 15 min or 3 h after injury, reduced parenchymal cell death at 12 h by 67% and 51%, respectively (Fig. 5d, i). These data indicate that ROS are a mediator of cell death after compression injury. The contribution of inflammation to cell death was assessed by transcranially inhibiting neutrophils and microglia with P2RX7 or P2RY6 antagonism, respectively. Inhibition of neutrophil recruitment through P2RX7 antagonism increased cell death in the meninges 12 h later, but had no impact on parenchymal cell death (Fig. 5h, i). Conversely, inhibition of microglia through P2RY6 antagonism increased parenchymal cell death at 12 h, but did not affect meningeal cell death (Fig. 5h, i). These data suggest that inflammation is neuroprotective within the first 12 h of compression injury.

TBI induces a complex reaction that can result in permanent damage and neurological dysfunction. In this study, we observed evidence of meningeal damage in ~50% of patients with mild head injury, indicating that this is a common pathology in humans. We sought mechanistic insights into this process by developing a novel closed-skull model of brain injury and imaging the acute cellular injury response from its inception. Importantly, we discovered that the skull bone is porous and permits the passage of small molecules (≤40,000 MW) and contrast agents by passive diffusion, which should facilitate local delivery

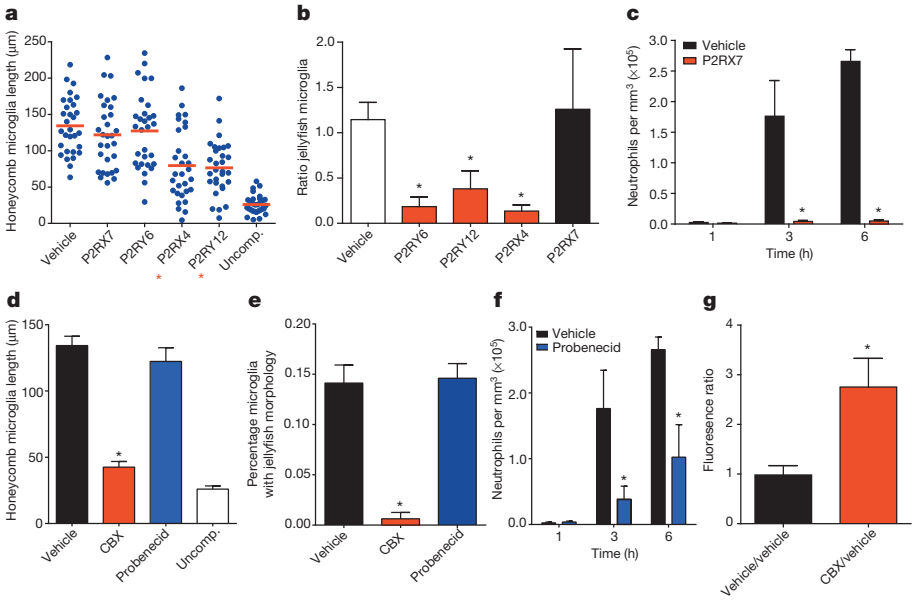



**Figure 4 | Purinergic receptor signalling mediates the innate immune response to compression injury.** **a**, Honeycomb microglia were quantified 3 h after compression injury in mice treated transcranially with P2RX7, P2RY6, P2RX4 and P2RY12 antagonists or vehicle. Uncompressed mice (Uncomp.) served as a negative control. Blue dots represent individual microglia, and the horizontal red line denotes the mean. **b**, Quantification of jellyfish microglia was performed 3 h after compression injury. Because two compression injuries were generated per mouse, data are represented as a ratio (mean ± s.d.) of purinergic receptor antagonist/vehicle compared with vehicle/vehicle. A ratio of one signifies no difference between the two hemispheres. **c**, Quantification of neutrophils per $mm^3$ tissue (mean ± s.d.) was performed at 1, 3 and 6 h after compression injury. **d**–**f**, Bar graphs (mean ± s.d.) show quantification of $CX3CR1^{gfp/+}$ microglia with a honeycomb (**d**) or jellyfish (**e**) morphology as well as the number of $LysM^{gfp/+}$ neutrophils (**f**) after transcranial administration of CBX or probenecid. **f**, Glial limitans permeability was quantified by generating two compression injuries per mouse. The mean SR101 fluorescence in the parenchyma beneath each injury was calculated and expressed as a ratio (CBX/vehicle or vehicle/vehicle). A ratio larger than one signifies increased permeability in the experimental group. Asterisks in all panels denote statistical significance ($P < 0.05$) relative to the vehicle control group. Data are representative of three mice per group and at least three independent experiments.

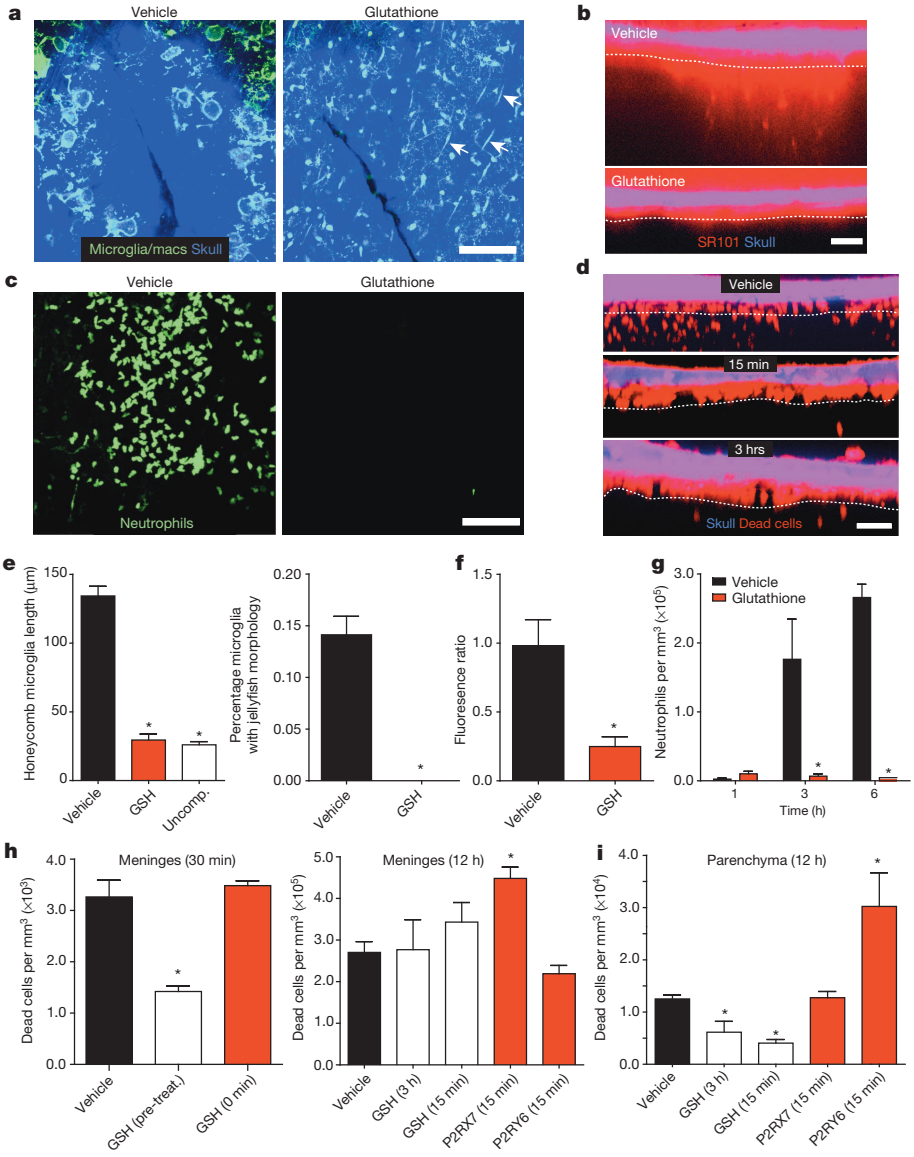



**Figure 5 | Transcranial administration of glutathione reduces inflammation and cell death after compression injury.** **a**–**i**, Twenty-five micrometre *xy* maximum projections (**a**, **c**) and 5 μm *xz* projections (**b**, **d**) were captured in CX3CR1$^{gfp/+}$ (**a**), B6 (**b**, **d**) or LysM$^{gfp/+}$ (**c**) mice after compression injury ($n = 3$, **a**–**c**, **e**–**g**; $n = 4$, **d**, **h**–**i**). **a**, GSH pre-treatment prevented jellyfish and honeycomb microglia formation (green) 1 h after a compression injury that resulted in a cracked skull (blue). GSH administration also promoted survival of meningeal macrophages (macs: green; white arrows). Scale bar, 100 μm. **b**, Relative to the vehicle control group, GSH pre-treatment prevented glial limitans breakdown observed 1 h after injury. SR101 (red) localizes above the glial limitans (indicated by a white dotted line) in the GSH-treated group. Scale bar, 50 μm. **c**, In GSH pre-treated mice, no neutrophil response (green) to compression injury was observed at 6 h. Scale bar, 100 μm. **d**, GSH administered 15 min or 3 h after a compression injury significantly reduced parenchymal cell death observed at 12 h. PI$^+$ dead cells (red) reside primarily in the meninges of GSH-treated mice. White dotted line indicates glial limitans. Scale bar, 50 μm. **e**–**g**, Bar graphs (mean ± s.d.) show quantification of honeycomb and jellyfish microglia (**e**), glial limitans permeability (**f**) and neutrophil recruitment (**g**) in vehicle- versus GSH-treated mice after compression injury at the time points denoted above. Uncomp., uncompressed skulls. **h**, The number of PI$^+$ dead cells per mm$^3$ (mean ± s.d.) was quantified in the meninges at 30 min or 12 h after compression injury. GSH significantly reduced meningeal cell death if applied before (pre-treatment (pre-treat.)), but not after (0 min, 15 min, 3 h) compression injury. P2RX7 blockade increased meningeal cell death when administered 15 min after compression injury. **i**, Parenchymal cell death was quantified 12 h after compression injury in the denoted groups. All data are representative of three independent experiments, and asterisks denote a statistically significant difference ($P < 0.05$) from the vehicle control group.

of therapeutics and other molecules into the central nervous system. Pathologically, compression injury initially caused meningeal cell death, vascular damage, ROS generation, and disruption of the glial limitans, which ultimately gave rise to indiscriminate parenchymal cell death. ROS are commonly observed in TBI lesions[7,8], and transcranial delivery of GSH preserved the glial limitans, reduced cell death, and eliminated the sterile injury response. GSH significantly reduced parenchymal cell death even when administered 3 h after injury, providing a therapeutic window for treatment of focal brain injury.

In the absence of GSH, the brain responded to damage by eliciting an anatomically partitioned sterile immune reaction[18]. Microglia first fortified the glial limitans through the generation of honeycomb and jellyfish structures. Honeycomb microglial networks circumscribed individual surviving astrocytes in the glial limitans and were induced to do so by the release of ATP from connexin hemichannels and its detection by purinergic receptors (P2RX4 and P2RY12)[16,17,19,20]. Phagocytic jellyfish microglia were similarly generated by purinergic signalling (P2RX4, P2RY6, P2RY12) and filled areas along the glial limitans in which astrocyte cell death had occurred. Transcranial inhibition of the microglial response through P2RY6 or connexin hemichannel antagonism increased the permeability of the glial limitans and parenchymal cell death after compression injury. Although microglia protected the parenchyma, myelomonocytic cells invaded the damaged meninges in a P2RX7- and pannexin-dependent manner, consistent with a recent study showing P2RX7-dependent neutrophil recruitment into the injured liver[21]. Collectively, our data suggest that the acute inflammatory reaction to brain injury is beneficial[6]. Moreover, ROS and purines represent major drivers of the injury response and are amenable to transcranial therapeutic manipulation.

## METHODS SUMMARY

Human patients presenting to the emergency room within 48 h of mild head injury were evaluated as part of an ongoing Traumatic Head Injury Neuroimaging Classification (THINC) study. MRI scans were obtained and evaluated after injection of a gadolinium-based contrast agent. To model the pathology of mild head injury, we surgically thinned the murine skull bone over the barrel cortex to a thickness of ~20–30 μm and then applied minimal downward pressure to promote concavity in the bone (meningeal compression) (Extended Data Fig. 1). As a control, a similar surgical procedure was conducted without applying downward pressure. All intravital imaging was performed using a Leica SP5 two-photon microscope and subsequent image analysis was performed using Imaris 7.0 software. Transcranial administration of compounds was achieved by placing compounds resuspended in artificial cerebral spinal fluid (aCSF) directly on top of the exposed skull bone (either thinned or completely intact). Compounds entered the CNS via passive diffusion through the bone. For these experiments, rodents were subsequently imaged using a Leica two-photon microscope or an 11.7-Tesla MRI scanner. Immunohistochemical stains of TBI lesions in mice were imaged using an Olympus FV1200 confocal microscope.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 27 April; accepted 21 October 2013.**
**Published online 8 December 2013.**

1. DeKosky, S. T., Blennow, K., Ikonomovic, M. D. & Gandy, S. Acute and chronic traumatic encephalopathies: pathogenesis and biomarkers. *Nature Rev. Neurol.* **9,** 192–200 (2013).
2. Roozenbeek, B., Maas, A. I. & Menon, D. K. Changing patterns in the epidemiology of traumatic brain injury. *Nature Rev. Neurol.* **9,** 231–236 (2013).
3. Zetterberg, H., Smith, D. H. & Blennow, K. Biomarkers of mild traumatic brain injury in cerebrospinal fluid and blood. *Nature Rev. Neurol.* **9,** 201–210 (2013).
4. McConeghy, K. W., Hatton, J., Hughes, L. & Cook, A. M. A review of neuroprotection pharmacology and therapies in patients with acute traumatic brain injury. *CNS Drugs* **26,** 613–636 (2012).
5. Xiong, Y., Mahmood, A. & Chopp, M. Animal models of traumatic brain injury. *Nature Rev. Neurosci.* **14,** 128–142 (2013).
6. Finnie, J. W. Neuroinflammation: beneficial and detrimental effects after traumatic brain injury. *Inflammopharmacology* **21,** 309–320 (2013).
7. Hall, E. D., Andrus, P. K. & Yonkers, P. A. Brain hydroxyl radical generation in acute experimental head injury. *J. Neurochem.* **60,** 588–594 (1993).
8. Lewén, A., Matz, P. & Chan, P. H. Free radical pathways in CNS injury. *J. Neurotrauma* **17,** 871–890 (2000).
9. Young, W. Role of calcium in central nervous system injuries. *J. Neurotrauma* **9** (suppl. 1), S9–S25 (1992).
10. Bullock, R. *et al.* Evidence for prolonged release of excitatory amino acids in severe human head trauma. Relationship to clinical events. *Ann. NY Acad. Sci.* **765,** 290–297 (1995).
11. Bullock, R. *et al.* Factors affecting excitatory amino acid release following severe human head injury. *J. Neurosurg.* **89,** 507–518 (1998).
12. Mazzeo, A. T., Beat, A., Singh, A. & Bullock, M. R. The role of mitochondrial transition pore, and its modulation, in traumatic brain injury and delayed neurodegeneration after TBI. *Exp. Neurol.* **218,** 363–370 (2009).
13. Janowitz, T. & Menon, D. K. Exploring new routes for neuroprotective drug development in traumatic brain injury. *Sci. Transl. Med.* **2,** 27rv1 (2010).
14. Mamourian, A. C., Hoopes, P. J. & Lewis, L. D. Visualization of intravenously administered contrast material in the CSF on fluid-attenuated inversion-recovery MR images: an *in vitro* and animal-model investigation. *AJNR Am. J. Neuroradiol.* **21,** 105–111 (2000).
15. Xu, H. T., Pan, F., Yang, G. & Gan, W. B. Choice of cranial window type for *in vivo* imaging affects dendritic spine turnover in the cortex. *Nature Neurosci.* **10,** 549–551 (2007).
16. Eltzschig, H. K., Sitkovsky, M. V. & Robson, S. C. Purinergic signaling during inflammation. *N. Engl. J. Med.* **367,** 2322–2333 (2012).
17. Davalos, D. *et al.* ATP mediates rapid microglial response to local brain injury *in vivo*. *Nature Neurosci.* **8,** 752–758 (2005).
18. Chen, G. Y. & Nunez, G. Sterile inflammation: sensing and reacting to damage. *Nature Rev. Immunol.* **10,** 826–837 (2010).
19. Nimmerjahn, A., Kirchhoff, F. & Helmchen, F. Resting microglial cells are highly dynamic surveillants of brain parenchyma *in vivo*. *Science* **308,** 1314–1318 (2005).
20. Haynes, S. E. *et al.* The P2Y12 receptor regulates microglial activation by extracellular nucleotides. *Nature Neurosci.* **9,** 1512–1519 (2006).
21. McDonald, B. *et al.* Intravascular danger signals guide neutrophils to sites of sterile inflammation. *Science* **330,** 362–366 (2010).

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

**Acknowledgements** The study was supported by the National Institutes of Health (NIH), the National Institute of Neurological Disorders and Stroke (NINDS), and the Center for Neuroscience and Regenerative Medicine (CNRM) at the Uniformed Services University of the Health Sciences—a collaborative effort between the NIH, the Department of Defense and the Walter Reed National Military Medical Center to develop innovative approaches for brain injury diagnosis and recovery.

**Author Contributions** T.L.R., D.N. and D.B.M. designed all murine experiments and interpreted the data. L.L.L. provided and interpreted the human TBI data. T.A. and A.P.K. contributed the transcranial manganese data. T.L.R. and D.B.M. 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 the paper. Correspondence and requests for materials should be addressed to D.B.M. (mcgavernd@mail.nih.gov).

LETTER

doi:10.1038/nature12884

# Structural basis for hijacking CBF-β and CUL5 E3 ligase complex by HIV-1 Vif

Yingying Guo[1]*, Liyong Dong[1]*, Xiaolin Qiu[1]*, Yishu Wang[1], Bailing Zhang[1], Hongnan Liu[1], You Yu[2], Yi Zang[1], Maojun Yang[2] & Zhiwei Huang[1]

**The human immunodeficiency virus (HIV)-1 protein Vif has a central role in the neutralization of host innate defences by hijacking cellular proteasomal degradation pathways to subvert the antiviral activity of host restriction factors[1–6]; however, the underlying mechanism by which Vif achieves this remains unclear. Here we report a crystal structure of the Vif–CBF-β–CUL5–ELOB–ELOC complex. The structure reveals that Vif, by means of two domains, organizes formation of the pentameric complex by interacting with CBF-β, CUL5 and ELOC. The larger domain (α/β domain) of Vif binds to the same side of CBF-β as RUNX1, indicating that Vif and RUNX1 are exclusive for CBF-β binding. Interactions of the smaller domain (α-domain) of Vif with ELOC and CUL5 are cooperative and mimic those of SOCS2 with the latter two proteins. A unique zinc-finger motif of Vif, which is located between the two Vif domains, makes no contacts with the other proteins but stabilizes the conformation of the α-domain, which may be important for Vif–CUL5 interaction. Together, our data reveal the structural basis for Vif hijacking of the CBF-β and CUL5 E3 ligase complex, laying a foundation for rational design of novel anti-HIV drugs.**

Human primary cells express restriction factors including APOBEC3 family members to block the replication and spread of HIV, a group of obligatory intracellular retroviruses[7–11]. One common strategy used by HIV-1 for its replication in host cells is to hijack cellular proteasomal degradation pathways to degrade the host restriction factors. A critical HIV protein involved in this process is HIV-1 virion infectivity factor (Vif) that is expressed in most lentiviruses[1–6]. Interaction with Vif results in recruitment of APOBEC3G to an E3 ubiquitin ligase complex containing the scaffold protein cullin5 (CUL5) and substrate adaptors elongin B (ELOB) and elongin C (ELOC), promoting APOBEC3G polyubiquitination and degradation and thereby damping APOBEC3G-mediated cellular defences[1–6]. A conserved Vif motif, called BC-box (residues 144–155), is required for Vif interaction with ELOB–ELOC through mimicking a conserved cellular SOCS-box motif of the SOCS-box proteins[12]. The HCCH (His 108, Cys 114, Cys 133 and His 139) motif of Vif, a unique zinc finger motif, is also important for Vif binding to CUL5, Vif-mediated degradation of APOBEC3G and HIV-1 infectivity[13]. Another host protein, core-binding factor subunit beta (CBF-β), was recently shown[14,15] to be simultaneously hijacked by HIV-1 Vif to form the Vif–CBF-β–CUL5–ELOB–ELOC complex. Vif is considered a good target for anti-HIV drugs because of its essential roles in HIV-1 infection; however, the structure of HIV-1 Vif alone or in the context of functional complexes is lacking.

To facilitate structural study, we reconstituted a Vif (residues 1–192)–CBF-β (residues 1–170)–CUL5 (residues 12–386, nCUL5)–ELOB (residues 1–102)–ELOC (residues 17–112) complex using purified proteins (Fig. 1a and Extended Data Fig. 1a). The absence of Vif–CBF-β reduced the interaction between the nCUL5 fragment and the ELOC–ELOB complex (Fig. 1a), indicating that the former two proteins have a critical role in promoting assembly of the pentameric complex. This is consistent with the observation[16] that SOCS2 is important for CUL5 interaction with ELOC–ELOB. The pentameric complex was crystallized and its structure was determined using molecular replacement (Extended Data Table 1 and Extended Data Figs 1b and 2). The overall complex structure has a U-shaped architecture, with nCUL5 and CBF-β–Vif corresponding to the two straight arms (Fig. 1b). Interaction of ELOC with nCUL5 and Vif forms the bent arm of the U-shaped structure. The Vif structure can be divided into a larger and a smaller domain, with a zinc ion binding between them (Fig. 1b). One side of CBF-β forms extensive contacts with the larger domain of Vif, whereas the carboxy-terminal peptide of CBF-β is sandwiched between the two domains of Vif. Vif also interacts with nCUL5. Thus, Vif has a central role in organizing the Vif–CBF-β–CUL5–ELOB–ELOC pentameric complex. Consistent with a previous study[12], CBF-β–Vif binding causes no marked conformational changes in CUL5–ELOB–ELOC (Extended Data Fig. 3). ELOB makes no interaction with components other than ELOC, but we cannot rule out the possibility that the C-terminal portion (residues 103–118) of ELOB that was absent from the truncated ELOB in our construct is involved in the formation of the pentameric complex.

CUL1 is a close homologue of CUL5 and forms a complex with SKP1, a substrate adaptor of cullin-RING E3 ubiquitin ligases[17]. The CUL1–SKP1 and nCUL5–ELOC complexes share a similar structural organization, with the amino termini of CUL1 and CUL5 binding to SKP1 and ELOC, respectively (Fig. 1c). This observation supports the idea that the cullin-RING E3 scaffold proteins have a conserved mode of binding to substrate adaptors. However, the amino acids at the nCUL5–ELOC interface differ from those at the CUL1–SKP1 interface (Extended Data Fig. 4a–c), indicating that these residues govern specific interaction of cullin-RING E3 scaffold proteins with their respective substrate receptors.

Vif in the complex maintains an elongated, cone-like shape (Fig. 2a) with highly positive charges on its surface (Fig. 2b). A database search using the DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server) identified no structures appreciably similar to that of Vif, indicating that the viral protein possesses a novel fold. The two-domain structure of Vif, however, is reminiscent of the cellular substrate receptors SOCS2 and VHL that share the common substrate adaptor ELOBC with Vif (refs 16, 18, 19). The larger domain (referred to α/β domain) of Vif contains a distorted five-stranded antiparallel β-sheet with three helices tightly packing against the convex side (Fig. 2a). Two loosely packing helices form the smaller domain (referred to as the α-domain) of Vif that harbours the BC-box motif (residues 144–155) found in VHL and SOCS-box proteins[16,18,19]. The zinc-finger motif, HCCH, makes tetrahedral coordination to $Zn^{2+}$ (Fig. 2c and Extended Data Fig. 2b) and stabilizes the three inter-domain loops (Fig. 2a). Stabilization of the three inter-domain loops is further strengthened by their interaction with each other, which in turn rigidifies the two helices from the α-domain of Vif. The amino acids from the two domain regions, in particular those involved in packing of the secondary structural elements, are conserved among the Vif family of proteins (Extended Data Fig. 5a).

[1]School of Life Science and Technology, Harbin Institute of Technology, Harbin 150080, China. [2]MOE Key Laboratory of Protein Sciences, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China.
*These authors contributed equally to this work.

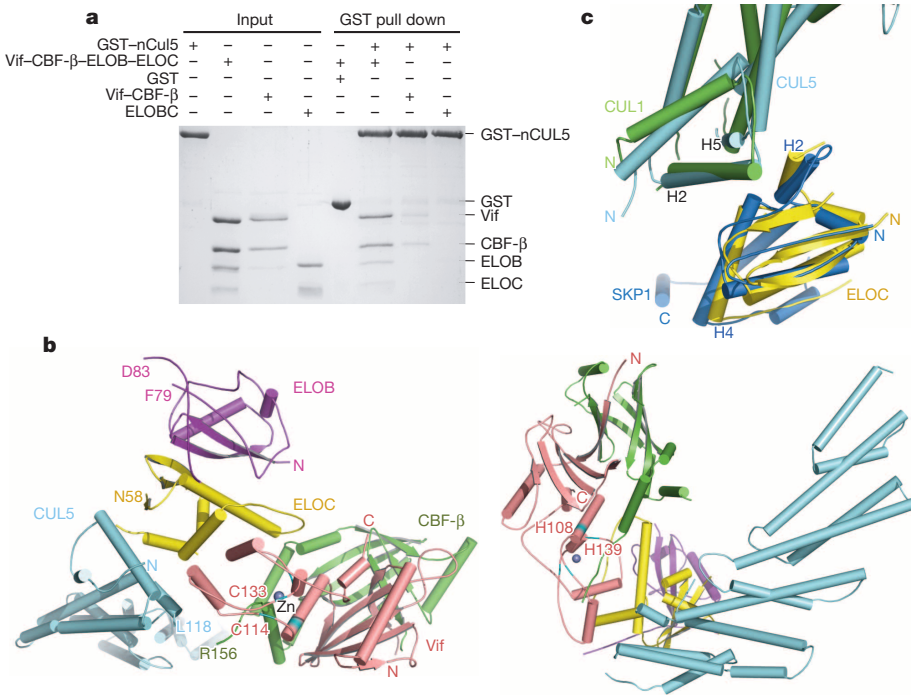



**Figure 1 | U-shaped structure of the Vif-organized Vif–CBF-β–CUL5–ELOB–ELOC complex.** **a**, Vif and CBF-β are important for ELOB–ELOC interaction with nCUL5 (residues 12–386). GST–nCUL5 was first bound to glutathione–sepharose and incubated with Vif–CBF-β–ELOB–ELOC, Vif–CBF-β or ELOB–ELOC protein as indicated. After extensive washing, the bound proteins were visualized by Coomassie blue staining after SDS–PAGE. **b**, Overall structures of Vif–CBF-β–nCUL5–ELOB–ELOC in two different orientations. Colour codes for the proteins are indicated. The grey sphere indicates $Zn^{2+}$. N and C represent N and C termini, respectively. Residue numbers are indicated. **c**, CUL5–ELOC and CUL1–SKP1 have a similar structural organization. Shown in the figure is structural superimposition of CUL1 (green)–SKP1 (blue) (Protein Data Bank code1LDK)[17] and CUL5–ELOC highlighting the conserved interface between the two complexes.

CBF-β and Vif form extensive interactions, burying a total surface area of 4,797 Å$^2$. CBF-β binds to a hydrophobic surface of Vif, leaving the positively charged surface exposed (Fig. 3a). The N-terminal peptide (residues 6–12) of Vif forms an antiparallel β-sheet with β-strand S3 from CBF-β (Fig. 3a, b). Vif residues Trp 5, Val 7 and Ile 9 from the peptide point to the central region of the β-barrel and make hydrophobic contacts with their respective neighbouring residues of CBF-β (Fig. 3b). A C-terminal peptide of CBF-β contributes to the CBF-β–Vif interaction via binding to a surface pocket formed between the two domains of Vif, rendering the Vif-bound $Zn^{2+}$ completely solvent inaccessible (Fig. 3a, c). Packing of the C-terminal portion of helix H5 from CBF-β against the central region of Vif also seems to be critical for the CBF-β–Vif

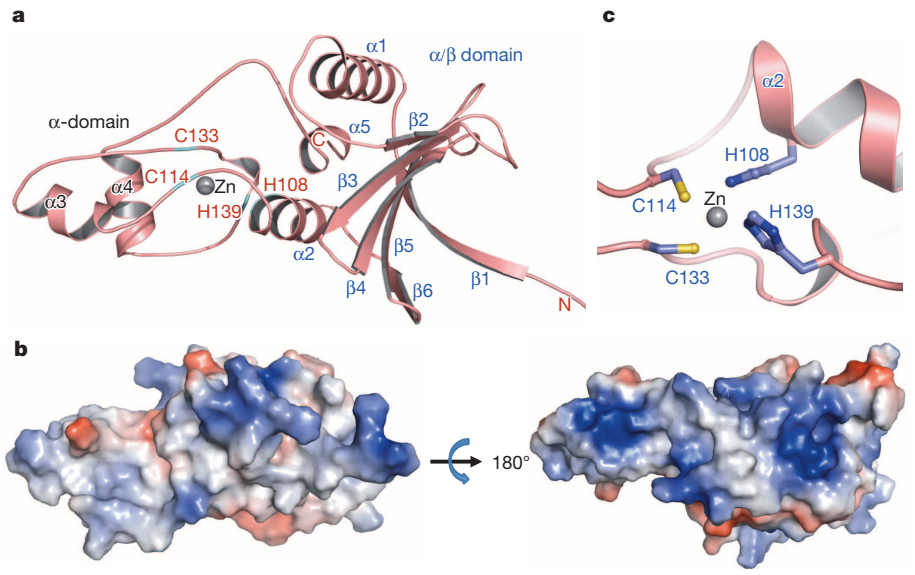



**Figure 2 | Vif contains two domains with zinc binding between them.** **a**, Overall structure of Vif shown in cartoon representation. Some of the secondary structural elements of the α-domain and α/β domain of Vif are labelled. **b**, Vif has a highly positively charged surface. Two views of the electrostatic surface potential map of Vif are shown. White, blue and red indicate neutral, positive and negative surfaces, respectively. **c**, The zinc-binding motif of Vif. The side chains of zinc-finger binding residues (His 108, Cys 114, Cys 133 and His 139) are shown as stick representation, and the $Zn^{2+}$ as a grey sphere.

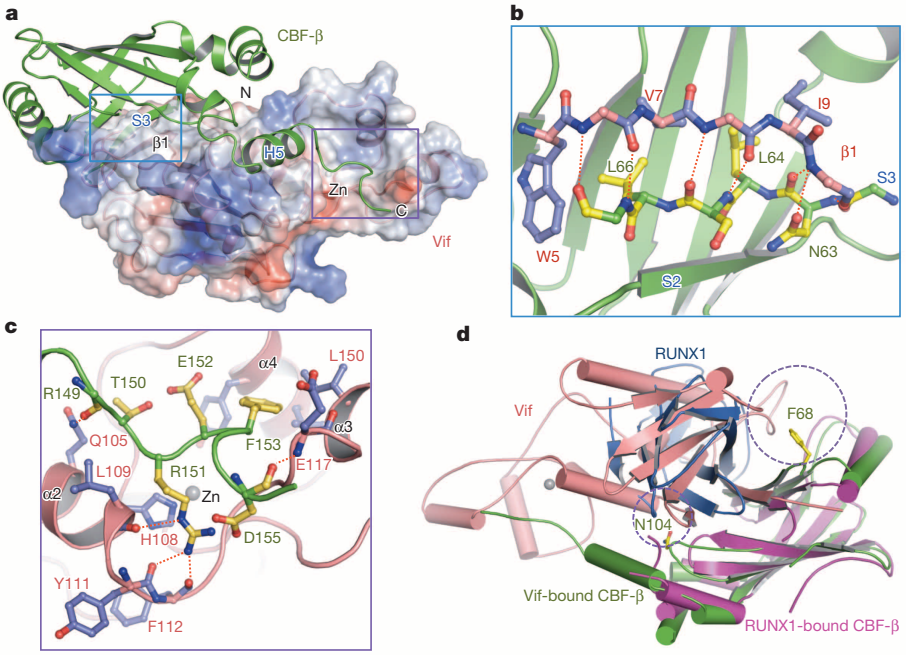



**Figure 3 | Vif and RUNX1 overlap when interacting with CBF-β.** **a**, Overall structure of the Vif–CBF-β complex (CBF-β in cartoon representation and Vif in electrostatic surface representation). Two regions of the Vif–CBF-β interaction are highlighted with purple and blue frames. Some of the secondary structural elements of CBF-β are labelled. **b**, A close-up view of the detailed interactions between Vif and CBF-β highlighted in **a**. Side chains from Vif involved in the interaction are shown in slate, and those from CBF-β in yellow. Red dashed lines represent hydrogen bonds. **c**, A close-up view of the detailed interactions between a Vif hydrophobic pocket and CBF-β C terminus highlighted in **a**. **d**, Vif and RUNX1 bind to the same side of CBF-β. Shown is the structural comparison of the Vif–CBF-β and RUNX1–CBF-β (Protein Data Bank code1E50)[22] complexes. Side chains of two residues (Phe 68 and Asn 104) from CBF-β are shown in stick representation and are highlighted. The Vif- and RUNX1-bound CBF-β is shown in green and purple, respectively.

interaction (Extended Data Fig. 6a). Consistent with the extensive CBF-β–Vif interactions, a CBF-β variant (residues 1–140) retained the ability to interact with Vif[20] (Extended Data Fig. 6b).

As a co-transcription factor of the RUNX family members, CBF-β associates with RUNX1, regulating expression of immunity-related genes[21]. Structural superposition of the CBF-β–RUNX1 (ref. 22) and CBF-β–Vif complexes showed that Vif completely overlaps with RUNX1 (Fig. 3d), indicating that Vif and RUNX1 are mutually exclusive for binding CBF-β. Supporting the structural observation, mutation of Asn 104 of CBF-β, which interacts with both Vif and RUNX1 (Fig. 3d), results in disruption of the Vif–CBF-β and RUNX1–CBF-β complexes[23]. In contrast, mutation of CBF-β Phe 68, which only interacts with Vif (Fig. 3d), impairs the Vif-binding activity of CBF-β but has no effect on the interaction of RUNX1 with CBF-β (ref. 24). A greater buried surface area as a result of Vif binding to CBF-β (4,797 Å$^2$) than RUNX1 (3,941 Å$^2$) indicates that CBF-β might have a higher affinity for Vif than RUNX1.

In addition to CBF-β, Vif also interacts with CUL5 and ELOC via its α-domain (Fig. 4a). As observed previously[12], the BC-box motif (helix α4) of Vif contacts H4 and its preceding loop of ELOC (Fig. 4b). Consistently, mutation of two equivalent HIV-1 Vif residues in HXB2 Vif[25]—I120S and L124S—or other key residues impaired Vif interaction with CUL5. The structural observations were further confirmed by mutations of other key residues involved in Vif–ELOC interaction (Extended Data Fig. 7). Structural comparison between the Vif–ELOC and SOCS2–ELOC[16] complexes (Fig. 4c) showed that, despite their low sequence homology, α4 of Vif is well aligned with H4 of SOCS2, indicating that Vif mimics SOCS2 for binding to ELOC. SKP1 shares sequence homology with ELOC[17]. Comparison of the Vif–CUL5–ELOC and CUL1–SKP1–SKP2 complexes[17] showed that α4 and α3 of Vif are also similarly positioned to the helix H6 of SKP1 and helix H1 of SKP2, respectively, to bind ELOC and CUL5 (Extended Data Fig. 4d), further supporting the idea that the cullin-RING E3 scaffold proteins have a conserved assembly mode.

CUL5 and CUL2 share the evolutionarily conserved adaptor protein ELOC, which assembles cullin-RING E3 ligase complexes[26]. Vif, however, preferentially uses CUL5, but not CUL2, as the scaffold protein to degrade APOBEC3G (ref. 3). Primary sequence alignment indicated that three Vif-interacting amino acids of CUL5—Leu 52, Trp 53 and Asp 55—are highly variable in CUL2 (Fig. 4d), indicating that they are the structural determinants for CUL5 selection by Vif. Simultaneous substitution in CUL5 of the three Vif-interacting residues Leu 52, Trp 53 and Asp 55 and the two ELOC-interacting residues Phe 41 and His 48 with their equivalents in CUL2 greatly impaired the ability of CUL5 to interact with Vif–CBF-β–ELOB–ELOC (Fig. 4e). A similar result was also obtained for the L52V and W53A double mutation of CUL5 (Fig. 4e), indicating that these two amino acids have a dominating role in CUL5 selection by Vif. CUL5 L52V and W53A are also the epitopes for preferential selection of CUL5 over CUL2 by SOCS2 (ref. 16). Structural comparison (Fig. 4c) revealed that α3 and its following loop (residues 116–131) from the α-domain of Vif are positioned similarly to H5 and H6 of SOCS2 (residues 177–192), a region termed the SOCS2 cullin box which is responsible for CUL5 binding, further supporting the structural and functional mimic of SOCS2 by Vif. These data strongly suggest that residues 116–131 in the α-domain of Vif act as a cullin box to preferentially recognize CUL5.

Our structure reveals that Vif has a critical role in organizing assembly of the Vif–CBF-β–nCUL5–ELOB–ELOC penatmeric complex by interacting with CBF-β, nCUL5 and ELOC (Fig. 1). Supporting this conclusion, nCUL5 displays a weaker interaction with ELOC–ELOB in the absence of Vif–CBF-β (Fig. 1a). Additionally, mutation of Vif Leu 145, which is important for the Vif–ELOC interaction, abolishes binding of Vif to CUL5 (ref. 12). Although not interacting with the other components in the pentameric complex, the zinc-finger motif may stabilize the conformation of the α-domain to promote Vif interaction with CUL5. Indeed, the F115A mutation of Vif predicted to perturb the α-domain (Extended Data Fig. 8) greatly reduces Vif interaction with CUL5 and compromises Vif-mediated APOBEC3G degradation[25]. The

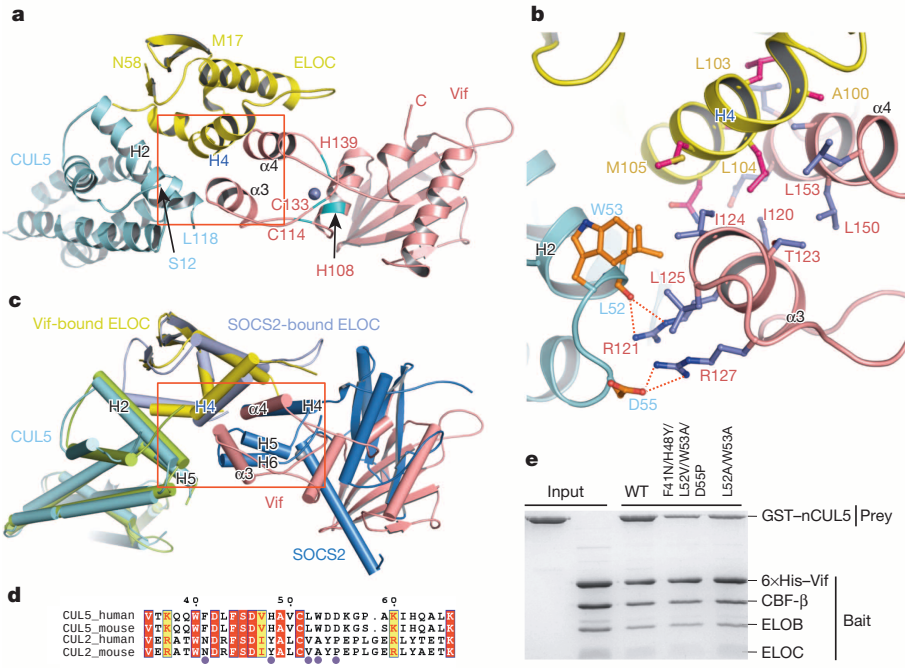



**Figure 4 | Interaction of Vif with nCUL5 and ELOC.** **a**, The α-domain of Vif interacts with nCUL5 and ELOC. Overall interaction of two helices of α-domain with nCUL5 and ELOC is shown. **b**, Specific interactions of Vif with nCUL5 and ELOC. A close-up view of the interactions (highlighted in **a**) of helices α3 and α4 of Vif with helix H2 and its following loop of nCUL5 and H4 of ELOC, respectively. **c**, Vif structurally mimics SOCS2. Structural comparison of Vif–ELOC–CUL5 and SOCS2 (blue)–ELOC (light blue)–CUL5 (lemon) (Protein Data Bank accession 4JGH)[18]; the interfaces of the two complexes are highlighted in red. **d**, Sequence alignment of CUL5 and CUL2 around the Vif-interacting region of CUL5. Conserved and similar residues are highlighted with red and yellow shading, respectively. Residues of CUL5 involved in Vif interaction are indicated with slate solid dots at the bottom. **e**, Structural determinants for preferential binding of Vif to nCUL5. Wild-type Vif–CBF-β–ELOB–ELOC complex in which Vif was tagged with poly-His was first bound to $Ni^{+}$ resin and incubated with wild-type (WT) or mutant GST–nCUL5 proteins. The bound proteins were visualized by Coomassie blue staining after SDS–PAGE.

C-terminal CBF-β peptide may further stabilize the conformation of the α-domain of Vif by binding between the two domains of Vif. Consistently, deletion of the CBF-β peptide greatly impaired the assembly of the Vif-containing pentameric complex (Extended Data Fig. 6b), although the peptide is not involved in binding to CUL5 or ELOC. The three Vif-interacting proteins selectively bind to the hydrophobic surfaces of Vif, leaving the positively charged patches solvent-exposed (Extended Data Fig. 5b), some of which are important for APOBEC3G binding[27,28]. Although the APOBEC3G-binding regions of Vif vary in different studies[27–30], the critical amino acids from them are exposed in the structure (Extended Data Fig. 5c), indicating that extensive interactions between Vif and APOBEC3G may be formed.

Our data (Fig. 4c) indicate that the α-domain of Vif structurally and functionally mimics SOCS2. Vif and SOCS2 probably use a conserved mechanism for selecting the adaptor protein CUL5, because mutation of the non-conserved CUL5 residues Val 52 and Trp 53, which are important for the interaction with SOCS2, also greatly reduces the efficiency for assembly of the Vif-organized complex (Fig. 4e). In addition to affinity (Extended Data Fig. 9), higher abundance of CUL5 than that of CUL2 in certain infected cells (http://biogps.org) may also contribute to Vif selection of CUL5 during HIV-1 infection. In line with this, VHL interacts with both CUL2 and CUL5 when overexpressed, although VHL selects CUL2 under physiological conditions[18].

We have solved the crystal structure of a Vif–CBF-β–nCUL5–ELOB–ELOC complex. Our data not only reveal the molecular mechanism underlying hijacking of the CBF-β and CUL5 E3 ligase complex by Vif, but also open the possibility of designing and developing novel antiviral drugs that target the Vif-containing complex.

## METHODS SUMMARY

6×His–Vif, CBF-β and ELOB–ELOC were co-expressed in *Escherichia coli* BL21 (DE3) cells and purified using Ni-NTA (Qiagen) followed by anion ion exchange chromatography (Source-15Q, Pharmacia). Glutathione *S*-transferase (GST)–nCUL5 was expressed and purified separately. Pentameric complex was formed by mixing Vif–CBF-β–ELOB–ELOC and nCUL5 proteins. After limited proteolysis with elastase, Vif–CBF-β–nCUL5–ELOB–ELOC was cleaned using an anion ion exchange column and subjected to size-exclusion chromatography (Superdex200 10/300 GL column, GE Healthcare). Crystals of Vif–CBF-β–nCUL5–ELOB–ELOC were grown using the hanging-drop vapour diffusion method and the structure was determined using molecular replacement.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 12 August; accepted 15 November 2013.**

1. Mariani, R. *et al.* Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* **114,** 21–31 (2003).
2. Sheehy, A. M., Gaddis, N. C. & Malim, M. H. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nature Med.* **9,** 1404–1407 (2003).
3. Yu, X. *et al.* Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302,** 1056–1060 (2003).
4. Stopak, K., de Noronha, C., Yonemoto, W. & Greene, W. C. HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Mol. Cell* **12,** 591–601 (2003).
5. Marin, M., Rose, K. M., Kozak, S. L. & Kabat, D. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nature Med.* **9,** 1398–1403 (2003).
6. Conticello, S. G., Harris, R. S. & Neuberger, M. S. The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr. Biol.* **13,** 2009–2013 (2003).
7. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418,** 646–650 (2002).
8. Mangeat, B. *et al.* Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424,** 99–103 (2003).
9. Harris, R. S. *et al.* DNA deamination mediates innate immunity to retroviral infection. *Cell* **113,** 803–809 (2003).
10. Zhang, H. *et al.* The cytidinedeaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **424,** 94–98 (2003).

11. Lecossier, D., Bouchonnet, F., Clavel, F. & Hance, A. J. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* **300,** 1112 (2003).
12. Stanley, B. J. *et al.* Structural insight into the human immunodeficiency virus Vif SOCS box and its role in human E3 ubiquitin ligase assembly. *J. Virol.* **82,** 8656–8663 (2008).
13. Luo, K. *et al.* Primate lentiviral virion infectivity factors are substrate receptors that assemble with cullin 5–E3 ligase through a HCCH motif to suppress APOBEC3G. *Proc. Natl Acad. Sci. USA* **102,** 11444–11449 (2005).
14. Zhang, W., Du, J., Evans, S. L., Yu, Y. & Yu, X. F. T-cell differentiation factor CBF-β regulates HIV-1 Vif-mediated evasion of host restriction. *Nature* **481,** 376–379 (2011).
15. Jager, S. *et al.* Vif hijacks CBF-β to degrade APOBEC3G and promote HIV-1 infection. *Nature* **481,** 371–375 (2011).
16. Kim, Y. K. *et al.* Structural basis of intersubunit recognition in elongin BC-cullin 5-SOCS box ubiquitin-protein ligase complexes. *Acta Crystallogr. D* **69,** 1587–1597 (2013).
17. Zheng, N. *et al.* Structure of the Cul1–Rbx1–Skp1–F box$^{Skp2}$ SCF ubiquitin ligase complex. *Nature* **416,** 703–709 (2002).
18. Stebbins, C. E., Kaelin, W. G. Jr & Pavletich, N. P. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science* **284,** 455–461 (1999).
19. Bullock, A. N., Debreczeni, J. E., Edwards, A. M., Sundstrom, M. & Knapp, S. Crystal structure of the SOCS2-elongin C-elongin B complex defines a prototypical SOCS box ubiquitin ligase. *Proc. Natl Acad. Sci. USA* **103,** 7637–7642 (2006).
20. Zhou, X., Evans, S. L., Han, X., Liu, Y. & Yu, X. F. Characterization of the interaction of full-length HIV-1 Vif protein with its key regulator CBFβ and CRL5 E3 ubiquitin ligase components. *PLoS ONE* **7,** e33495 (2012).
21. Taniuchi, I. *et al.* Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. *Cell* **111,** 621–633 (2002).
22. Warren, A. J., Bravo, J., Williams, R. L. & Rabbitts, T. H. Structural basis for the heterodimeric interaction between the acute leukaemia-associated transcription factors AML1 and CBFβ. *EMBO J.* **19,** 3004–3015 (2000).
23. Kim, D. Y. *et al.* CBF-β stabilizes HIV Vif to counteract APOBEC3 at the expense of RUNX1 target gene expression. *Mol. Cell* **49,** 632–644 (2013).
24. Hultquist, J. F., McDougle, R. M., Anderson, B. D. & Harris, R. S. HIV type 1 viral infectivity factor and the RUNX transcription factors interact with core binding factor β on genetically distinct surfaces. *AIDS Res. Hum. Retroviruses* **28,** 1543–1551 (2012).
25. Xiao, Z. *et al.* Assembly of HIV-1 Vif-Cul5 E3 ubiquitin ligase through a novel zinc-binding domain-stabilized hydrophobic interface in Vif. *Virology* **349,** 290–299 (2006).
26. Zimmerman, E. S., Schulman, B. A. & Zheng, N. Structural assembly of cullin-RING ubiquitin ligase complexes. *Curr. Opin. Struct. Biol.* **20,** 714–721 (2010).
27. Chen, G., He, Z., Wang, T., Xu, R. & Yu, X. F. A patch of positively charged amino acids surrounding the human immunodeficiency virus type 1 Vif SLVx4Yx9Y motif influences its interaction with APOBEC3G. *J. Virol.* **83,** 8674–8682 (2009).
28. Russell, R. A. & Pathak, V. K. Identification of two distinct human immunodeficiency virus type 1 Vif determinants critical for interactions with human APOBEC3G and APOBEC3F. *J. Virol.* **81,** 8201–8210 (2007).
29. Zhang, H., Pomerantz, R. J., Dornadula, G. & Sun, Y. Human immunodeficiency virus type 1 Vif protein is an integral component of an mRNP complex of viral RNA and could be involved in the viral RNA folding and packaging process. *J. Virol.* **74,** 8252–8261 (2000).
30. He, Z., Zhang, W., Chen, G., Xu, R. & Yu, X. F. Characterization of conserved motifs in HIV-1Vif required for APOBEC3G and APOBEC3F interaction. *J. Mol. Biol.* **381,** 1000–1011 (2008).


**Acknowledgements** We thank F. Yu and J. He at Shanghai Synchrotron Radiation Facility (SSRF) for help with data collection. We thank J. Chai for critical reading of the manuscript. This research was funded by the National Natural Science Foundation of China grant 31300605 and the Program for New Century Excellent Talents in University grant AUGA5710060713 to Z.H.; Ministry of Science and Technology of China grants 2012CB911101 and 2011CB910502, and the National Natural Science Foundation of China grants 31030020 and 31170679 to M.Y.

**Author Contributions** Z.H. designed the experiments. Y.G., L.D. and Z.H. performed the bulk of the experiments. Data were analysed by Z.H., Y.G., L.D., X.Q.; Y.W., B.Z., H.L., Y.Y., Y.Z. and M.Y. contributed to some experiments and discussions. Z.H., Y.G., L.D. and X.Q. wrote the paper.

**Author Information** The atomic coordinates and structure factors for Vif–CBF-β–nCUL5–ELOB–ELOC complex have been deposited in the Protein Data Bank under the accession code 4N9F. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.H. (huangzhiwei@hit.edu.cn).

LETTER

doi:10.1038/nature12815

# Structural basis of lentiviral subversion of a cellular protein degradation pathway

David Schwefel[1], Harriet C. T. Groom[2], Virginie C. Boucherit[2], Evangelos Christodoulou[1], Philip A. Walker[1], Jonathan P. Stoye[2], Kate N. Bishop[2] & Ian A. Taylor[1]

**Lentiviruses contain accessory genes that have evolved to counteract the effects of host cellular defence proteins that inhibit productive infection. One such restriction factor, SAMHD1, inhibits human immunodeficiency virus (HIV)-1 infection of myeloid-lineage cells[1,2] as well as resting $CD4^+$ T cells[3,4] by reducing the cellular deoxynucleoside 5′-triphosphate (dNTP) concentration to a level at which the viral reverse transcriptase cannot function[5,6]. In other lentiviruses, including HIV-2 and related simian immunodeficiency viruses (SIVs), SAMHD1 restriction is overcome by the action of viral accessory protein x (Vpx) or the related viral protein r (Vpr) that target and recruit SAMHD1 for proteasomal degradation[7,8]. The molecular mechanism by which these viral proteins are able to usurp the host cell's ubiquitination machinery to destroy the cell's protection against these viruses has not been defined. Here we present the crystal structure of a ternary complex of Vpx with the human E3 ligase substrate adaptor DCAF1 and the carboxy-terminal region of human SAMHD1. Vpx is made up of a three-helical bundle stabilized by a zinc finger motif, and wraps tightly around the disc-shaped DCAF1 molecule to present a new molecular surface. This adapted surface is then able to recruit SAMHD1 via its C terminus, making it a competent substrate for the E3 ligase to mark for proteasomal degradation. The structure reported here provides a molecular description of how a lentiviral accessory protein is able to subvert the cell's normal protein degradation pathway to inactivate the cellular viral defence system.**

HIV-1 infection of myeloid and $CD4^+$ T cells is inhibited by the post-entry restriction factor SAMHD1. In other primate lentiviruses, including HIV-2 and SIV, this block is overcome by the expression of the Vpx accessory protein. Vpx recruits SAMHD1 to the DDB1–CUL4A–ROC1 E3 ubiquitin ligase complex through interaction with the substrate-adaptor protein DCAF1, and facilitates its degradation through the proteasomal pathway[1,2,7,8]. To understand the mechanism of Vpx-mediated recruitment of SAMHD1, we assessed which regions of each molecule (Fig. 1a) are required for the interaction. These data reveal that only SAMHD1 molecules containing a C-terminal region (residues 582–626) are able to support ternary complex formation (Fig. 1b, compare centre and left panels) and that this region alone is sufficient for the interaction (Fig. 1b, right). We therefore determined the crystal structure of the ternary complex of the C-terminal WD40 domain of DCAF1 (DCAF1-CtD) with the Vpx of SIV from the sooty mangabey (*Cercocebus atys*) ($Vpx_{sm}$) and the C-terminal region of SAMHD1 (SAMHD1-CtD). The crystal structure was solved by single-wavelength anomalous dispersion (SAD) (Extended Data Fig. 1 and Extended Data Table 1) and is shown in Fig. 1c. DCAF1-CtD comprises a seven-bladed β-propeller disc-shaped molecule 45 Å in diameter and 20 Å in depth. $Vpx_{sm}$ comprises an antiparallel V-shaped three-helical bundle that wraps around one side and the top of DCAF1-CtD. This arrangement of helices is conserved in the HIV-1 Vpr solution structure[9]. However, the structures differ significantly at the helical termini and, in $Vpx_{sm}$, zinc coordinated by His 39, His 82, Cys 87 and Cys 89 (Figs 1c and 2a) brings together the C termini of helices 1 and 3 to stabilize the structure. Residues Asn 606 to Asp 624 of SAMHD1-CtD are also well ordered. They form two short perpendicular α-helices, helix A (Leu 610–Ala 613) and helix B (Arg 617–Lys 622), connected by a three-residue linker (S614–S616), and pack into a cleft between $Vpx_{sm}$ and DCAF1-CtD (Fig. 1c).

The complex contains four interfacial regions (Fig. 2b), a combined $Vpx_{sm}$–SAMHD1–DCAF1 ternary interface (Fig. 2c) and a more extensive DCAF1–$Vpx_{sm}$-binding surface with three sites of interaction (Fig. 2d–f). The $Vpx_{sm}$–SAMHD1 interaction buries 700 $Å^2$ of molecular surface. At the interface the hydrophobic side chains of Leu 610, Val 618, Leu 620 and Phe 621 from SAMHD1 helix A and B pack into a hydrophobic pocket between the amino termini of $Vpx_{sm}$ helix 1 and 3 (Fig. 2c). The interface also contains electrostatic interactions between acidic residues Glu 15 and Glu 16 at the N terminus of $Vpx_{sm}$ helix 1 and Arg 609 and Arg 617, part of a tribasic Arg 609-Arg 617-Lys 622 motif, in SAMHD1. By contrast, the contact area between SAMHD1-CtD and DCAF1-CtD is small, just 210 $Å^2$. The only direct interaction is between Lys 622 in the SAMHD1 tribasic motif and Asp 1092 of DCAF1 in the acidic Glu 1091-Asp-Glu 1093 loop connecting blade 7 and 1 of the DCAF1-CtD propeller (Fig. 2c). Asp 1092 of SAMHD1-CtD is also hydrogen bonded to Tyr 66 of $Vpx_{sm}$ in an interaction that bridges SAMHD1-CtD and $Vpx_{sm}$. Several other key residues in $Vpx_{sm}$ also mediate bridging interactions. These include Trp 24, which stacks against Arg 617 of SAMHD1-CtD as well as hydrogen bonding to Asn 1135 of DCAF1 (Fig. 2e), and Tyr 69, which packs with Val 618 of SAMHD1-CtD and is also hydrogen bonded to the side chain of Glu 1091 in the DCAF1-CtD Glu-Asp-Glu loop.

The DCAF1–$Vpx_{sm}$ interface is much larger, made up from three sites of interaction burying 2,000 Å of surface (Fig. 2d–f and Extended Data Table 2). The first involves the N-terminal extended region of $Vpx_{sm}$ (Glu 6–Ser 13), which packs against the concave surface of DCAF1 blade 1 spanning from the underside to the topside of the disc, making several hydrogen bonds and hydrophobic interactions (Fig. 2d). Further interactions involve Trp 24, Thr 28, Ile 32 and Gln 76 in helix 1 and 3 of $Vpx_{sm}$, which contact residues on blade 1 and 2 of DCAF1-CtD and the interspersing loop (Fig. 2e). Helix 3 of $Vpx_{sm}$ lies in a groove between blade 1 and 7 on the upper face of DCAF1-CtD. Interactions between the charged and hydrophobic side chains of Lys 77, Arg 70, Phe 80 and Met 81 along the length of helix 3 with residues in the intra-strand loops of blade 1 and 7 of DCAF-CtD, including the Glu-Asp-Glu loop, comprise the third region of the DCAF1–$Vpx_{sm}$ interface (Fig. 2f).

Mutation of many of the residues at these interfaces have been shown previously to reduce viral infectivity, disrupt binding of Vpx proteins to DCAF1 and interfere with proteasomal degradation of SAMHD1 (Extended Data Table 3). Examples include: the bridging residue of $Vpx_{sm}$, Trp 24 (ref. 10; Fig. 2c, e); $Vpx_{sm}$ Gln 76 (ref. 11), which is hydrogen bonded to Asn 1135 and Trp 1156 of DCAF1 (Fig. 2e); and $Vpx_{sm}$ Lys 77 (ref. 12), which is integral to an extensive salt-bridge network that links residues in the DCAF1 Glu-Asp-Glu loop with Arg 70, Tyr 69 and Tyr 66 in helix 3 of $Vpx_{sm}$ (Fig. 2f).

[1]Division of Molecular Structure, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK. [2]Division of Virology, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.

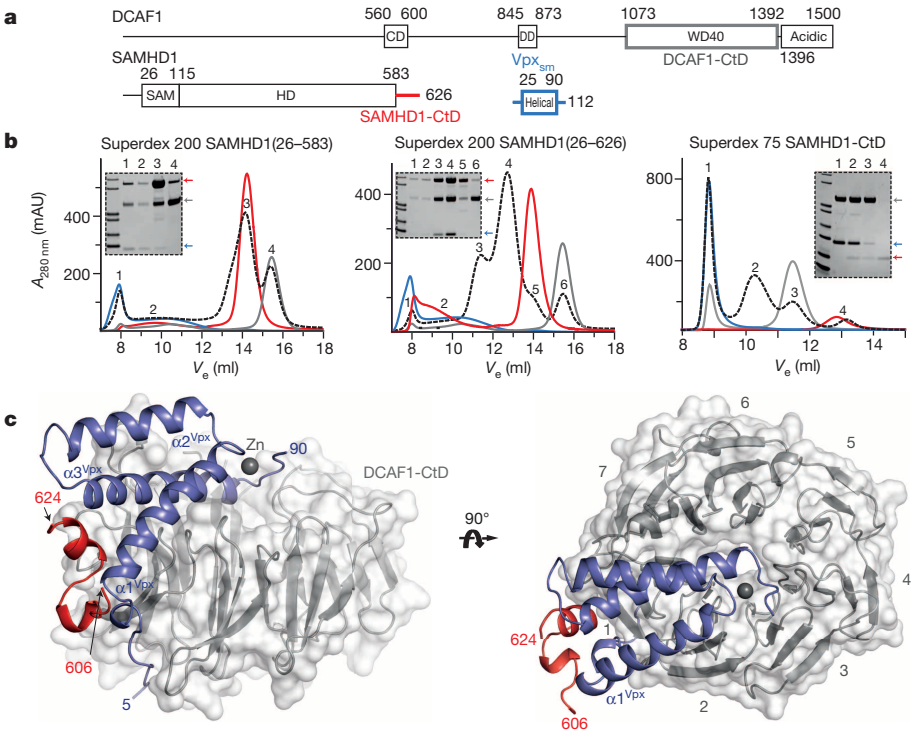



**Figure 1 | The SAMHD1-CtD–Vpx$_{sm}$–DCAF1-CtD complex.** **a**, Schematic of proteins. CD, chromo domain; DD, dimerization domain; HD, His/Asp domain; SAM, sterile α-motif. Regions coloured grey (DCAF1, 1058–1396), red (SAMHD1, 582–626) and blue (Vpx$_{sm}$, 1–112) were used for crystallization. **b**, Size-exclusion chromatograms (black) of equimolar mixtures of Vpx$_{sm}$, DCAF1-CtD and SAMHD1(26–583) (left), SAMHD1(26–626) (middle) and SAMHD1(582–626) (right). Chromatograms from individual components are also shown: Vpx$_{sm}$ (blue), DCAF1-CtD (grey) and SAMHD1 (red). SDS–PAGE analyses of peaks are inset. Peak 1 (void volume) contains unspecific aggregates. $A_{280\,nm}$, absorbance at 280 nm, $V_e$ is elution volume. **c**, Cartoon representation of the ternary complex. DCAF1-CtD is shown in a grey surface, β-propeller blades are numbered. SAMHD1-CtD is red, Vpx$_{sm}$ is blue and a zinc ion is shown as a grey sphere.

In the structure, SAMHD1-CtD makes salt bridges at both the Vpx$_{sm}$ and DCAF1-CtD interfaces through charged side chains in the tribasic motif. *In vitro* binding studies show that full-length SAMHD1 and SAMHD1-CtD have comparable affinity for the DDB1–Vpx–DCAF1 complex[13]. However, to test whether SAMHD1-CtD alone is sufficient for recruitment by Vpx and to assess the contribution of the tribasic motif, an *in vivo* reporter assay was used. SAMHD1-CtD was fused to the C terminus of a tandem green fluorescent protein (GFP)-tagged nuclear localization signal (NLS) protein that localizes to the nucleus[14]. Human 293T cells transduced with the NLS–GFP–SAMHD1-CtD fusion (Fig. 3a) display GFP fluorescence in their nuclei (Fig. 3b). Delivery of Vpx by infection with SIV$_{mac}^{Vpx+}$ (SIV isolated from *Macaca mulatta*) virus greatly reduces the number of GFP$^+$ cells (Fig. 3b, c). By contrast, infection with SIV$_{mac}^{Vpx-}$ virus or addition of the proteasomal inhibitor MG132 with Vpx$^+$ virus does not reduce the population of GFP$^+$ cells (Fig. 3b, c), indicating that loss of GFP fluorescence results from Vpx-mediated proteasomal degradation of NLS–GFP–SAMHD1-CtD. In addition, mutation of residues in the SAMHD1 tribasic motif (Arg 609-Arg 617-Lys 622) to alanine or glutamate also abolishes or severely

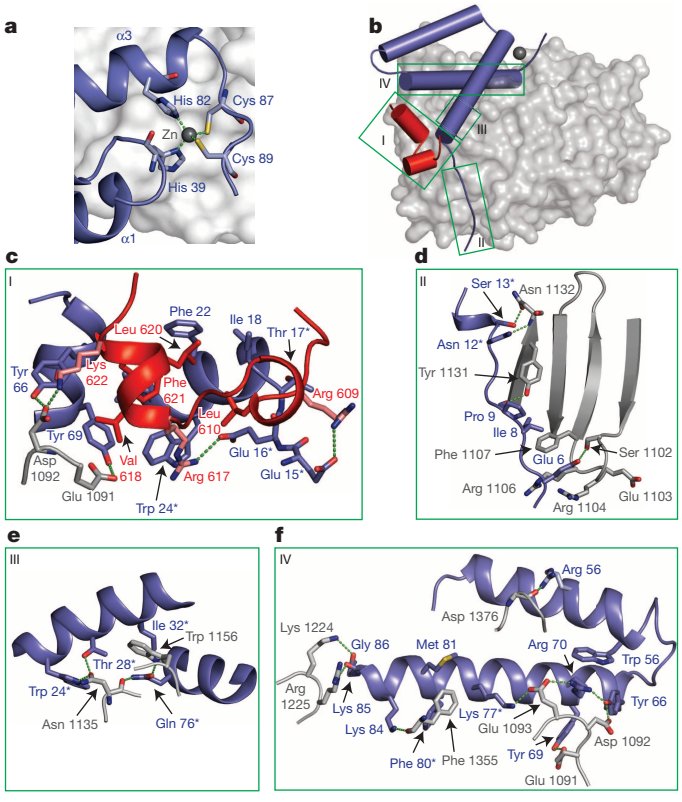



**Figure 2 | Intermolecular interfaces.** **a**, Zn ion (grey sphere) and surrounding residues. Co-ordinating residues are displayed as sticks, co-ordinate bonds as green dashes. **b**, Cartoon representation of SAMHD1-CtD–Vpx$_{sm}$–DCAF1-CtD. DCAF1-Ctd is shown in grey, cylinders represent α-helices in SAMHD1-CtD (red) and Vpx$_{sm}$ (blue), intermolecular interfaces are highlighted by green boxes (I, II, III, IV). **c**–**f**, Views of the interface between SAMHD1-CtD–Vpx$_{sm}$–DCAF1-CtD (box I) and Vpx$_{sm}$–DCAF1-CtD (box I–IV). Residues contributing to the interface are shown as sticks, hydrogen-bonding interactions as dashed lines and residues important for Vpx function are highlighted with an asterisk.

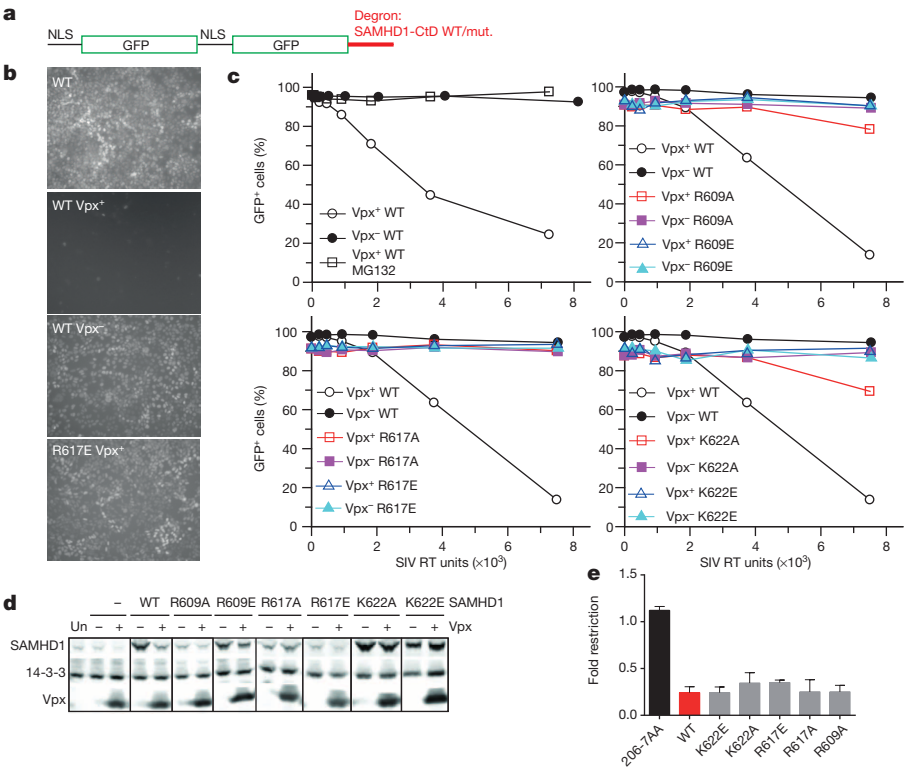



**Figure 3 | SAMHD1 C-terminal region.** **a**, The NLS–GFP–SAMHD1-CtD degron. WT, wild type; mut., mutant. **b**, Microscopic images (×20 magnification) of uninfected and SIV $Vpx^{+/-}$ infected 293T cells expressing NLS–GFP–SAMHD1-CtD wild type or R617E mutant. **c**, Quantitative analysis of GFP–SAMHD1 degron expression after infection with SIV $Vpx^{+/-}$ viruses. Percentages of $GFP^+$ cells are plotted against SIV virus titre. RT units, reverse transcriptase activity ($mU\ ml^{-1}$). MG132 was added where indicated. Top left panel is representative of three independent experiments. **d**, Immunoblots of 293T cells co-transfected with SAMHD1 and Vpx plasmids (Un., untransfected; −, empty vector; +, Vpx) probed with antibodies to SAMHD1 (top), 14-3-3 (middle) or $Vpx_{HIV-2}$ (bottom). Representative of two independent experiments. **e**, Restriction of HIV-1 infection by wild-type SAMHD1 and mutants. Restriction is expressed as a ratio of HIV-infected SAMHD1-transduced to infected SAMHD1-negative cells (>3 independent experiments with different viral stocks). Data represent the mean value and error bars are standard deviation.

diminishes the capacity of Vpx to induce degradation of NLS–GFP–SAMHD1-CtD (Fig. 3b, c). These data reveal that SAMHD1-CtD in isolation acts as a Vpx-dependent degron to induce the proteasomal turnover of a heterologous protein, and that disruption of the protein–protein interactions observed in the crystal structure prevent Vpx-mediated degradation. When the same mutations are introduced into SAMHD1 they also reduce the capacity of Vpx to induce degradation, albeit to varying degrees (Fig. 3d). However, all mutants show wild-type restriction of HIV-1, indicating that the CtD is not required for SAMHD1 anti-HIV-1 activity (Fig. 3e).

To put the DCAF1-CtD–$Vpx_{sm}$–SAMHD1-CtD structure into the context of the E3 ligase, a molecular model of the entire CUL4A–DDB1–ROC1–DCAF–SAMHD1–$Vpx_{sm}$ assembly was constructed by superposition of DCAF1-CtD–$Vpx_{sm}$–SAMHD1-CtD onto existing structures in the Protein Data Bank (PDB). First, the β-propellers of DCAF1-CtD and the related substrate adaptor, DDB2, were aligned (Extended Data Fig. 2, inset), facilitating substitution of DCAF1-CtD for DDB2 in the existing DDB1–DDB2 structure[15]. The DDB1–CUL4A–ROC1 interface is highly flexible[16]. Therefore, the DCAF1-CtD–$Vpx_{sm}$–SAMHD1-CtD–DDB1 model was superposed onto the two most extreme conformations available, allowing the range of orientations that CUL4A–ROC1 can adopt with respect to $Vpx_{sm}$ and SAMHD1 to be visualized (Extended Data Fig. 2). The model places $Vpx_{sm}$ and SAMHD1-CtD on the opposite face of the DCAF1-CtD disk from the DCAF1–DDB1-binding site, accessible to the RING domain of ROC1. Moreover, in both conformations SAMHD1 is placed in the proximity of the ROC1 RING domain, ideally located for ubiquitin transfer. Notably, regions of SAMHD1 proximal to the bound SAMHD-CtD are required for catalytic activation/tetramerization[17] and association with Vpx–DCAF1–DDB1 inhibits SAMHD1 catalysis[13], suggesting that recruitment to the CUL4A–DDB1–ROC1 complex might additionally downregulate SAMHD1 activity through tetramer disassembly.

Reprogramming of the CUL4–DDB1–ROC1 E3 ubiquitin ligase is also used by paramyxovirus and hepatitis B virus to subvert the cellular antiviral response. These viruses usurp the interaction of DDB1 with DCAF1 by installing the viral substrate recruitment factors V or X in its place[18,19]. In lentiviruses, a different strategy has evolved in which the substrate recruitment factor DCAF1 is itself adapted by association with accessory proteins to create a new binding pocket, in the case of $Vpx_{sm}$ and $Vpx_{HIV-2}$ to recruit SAMHD1 through its C-terminal region (Extended Data Fig. 3). Notably, the SAMHD1 tribasic motif is conserved among primates but is absent in species that are not HIV/SIV hosts (Fig. 4a). Similarly, the N-terminal $Vpx_{sm}$ sequence $_9$PPGNSGEET$_{17}$, containing Glu 15 and Glu 16 that make salt bridges with Arg 609 and Arg 617, is conserved among all Vpx proteins that target human SAMHD1 for degradation[7] (Fig. 4b), suggesting that the complementarity between these motifs is a driver of the specificity of the SAMHD1-CtD–Vpx interaction. In Vpx proteins that do not induce degradation of human SAMHD1 (refs 7, 20), present in viruses isolated from red-capped mangabey (*C. torquatus*) ($Vpx_{rcm}$) and mandrill (*Mandrillus sphinx*) ($Vpx_{mnd-2}$), $_9$PPGNSGEET$_{17}$ is not conserved (Fig. 4b). However, these Vpx proteins still induce degradation of SAMHD1 but do so by targeting it to DCAF1 through sequences located towards the SAMHD1 N terminus[20].

In some SIVs, the evolutionarily related accessory protein Vpr recruits SAMHD1 for degradation[7]. By contrast, $Vpr_{HIV-1}$ still associates with DCAF1 and the CUL4A–DDB1–ROC1 E3 ligase but results in cell cycle arrest at G2, probably through recruitment of an unidentified cellular factor to the E3 ligase[21–23]. In the complex, we identified a structural zinc-binding site in $Vpx_{sm}$ (Fig. 2a) that is conserved in both Vpx and Vpr proteins (Fig. 4b). Mutation of His 71 in $Vpr_{HIV-1}$, the equivalent of the $Vpx_{sm}$ zinc-coordinating His 82, abrogates DDB1–DCAF1 binding and Vpr-induced cell cycle arrest[24]. Furthermore, most of the other conserved residues map to the DCAF1 interface (Fig. 4c) and mutation of two of these (Vpr Gln 65 and Trp 18, the equivalents of Vpx Gln 76 and Trp 24) also results in loss of Vpr function[23,25] (Extended Data Table 3). These observations suggest a strong structural conservation between these related accessory proteins. Moreover, given that Vpr-induced cell cycle arrest is also mediated through association with DCAF1 and the actions of the CUL4A–DDB1–ROC1 E3 ligase[21–23], it is likely that both factors use a similar mechanism to target cellular proteins to the CUL4A complex (Extended Data Fig. 3). Consequently, although the cellular

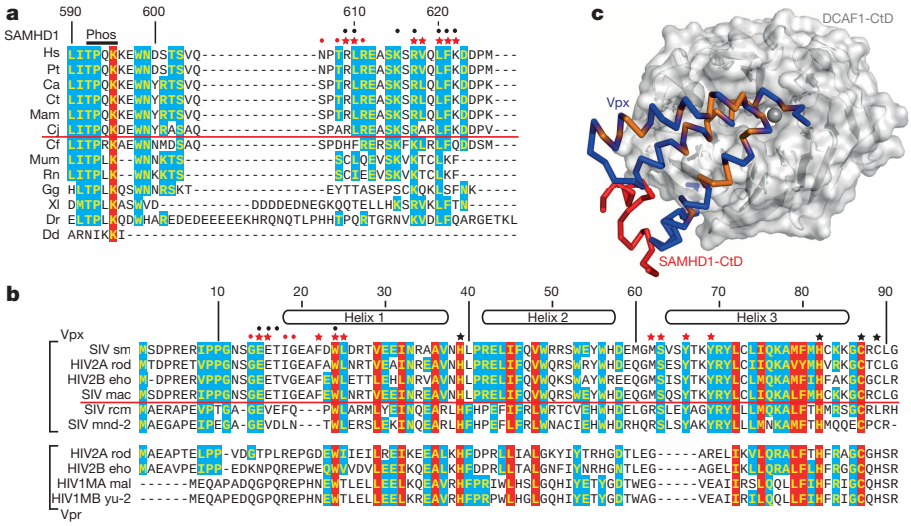



**Figure 4 | Species specificity of the SAMHD1–Vpx interaction.** **a**, Sequence alignment of the C termini of SAMHD1. One-hundred per cent type-conserved residues are highlighted red, 60% are highlighted cyan. Residue numbers refer to human SAMHD1-CtD. Red stars indicate side chains involved in $Vpx_{sm}$–DCAF1-CtD binding, red dots indicate main-chain interactions. Black dots indicate residues whose mutation impairs DDB1–DCAF1–Vpx binding[8]. Species above the red line are HIV-2/SIV hosts. Phos, phosphorylation site. Ca, *Cercocebus atys*; Cf, *Canis familiaris*; Cj, *Callithrix jacchus*; Ct, *Cercocebus torquatus*; Dd, *Dictyostelium discoideum*; Dr, *Danio rerio*; Gg, *Gallus gallus*; Hs, *Homo sapiens*; Mam, *Macaca mulatta*; Mum, *Mus musculus*; Pt, *Pan troglodytes*; Rn, *Rattus norvegicus*; Xl, *Xenopus laevis*. See Methods for sequence accession numbers. **b**, Alignment of Vpx and Vpr proteins, coloured and annotated as in **a**. Numbering refers to $Vpx_{sm}$. Red stars indicate side chains involved in SAMHD1-CtD binding, red dots indicate main-chain interactions and black stars indicate zinc ligands. Black dots indicate residues whose mutation impairs SAMHD1 degradation. Proteins above the red line induce degradation of human SAMHD1, proteins below cannot. See Methods for sequence accession numbers. **c**, Type-conserved amino acid residues in $Vpx_{sm}$ and $Vpr_{HIV-1}$ (orange highlight) mapped onto the $Vpx_{sm}$–DCAF1-CtD structure. $Vpx_{sm}$ is shown as a blue ribbon, DCAF1-CtD as a white surface, and SAMHD1-CtD as a red ribbon.

target(s) of Vpr are unknown, the ternary complex presented here provides a structural model for the design of therapeutic agents that target the Vpx– and Vpr–DCAF1 interaction.

## METHODS SUMMARY

Strep-II-tagged SAMHD1 residues 582–626 (SAMHD1-CtD) and glutathione *S*-transferase (GST)-tagged $Vpx_{sm}$ were expressed in *Escherichia coli* strain Rosetta 2 (DE3) and purified by affinity and gel-filtration chromatography. His-tagged DCAF1, residues 1058–1396 (DCAF1-CtD), was expressed in insect cells and purified using Ni-NTA Sepharose and gel-filtration chromatography. For structure determination, selenium was incorporated into $Vpx_{sm}$ and DCAF1-CtD by supplementing culture media with seleno-methionine. The ternary complex was assembled by incubating the components in a molar ratio of DCAF1-CtD:$Vpx_{sm}$:SAMHD1-CtD = 1:1.5:1.5 overnight on ice and purified by gel-filtration chromatography. Details of protein crystallization, data processing, structure determination and refinement are provided in Methods. Structure figures were prepared in PyMOL (http://pymol.sourceforge.net/). The degradation of tandem NLS–GFP fused to SAMHD1-CtD stably expressed in 293T cells was analysed by fluorescence microscopy and flow cytometry. SAMHD1 degradation by Vpx was assessed in 293T cells by transfection and immunoblotting. Infectivity of HIV-1 in U937 cells in the presence of SAMHD1 mutants was determined using two-colour flow cytometry.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 11 July; accepted 28 October 2013.**
**Published online 15 December 2013.**

1. Laguette, N. *et al.* SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* **474,** 654–657 (2011).
2. Hrecka, K. *et al.* Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* **474,** 658–661 (2011).
3. Baldauf, H. M. *et al.* SAMHD1 restricts HIV-1 infection in resting $CD4^+$ T cells. *Nature Med.* **18,** 1682–1689 (2012).
4. Descours, B. *et al.* SAMHD1 restricts HIV-1 reverse transcription in quiescent $CD4^+$ T-cells. *Retrovirology* **9,** 87 (2012).
5. Goldstone, D. C. *et al.* HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature* **480,** 379–382 (2011).
6. Lahouassa, H. *et al.* SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nature Immunol.* **13,** 223–228 (2012).
7. Lim, E. S. *et al.* The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. *Cell Host Microbe* **11,** 194–204 (2012).
8. Ahn, J. *et al.* HIV/simian immunodeficiency virus (SIV) accessory virulence factor Vpx loads the host cell restriction factor SAMHD1 onto the E3 ubiquitin ligase complex $CRL4^{DCAF1}$. *J. Biol. Chem.* **287,** 12550–12558 (2012).
9. Morellet, N., Bouaziz, S., Petitjean, P. & Roques, B. P. NMR structure of the HIV-1 regulatory protein VPR. *J. Mol. Biol.* **327,** 215–227 (2003).
10. Wei, W. *et al.* A novel DCAF1-binding motif required for Vpx-mediated degradation of nuclear SAMHD1 and Vpr-induced G2 arrest. *Cell. Microbiol.* **14,** 1745–1756 (2012).
11. Srivastava, S. *et al.* Lentiviral Vpx accessory factor targets VprBP/DCAF1 substrate adaptor for cullin 4 E3 ubiquitin ligase to enable macrophage infection. *PLoS Pathog.* **4,** e1000059 (2008).
12. Bergamaschi, A. *et al.* The human immunodeficiency virus type 2 Vpx protein usurps the CUL4A–DDB1 DCAF1 ubiquitin ligase to overcome a postentry block in macrophage infection. *J. Virol.* **83,** 4854–4860 (2009).
13. DeLucia, M., Mehrens, J., Wu, Y. & Ahn, J. HIV-2 and SIVmac accessory virulence factor Vpx down-regulates SAMHD1 enzyme catalysis prior to proteasome-dependent degradation. *J. Biol. Chem.* **288,** 19116–19126 (2013).
14. Bennett, E. J., Bence, N. F., Jayakumar, R. & Kopito, R. R. Global impairment of the ubiquitin-proteasome system by nuclear or cytoplasmic protein aggregates precedes inclusion body formation. *Mol. Cell* **17,** 351–365 (2005).
15. Scrima, A. *et al.* Structural basis of UV DNA-damage recognition by the DDB1–DDB2 complex. *Cell* **135,** 1213–1223 (2008).
16. Fischer, E. S. *et al.* The molecular basis of $CRL4^{DDB2/CSA}$ ubiquitin ligase architecture, targeting, and activation. *Cell* **147,** 1024–1039 (2011).
17. Yan, J. *et al.* Tetramerization of SAMHD1 is required for biological activity and inhibition of HIV infection. *J. Biol. Chem.* **288,** 10406–10417 (2013).
18. Angers, S. *et al.* Molecular architecture and assembly of the DDB1–CUL4A ubiquitin ligase machinery. *Nature* **443,** 590–593 (2006).
19. Li, T., Robert, E. I., van Breugel, P. C., Strubin, M. & Zheng, N. A promiscuous α-helical motif anchors viral hijackers and substrate receptors to the CUL4–DDB1 ubiquitin ligase machinery. *Nature Struct. Mol. Biol.* **17,** 105–111 (2010).
20. Fregoso, O. I. *et al.* Evolutionary toggling of Vpx/Vpr specificity results in divergent recognition of the restriction factor SAMHD1. *PLoS Pathog.* **9,** e1003496 (2013).
21. Wen, X., Duus, K. M., Friedrich, T. D. & de Noronha, C. M. The HIV1 protein Vpr acts to promote G2 cell cycle arrest by engaging a DDB1 and Cullin4A-containing ubiquitin ligase complex using VprBP/DCAF1 as an adaptor. *J. Biol. Chem.* **282,** 27046–27057 (2007).
22. Schröfelbauer, B., Hakata, Y. & Landau, N. R. HIV-1 Vpr function is mediated by interaction with the damage-specific DNA-binding protein DDB1. *Proc. Natl Acad. Sci. USA* **104,** 4130–4135 (2007).
23. Le Rouzic, E. *et al.* HIV1 Vpr arrests the cell cycle by recruiting DCAF1/VprBP, a receptor of the Cul4–DDB1 ubiquitin ligase. *Cell Cycle* **6,** 182–188 (2007).

24. Hrecka, K. *et al.* Lentiviral Vpr usurps Cul4–DDB1[VprBP] E3 ubiquitin ligase to modulate cell cycle. *Proc. Natl Acad. Sci. USA* **104,** 11778–11783 (2007).
25. Barnitz, R. A., Chaigne-Delalande, B., Bolton, D. L. & Lenardo, M. J. Exposed hydrophobic residues in human immunodeficiency virus type 1 Vpr helix-1 are important for cell cycle arrest and cell death. *PLoS ONE* **6,** e24924 (2011).

**Acknowledgements** We thank L. Haire and R. Ogrodowicz for help with crystallization, I. Jones for the provision of the modified *A. californica* baculovirus bacmid, BAC10:1629$_{KO}$, S. Smerdon and S. Gamblin for comments on the manuscript. We gratefully acknowledge the Diamond Light Source for synchrotron access (grant no. 7707). This work was supported by the UK Medical Research Council, file references U117565647 (I.A.T.), U117592729 (K.N.B.) and U117512710 (J.P.S.); the Wellcome Trust, ref. 084955 (K.N.B.); and by an EMBO long-term fellowship co-funded by the European Commission Marie Curie Actions (EMBOCOFUND2010, GA-2010-267146) (D.S.).

**Author Contributions** D.S., H.C.T.G., V.C.B., E.C. and P.A.W. performed experiments. D.S., H.C.T.G., V.C.B., E.C., P.A.W., J.P.S., K.N.B. and I.A.T. contributed to experimental design, data analysis and manuscript writing.

**Author Information** The coordinates of DCAF1-CtD–Vpx$_{sm}$–SAMHD1-CtD have been deposited in the PDB under accession 4CC9. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.A.T. (itaylor@nimr.mrc.ac.uk).

## METHODS

**Cloning, protein expression and purification.** The DNA sequence coding for human SAMHD1 residues Q582–M626 (SAMHD1-CtD) was amplified by PCR from a cDNA template using the oligonucleotide primer sequences: forward, GGC GGATCCTCAGGATGGCGATGTTATAGCCCC; reverse, GGCGCGGCCGCT CATCACATTGGGTCATCTTTAAAAAGCTGG. The PCR product was gel purified and ligated into the pET52b plasmid (Merck Millipore) using standard restriction enzyme cloning to generate an N-terminal Strep-II-tagged fusion protein. N-terminally GST-tagged sooty mangabey Vpx residues M1–A112 ($Vpx_{sm}$) in the pET49b plasmid (Merck Millipore) was a gift from D. Goldstone. Human DCAF1 residues A1058–E1396 (DCAF1-CtD) were amplified by PCR from a cDNA template using the oligonucleotide primer sequences: forward, GGCCCATGGCATC ATTTCCAAAGTATGGAGGGG; reverse, GGCGAGCTCCTCTGCCAGACGC TGCCTGCC. The PCR product was gel purified and ligated into the pTriEx-6 plasmid (Merck Millipore) using standard restriction enzyme cloning to generate a C-terminal 10×His-tagged fusion protein. Insert sequences were verified by DNA sequencing.

SAMHD1-CtD and GST–$Vpx_{sm}$ were expressed in the *E. coli* strain Rosetta 2 (DE3) (Merck Millipore). Bacterial cultures were grown in an incubator shaker at 37 °C. Protein expression was induced by the addition of 0.1 mM isopropyl-β-D-thiogalactoside (IPTG) at $A_{600\,nm} = 0.5$. Afterwards, cultures were cooled to 18 °C and grown for a further 20 h. To produce seleno-methionine (SeMet)-labelled $Vpx_{sm}$, cells were grown to $A_{600\,nm} = 0.5$ at 37 °C in M9 minimal medium. Then, an amino acid supplement (L-lysine, L-phenylalanine, L-threonine to a final concentration of 100 mg $l^{-1}$, L-isoleucine, L-leucine, L-valine and L-SeMet to a final concentration of 40 mg $l^{-1}$, respectively) was added to inhibit endogenous methionine biosynthesis and start SeMet incorporation[26]. Fifteen minutes after addition, the culture was cooled to 18 °C. Protein expression was induced by the addition of 0.5 mM IPTG, and cells were grown for a further 20 h.

Cultures were centrifuged for 20 min at 4,500*g* and 4 °C. Cell pellets were resuspended in 30 ml lysis buffer (50 mM Tris-Hcl pH 7.8, 500 mM NaCl, 4 mM $MgCl_2$, 0.5 M TCEP, 1× EDTA-free mini complete protease inhibitors (Roche), 0.1 U $ml^{-1}$ Benzonase (Novagen) per pellet of 1 litre bacteria culture. Cells were lysed in an EmulsiFlex-C5 (Avestin). The lysate was cleared by centrifugation for 1 h at 48,000*g* and 4 °C. All further purification steps were performed at 4 °C or on ice. The cleared lysates were applied to 10 ml StrepTactin (IBA, for SAMHD1-CtD) or to 10 ml Glutathione-Sepharose (GE Healthcare, for GST–$Vpx_{sm}$) columns. Columns were washed with 600 ml wash buffer (50 mM Tris-HCl pH 7.8, 500 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP). Column-immobilized GST–$Vpx_{sm}$ was additionally washed with 250 ml of wash buffer supplemented with 5 mM ATP, followed by 250 ml wash buffer supplemented with 1% CHAPS. SAMHD1-CtD was eluted from the column with 25 mM ammonium bicarbonate pH 7.5, 2.5 mM desthiobiotin. The elution peak was lyophilized, resuspended in 500 µl of 25 mM ammonium bicarbonate pH 7.5 and applied to a Superdex 75 16/60 gel-filtration column (GE Healthcare) equilibrated with 25 mM ammonium bicarbonate pH 7.5. The peak fractions were pooled, lyophilized and resuspended in 10 mM Bis-Tris propane pH 8.5, 150 mM NaCl, 4 mM $MgCl_2$, 0.5 M TCEP. Small aliquots at a concentration of approximately 2 mM were flash frozen in liquid nitrogen and stored at −80 °C. GST–$Vpx_{sm}$ was eluted from the column with wash buffer supplemented with 20 mM glutathione. The elution peak was concentrated to 5 ml and incubated overnight with 1 mg GST-3C protease. Cleaved $Vpx_{sm}$ was further purified on a Superdex 200 26/60 gel-filtration column (GE Healthcare) equilibrated with 10 mM Tris-HCl pH 7.8, 150 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP. Peak fractions containing $Vpx_{sm}$ were concentrated to approximately 20 mg $ml^{-1}$ and flash frozen in liquid nitrogen in small aliquots for further storage at −80 °C. Seleno-methionine-substituted $Vpx_{sm}$ was purified in the same way.

For the production of DCAF1-CtD, recombinant baculovirus was generated by co-transfecting Sf9 cells with pTriEx-DCAF1-CtD and linearized $BAC10{:}1629_{KO}$ (ref. 27). Sf9 cells were cultured in SF900 II serum-free medium (Invitrogen) at 28 °C. In a typical preparation, 2 litres of Sf9 cells at $2 \times 10^6$ cells $ml^{-1}$ density were infected with 4 ml of high-titre DCAF1-CtD virus for 48 h. For structure determination, selenium was incorporated into DCAF1-CtD by supplementing 921Δ series medium (Expression Systems, LLC) with 50 mg $l^{-1}$ seleno-methionine.

Sf9 cultures were centrifuged for 20 min at 4,500*g* and 4 °C. Pellets were resuspended in 30 ml lysis buffer (50 mM Tris-HCl pH 7.8, 500 mM NaCl, 4 mM $MgCl_2$, 30 mM imidazole-HCl pH 7.8, 0.5 mM TCEP, 1× EDTA-free mini complete protease inhibitors (Roche), 0.1 U $ml^{-1}$ Benzonase (Novagen)) per pellet of 1 litre culture. Cells were lysed in an EmulsiFlex-C5 (Avestin). The lysate was cleared by centrifugation for 1 h at 48,000*g* and 4 °C. All further purification steps were performed at 4 °C or on ice. The cleared lysate was applied to a 1 ml Ni-NTA Sepharose column (GE Healthcare). The column was washed with 500 ml wash buffer (50 mM Tris-HCl pH 7.8, 500 mM NaCl, 4 mM $MgCl_2$, 30 mM imidazole-HCl pH 7.8, 0.5 mM TCEP). Protein was eluted with wash buffer supplemented with 300 mM imidazole pH 7.8. Peak fractions were pooled, concentrated to 2 ml and applied on a Superdex 200 16/60 gel-filtration column (GE Healthcare) equilibrated with 10 mM Tris-HCl pH 7.8, 150 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP. Peak fractions containing DCAF1-CtD were pooled, concentrated to approximately 20 mg $ml^{-1}$ and flash frozen in liquid nitrogen in small aliquots for further storage at −80 °C. Seleno-methionine-substituted DCAF1-CtD was purified in the same way.

**Assembly of the ternary complex.** Initially, the complex was prepared by mixing 100 µg of DCAF1-CtD, $Vpx_{sm}$ and SAMHD1-CtD in 10 mM Bis-Tris propane pH 8.5, 150 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP, followed by incubation overnight on ice. Complex formation was assessed by applying the mixture on an analytical Superdex 75 10 300 GL gel-filtration column (GE Healthcare), equilibrated with 10 mM Bis-Tris propane pH 8.5, 150 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP and SDS–PAGE analysis (molecular weight of markers 97, 66.3, 55.4, 36.5, 31.0, 21.5, 14.4, 6.0 and 3.5 kDa). After optimization, for large-scale preparation of the complex, DCAF1-CtD, $Vpx_{sm}$ and SAMHD1-CtD were mixed in a molar ratio of 1:1.5:1.5 in 0.1 M Bis-Tris propane pH 8.5, 150 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP and incubated overnight on ice. The mixture was applied to a Superdex 75 16/60 gel-filtration column (GE Healthcare), equilibrated with 10 mM Bis-Tris propane pH 8.5, 150 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP. Peak fractions containing all components were pooled, concentrated to approximately 5 mg $ml^{-1}$ and directly used for crystallization experiments, after addition of an equimolar amount of SAMHD1-CtD.

**Crystallization and structure determination.** The native DCAF1-CtD–$Vpx_{sm}$–SAMHD1-CtD complex was crystallized by vapour diffusion using an Oryx crystallization robot (Douglas Instruments). Crystals were obtained from 0.1 µl droplets containing an equal volume of 2.5 mg $ml^{-1}$ protein complex in 10 mM Bis-Tris propane pH 8.5, 150 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP mixed with 0.2 M magnesium chloride, 0.1 M HEPES pH 7.5, 15% PEG 400. For data collection, crystals were adjusted to 25% PEG 400 and cryo-cooled in liquid nitrogen. Native crystals diffracted up to 2.5 Å resolution on beamline I03 at the Diamond Light Source, and belong to the space group $P2_12_12_1$ with cell dimensions of $a = 73.79$, $b = 82.03$, $c = 113.29$, with a single copy of the complex in the asymmetric unit. Crystals of the complex containing seleno-methionine-substituted DCAF1-CtD and $Vpx_{sm}$ were grown using the vapour diffusion method by mixing 1 µl complex at 5 mg $ml^{-1}$ with 1 µl reservoir solution containing 0.4 M magnesium sulphate, 0.1 M MES pH 6.5. Crystals were adjusted to 25% glycerol and cryo-cooled in liquid nitrogen. A SAD data set was collected on beamline I04 at the Diamond Light Source at a wavelength of 0.97972 Å, corresponding to the anomalous $f''$ peak wavelength for selenium determined from an X-ray fluorescence scan. The crystal diffracted up to 3.5 Å resolution and was nearly isomorphous to the native crystal with the same space group and unit cell dimensions of $a = 74.25$, $b = 82.88$, $c = 115.56$. Data were reduced using the XDS suite[28]. An initial set of five selenium sites was found using the programs SHELXC and SHELXD[29]. These sites were used as input for the program autoSHARP[30], which added two more sites and performed density modification, leading to an interpretable electron density map. A nearly complete model for $Vpx_{sm}$ and a partial polyalanine trace of DCAF1-CtD were located in this initial map using the buccaneer chain-tracing program[31]. Completion of the polyalanine trace of DCAF1-CtD and placement of the protein backbone of SAMHD1-CtD was then undertaken manually in Coot[32]. A round of refinement against native data recorded to 2.5 Å produced an improved map that allowed further building of the SAMHD1-CtD and DCAF1-CtD side chains. Subsequent incorporation of a $Zn^{2+}$ ion and ligands in Coot combined with positional, real-space, individual B-factor and TLS refinement in phenix.refine[33] produced a final model for residues 1073–1314, 1328–1392 of DCAF1-CtD, 5–90, 100–111 of $Vpx_{sm}$ and 606–624 of SAMHD1-CtD with $R/R_{free}$ factors of 17.8%/21.6%. In the model, 97.1% of residues have backbone dihedral angles in the favoured region of the Ramachandran plot, 2.66% fall in the allowed regions and 0.24% are outliers. Details of data collection and refinement statistics are presented in Extended Data Table 1.

**Multiple sequence alignment.** Amino acid sequences were aligned using the ClustalW server and adjusted manually. Uniprot accessions are as follows. SAMHD1: *Homo sapiens*, Q9Y3Z3; *Pan troglodytes*, H6WE97; *Cercocebus atys*, H6WEA6; *Cercocebus torquatus*, H6WEA7; *Macaca mulatta*, F7CA64; *Callithrix jacchus*, F7IGP7; *Canis familiaris*, E2QTR2; *Mus musculus*, Q60710; *Rattus norvegicus*, D3Z898; *Gallus gallus*, Q5ZJL9; *Xenopus laevis*, Q6INN8; *Danio rerio*, Q502K2; *Dictyostelium discoideum*, B0G107. Vpx: *Cercocebus atys* (isolate sm), P19508; *Macaca mulatta* (isolate mac), P05917; *Cercocebus torquatus* (isolate rcm), NCBI GenBank HM803689; *Mandrillus sphinx* (isolate mnd-2), NCBI reference sequence NP_758889; HIV-2A (isolate rod), P06939; HIV-2B (isolate eho), Q89721. Vpr: HIV-2A (isolate rod), P06938; HIV-2B (isolate eho), P0C1P6; HIV-1MA (isolate mal), P05955; HIV-1MB (isolate yu-2), P35967.

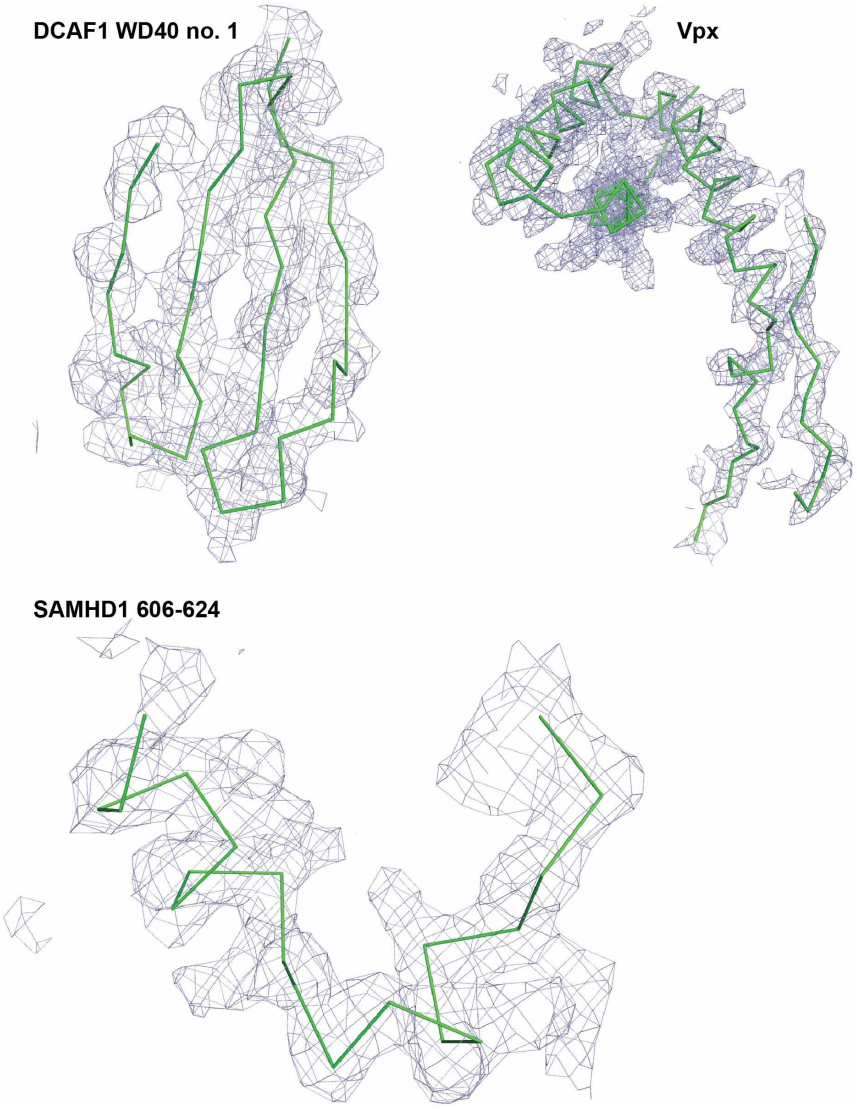
**Modelling of the cullin/RING ubiquitin ligase assembly.** Structural superposition was carried out using SSM[34] implemented in PDBefold (http://www.ebi.ac.uk/msd-srv/ssm/). DCAF1-CtD was first superposed with the DDB2 β-propeller in the DDB1–DDB2 structure (PDB accession 3EI3) such that the N-terminal residues and first β-strand of DCAF1-CtD align with the equivalent sequence just C terminal to the helical DDB1-binding element of DDB2, termed the H-box. The second β-propeller domain of DDB1 (BPB), which serves as anchor for the CUL4 scaffold (wheat half-transparent surface), is mobile with respect to the DDB1 BPA and BPC. The two most extreme BPB conformations available in the PDB database were used to model the rotational range of the CUL4 arm with respect to the DDB1 BPB (PDB accessions 2HYE, to the left, and 3I7H, to the right).

**Restriction assay.** SAMHD1 wild-type sequence was inserted into pLGatewayIeYFP[35] and mutations created by PCR-based site-directed mutagenesis. MoMLV-based YFP vectors were made by co-transfecting pVSV-G (ref. 36), pKB4 (ref. 37) and pLgatewaySN_SAMHD1 (wild type or mutants) into 293T cells, harvesting 48 h after transfection. HIV-1–GFP was made as described earlier by co-transfecting pVSV-G, p8.91 (ref. 38) and pCSGW (ref. 39). U937 cells[40] were maintained in RPMI plus [L]glutamine (GIBCO) with 10% fetal calf serum (Biosera), penicillin and streptomycin. Cells ($1 \times 10^6$) were transduced by spinoculation at 1,700 r.p.m. for 90 min with 0.5 ml neat virus in the presence of 1 μg ml$^{-1}$ polybrene. Cells were differentiated by addition of 100 nM phorbol myristate acetate for 72 h and infected with HIV-1–GFP. Restriction was assessed by two-colour flow cytometry after 72 h.

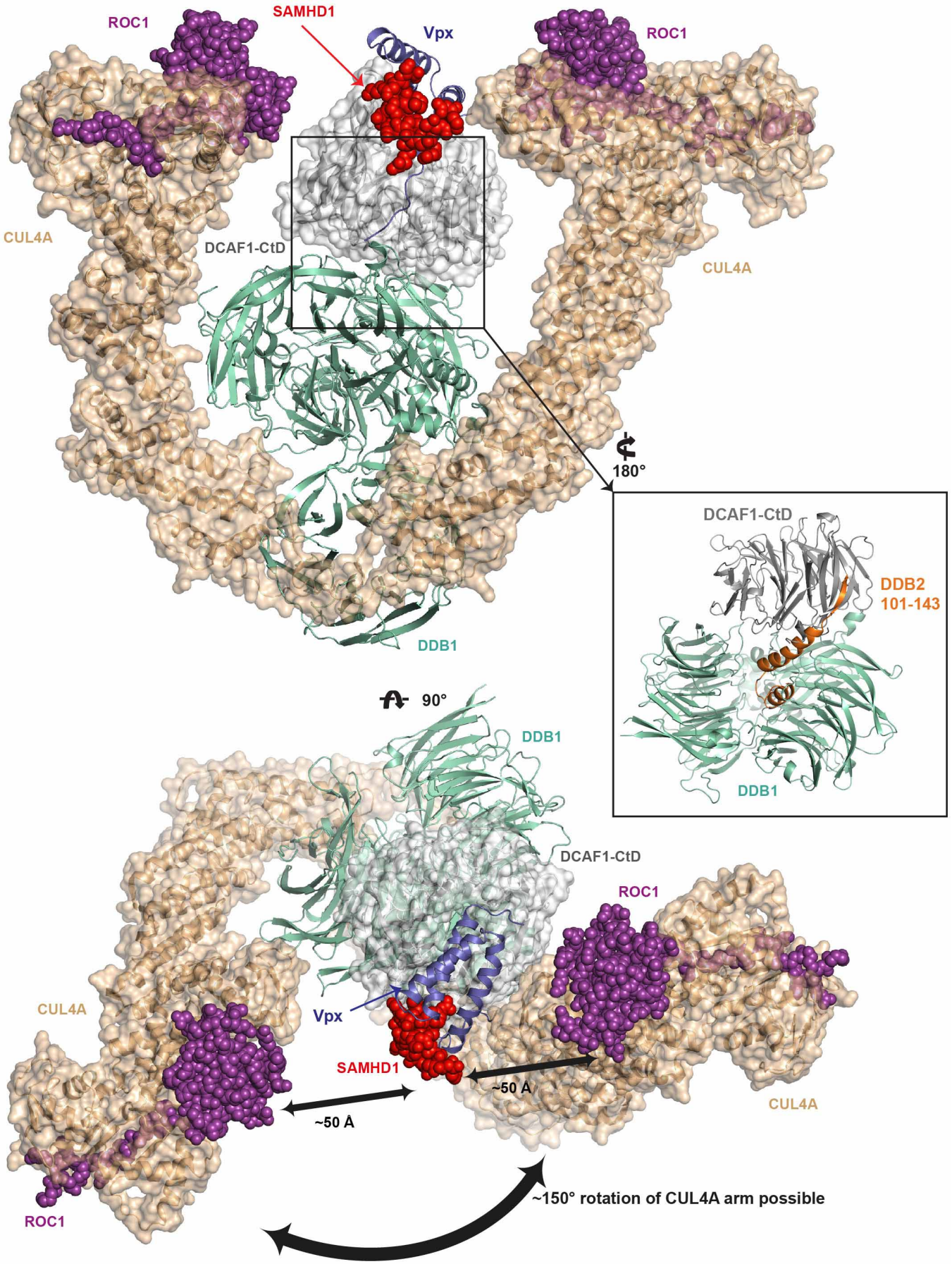
**Transfection and immunoblotting.** SAMHD1 and HIV-2 Vpx sequences were amplified from pLgatewaySN_SAMHD1 and pIRES2-EGFP-Vpx (gift from M. Stevenson), respectively, and cloned into pCMS28 (ref. 41). Point mutations were created by PCR-based site-directed mutagenesis. 293T cells were co-transfected with 2 μg each of pCMS28-SAMHD1 and -Vpx, wild type, mutant proteins or empty vector. Cells were harvested 24 h after transfection and analysed by SDS–PAGE followed by immunoblotting with anti-SAMHD1 3295 (generated in house), anti-HIV-2 Vpx (hybridoma 6D2.6 supernatant, NIH AIDS Reagent Program 2739) and anti-14-3-3 (C-16, Santa Cruz sc-731).

**Degron assay.** A reporter construct comprising two copies of GFP with NLS fused to residues 600–626 of SAMHD1 was synthesized (GenArt; NLS–GFP–SAMHD1-CtD; Fig. 3a). NLS–GFP–SAMHD1-CtD was inserted into pCMS28 using BglII and EcoRI. All mutants were created by PCR-based site-directed mutagenesis. Puromycin virus-like particles were generated by co-transfection of pVSVG, pKB4 and pCMS28-NLS-GFP-SAMHD1-CtD wild type or mutant as above. The stable cell lines were made by transduction of 293T cells followed by puromycin selection. VLPs SIV Vpx$^+$ and Vpx$^-$ were generated by co-transfection of pVSVG and a SIV Gag-Pol expression plasmid pSIV3$^+$ (ref. 42) or pSIV3$^+_{vpx-}$ (gift from C. Goujon and A. Cimarelli). Viral titres were quantified using a modified ELISA for reverse transcriptase activity (Cavidi). 293T cells stably expressing SAMHD1–GFP wild type or mutant were seeded at $5 \times 10^4$ cells per well in a 24-well plate 1 day before infection. Cells were infected with twofold serial dilutions of replication defective SIV Vpx$^+$ virus in the presence of 1 μg ml$^{-1}$ polybrene. For each experiment, SIV Vpx$^-$ VLPs and proteasome inhibitor MG132 (25.2 μM) were included as negative controls. After 48 h, cells were harvested and the percentage of GFP-positive cells was determined by flow cytometry using a FACSVerse analyser (BD Biosciences). In parallel, GFP expression was analysed by microscopy using an inverted fluorescent microscope (LEICA).

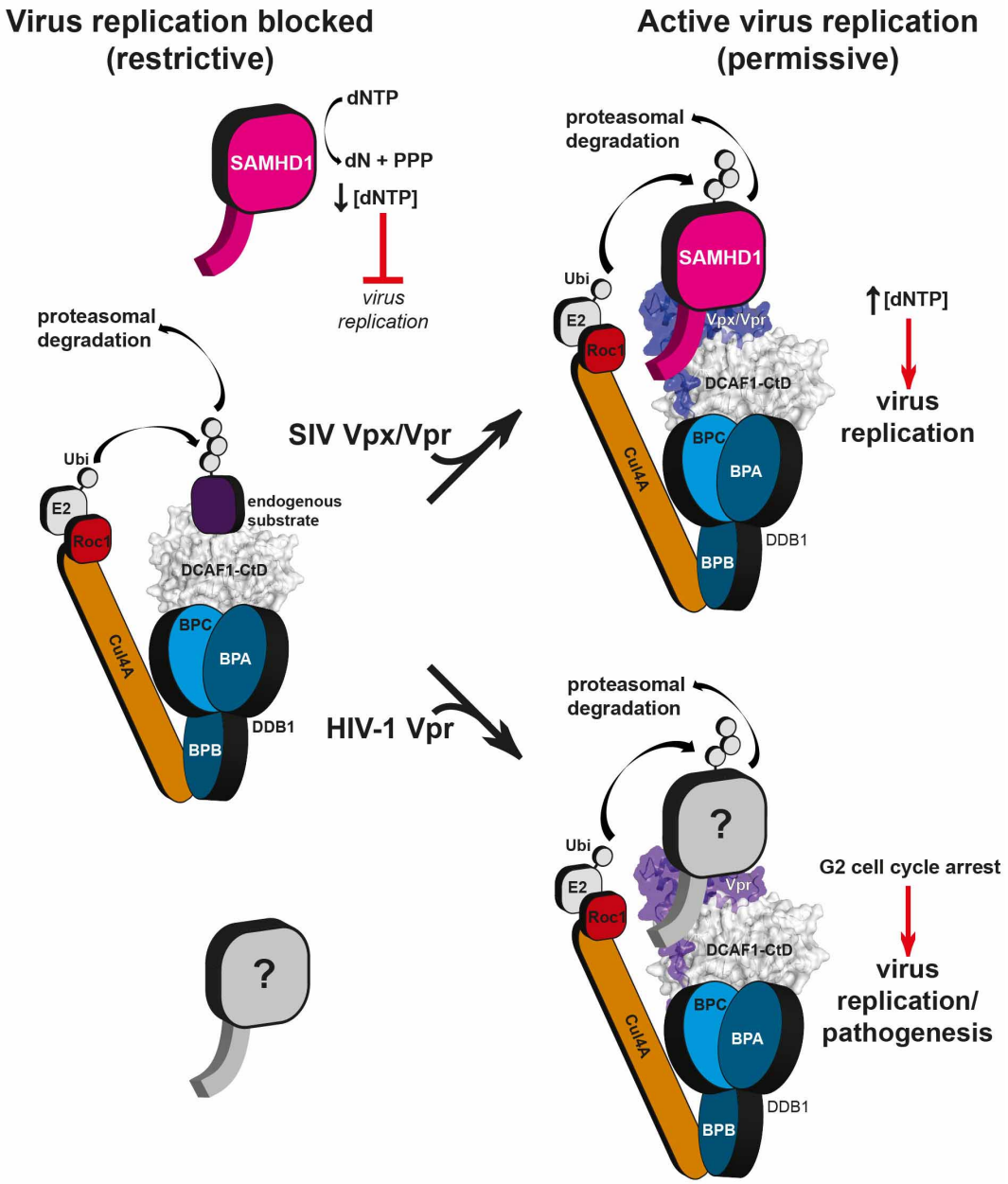

26. Doublié, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276,** 523–530 (1997).
27. Zhao, Y., Chapman, D. A. & Jones, I. M. Improving baculovirus recombination. *Nucleic Acids Res.* **31,** e6 (2003).
28. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* **66,** 133–144 (2010).
29. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64,** 112–122 (2008).
30. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364,** 215–230 (2007).
31. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62,** 1002–1011 (2006).
32. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
33. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68,** 352–367 (2012).
34. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* **60,** 2256–2268 (2004).
35. Yap, M. W., Nisole, S., Lynch, C. & Stoye, J. P. Trim5α protein restricts both HIV-1 and murine leukemia virus. *Proc. Natl Acad. Sci. USA* **101,** 10786–10791 (2004).
36. Bock, M., Bishop, K. N., Towers, G. & Stoye, J. P. Use of a transient assay for studying the genetic determinants of Fv1 restriction. *J. Virol.* **74,** 7422–7430 (2000).
37. Groom, H. C. *et al.* Absence of xenotropic murine leukaemia virus-related virus in UK patients with chronic fatigue syndrome. *Retrovirology* **7,** 10 (2010).
38. Naldini, L. *et al. In vivo* gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272,** 263–267 (1996).
39. Bainbridge, J. W. *et al. In vivo* gene transfer to the mouse eye using an HIV-based lentiviral vector; efficient long-term transduction of corneal endothelium and retinal pigment epithelium. *Gene Ther.* **8,** 1665–1668 (2001).
40. Sundström, C. & Nilsson, K. Establishment and characterization of a human histiocytic lymphoma cell line (U-937). *Int. J. Cancer* **17,** 565–577 (1976).
41. Gallois-Montbrun, S. *et al.* Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules. *J. Virol.* **81,** 2165–2178 (2007).
42. Nègre, D. *et al.* Characterization of novel safe lentiviral vectors derived from simian immunodeficiency virus (SIVmac251) that efficiently transduce mature human dendritic cells. *Gene Ther.* **7,** 1613–1623 (2000).




**Extended Data Figure 1 | Experimental electron density.** Experimental electron density observed after solvent flattening for DCAF1, $Vpx_{sm}$ and SAMHD1 is shown as light blue wireframe, contoured at 1σ. The backbone Cα traces of the final refined protein structure are shown in green ribbon representation.



**Extended Data Figure 2 | Model of the hijacked CUL4A–DDB1–ROC1 E3 ubiquitin ligase complex.** $Vpx_{sm}$ (blue) bound to the substrate-specificity module DCAF1-CtD (grey) is shown in cartoon representation as in Fig. 1c. SAMHD1-CtD is shown as red spheres. The DDB1 adaptor is shown as green cartoon. The inset to the right shows the superposition of the DDB2 helical hairpin (H-box, orange), which inserts into the binding groove created by DDB1 β-propellers 1 and 3 (BPA, BPC), and the N terminus of DCAF1-CtD presented in this study. The CUL4A scaffold is represented as an orange semi-transparent surface, the ROC1 RING module as purple spheres. Owing to the conformational freedom of the DDB1–CUL4A connection, the two most extreme conformations of CUL4A with respect to DDB1 available in the PDB were modelled. See Methods for modelling procedures and PDB accessions. The model clearly shows that in both extreme CUL4 conformations, the ROC1 RING finger (purple spheres) is well positioned to reach the SAMHD1 protein, which would be attached at the N-terminal end of the SAMHD1-CtD. The SAMHD1 globular fold is probably mobile with respect to the fixed position of SAMHD1-CtD owing to the flexibility of the sequence stretch between SAMHD1 residues 583 (the last ordered residue of PDB accession 3U1N[5]) and 606 (the first ordered residue of SAMHD1-CtD presented here).




**Extended Data Figure 3 | Analogous mechanism of restriction factor counteraction in SIV/HIV-2 and HIV-1 Vpx and Vpr.** SAMHD1 provides a potent post-entry block against immunodeficiency viruses in non-cycling cells. Its dNTP-triphosphohydrolase activity lowers the cellular dNTP pool, preventing viral reverse transcription. HIV-2/SIVs use their Vpx and Vpr accessory proteins to modify the host cell's CUL4A–DDB1–DCAF1 ubiquitin ligase specificity towards SAMHD1, resulting in its proteasomal degradation and ultimately raising dNTP levels, making the cells permissive to viral replication. Sequence similarity and comparative functional analysis suggest that the ancestral HIV-1 accessory protein Vpr uses a similar mechanism to exploit the CUL4A–DDB1–DCAF1 system to induce proteasomal degradation of an as yet undiscovered cellular factor whose absence causes cell cycle arrest in the G2 phase, promoting viral replication and pathogenesis *in vivo*.

**Extended Data Table 1 | Data collection and refinement statistics**

| | Se peak | Native |
|---|---|---|
| **Data collection** | | |
| Space group | $P2_12_12_1$ | $P2_12_12_1$ |
| Cell dimensions | | |
| *a*, *b*, *c* (Å) | 74.25, 82.88, 115.56 | 73.79, 82.03, 113.29 |
| α, ß, γ (°) | 90, 90, 90 | 90, 90, 90 |
| Resolution (Å) | 30-3.5 (3.71-3.5)* | 30-2.46 (2.61-2.46) |
| $R_{sym}$ or $R_{merge}$ | 5.3 (9.9) | 9.1 (52.9) |
| $I$ / σ($I$) | 20.0 (11.5) | 15.9 (3.6) |
| Completeness (%) | 98.6 (96) | 99.3 (97.8) |
| Redundancy | 3.5 (3.5) | 6.5 (6.4) |
| | | |
| FOM (acentric/centric) | 0.27/0.06 | |
| Anomalous phasing power | 0.99 | |
| | | |
| **Refinement** | | |
| Resolution (Å) | | 30-2.47 |
| No. reflections | | 25518 |
| $R_{work}$ / $R_{free}$ | | 17.6 / 21.6 |
| No. atoms | | |
| Protein | | 3407 |
| Ligand PG4 | | 26 |
| Ligand Zn | | 1 |
| Water | | 25 |
| *B*-factors (Å$^2$) | | |
| Protein | | 35.9 |
| Ligand PG4 | | 62.5 |
| Ligand Zn | | 43.9 |
| Water | | 27.8 |
| r.m.s.d. | | |
| Bond lengths (Å) | | 0.008 |
| Bond angles (°) | | 1.168 |

r.m.s.d., root mean squared deviation.
* Values in parentheses are for highest-resolution shell.

**Extended Data Table 2 | Intermolecular interactions**

(**A**) Vpx-DCAF1 (1987Å$^2$)

| Vpx | Region | Contact type | DCAF1 | Region |
|---|---|---|---|---|
| E6 | Nt | HB | S1102, R1104$_{MC}$, R1106$_{MC}$ | WD40 1 |
| I8 | Nt | HI | F1107, L1119 | WD40 1 |
| P9$_{MC}$ | Nt | HB | Y1131 | WD40 1 |
| N12 | Nt | HB | S1130$_{MC}$, N1132$_{MC}$ | WD40 1 |
| S13 | Nt | HB | S1130, N1132 | WD40 1 |
| W24 | α1 | HB | N1135 | WD40 1-2 |
| T28 | α1 | HB | N1135 | WD40 1-2 |
| I32 | α1 | HI | W1156 | WD40 2 |
| A35 | α1 | HI | W1156 | WD40 2 |
| L44 | α2 | HI | M1375 | WD40 7 |
| L48 | α2 | HI | M1375 | WD40 7 |
| R51 | α2 | HB | D1376$_{MC}$ | WD40 7 |
| W56 | α2 | HI | L1378 | WD40 7 |
| Y66 | α3 | HB | D1092 | WD40 7-1 |
| Y69 | α3 | HB | E1091 | WD40 7-1 |
| R70 | α3 | SB | E1093 | WD40 7-1 |
| L74 | α3 | HI | M1375, L1378 | WD40 7 |
| Q76 | α3 | HB | N1135, W1156 | WD40 1-2, 2 |
| K77 | α3 | SB, HB | E1093, M1380$_{MC}$ | WD40 7-1, 7 |
| A78 | α3 | HI | M1375 | WD40 7-1 |
| F80 | α3 | HI | F1330, F1355 | WD40 6, 7 |
| M81 | α3 | HI | F1330, F1355 | WD40 6, 7 |
| K84 | α3 | HI, HB | L1313, F1330, F1355$_{MC}$ | WD40 6, 7 |
| K85$_{MC}$ | α3 | HB | K1224, R1225 | WD40 4 |
| G86$_{MC}$ | α3 | HB | K1224 | WD40 4 |

(**B**) Vpx-SAMHD1 (703Å$^2$)

| Vpx | Region | Contact type | SAMHD1 |
|---|---|---|---|
| G14$_{MC}$ | Nt | HB | L610$_{MC}$, R611$_{MC}$ |
| E15 | Nt | SB | R609 |
| E16 | Nt | SB | R617 |
| I18$_{MC}$ | α1 | HB | N606$_{MC}$, T608$_{MC}$ |
| G19$_{MC}$ | α1 | HB | N606$_{MC}$ |
| A21 | α1 | HI | R617, L620 |
| F22 | α1 | HI | L620, F621 |
| W24 | α1 | HI | R617, F621 |
| L25 | α1 | HI | F621 |
| M62 | α2- α3 | HI | K622 |
| S63 | α2- α3 | HB | F621$_{MC}$ |
| Y66 | α3 | HB, HI | F621$_{MC}$, K622 |
| Y69 | α3 | HI | V618, F621 |

(**C**) DCAF1-SAMHD1 (207Å$^2$)

| DCAF1 | Region | Contact type | SAMHD1 |
|---|---|---|---|
| D1092 | WD40 7-1 | SB | K622 |

HB, hydrogen bond; HI, hydrophobic interaction; SB, salt bridge. Bi-functional residues are highlighted in red.

**Extended Data Table 3 | Vpx and Vpr mutations**

| SIVsm Vpx residue | HIV-1 Vpr residue | SIVsm Vpx phenotype | HIV-1 Vpr phenotype | Reference |
|---|---|---|---|---|
| N12 | n.a. | N12A/no SAMHD1 degradation | n.a. | PMID 22362772 |
| S13, T17, T28 | n.a. | S13A,T17A,T28A/no SAMHD1 degradation, reduced HIV-1 infectivity | n.a. | PMID 23076149 |
| E15 | n.a. | E15A/no SAMHD1 degradation | n.a. | PMID 22362772 |
| E16 | n.a. | E16A/no SAMHD1 degradation | n.a. | PMID 22362772 |
| T17 | n.a. | T17A/no SAMHD1 degradation | n.a. | PMID 22362772 |
| W24 | W18 | W24A/no SAMHD1 degradation | W18A/reduced cell cycle arrest | PMID 22776683<br>PMID 21949789 |
| I32 | n.a. | I32S/no DCAF1 binding, no SAMHD1 degradation | n.a. | PMID 22776683 |
| H39 (Zn finger) | n.a. | H39A/no DCAF1 binding, reduced SAMHD1 degradation, reduced HIV-1 infectivity | n.a. | PMID 18829761<br>PMID 22776683 |
| Q76 | Q65 | Q76R/no DCAF1 binding, reduced HIV-2 infectivity<br>Q76A/no DCAF1 binding, reduced HIV-1 infectivity, reduced SIV infectivity, SIV RT activity compromised | Q65R/no DCAF1 binding, no cell cycle arrest | PMID 21720370<br>PMID 19264781<br>PMID 19923175<br>PMID 18464893<br>PMID 17314515 |
| K77 | n.a. | K77A/ no DCAF1 binding, reduced HIV-2 infectivity | n.a. | PMID 19264781 |
| F80 | n.a. | F80A/no DCAF1 binding, reduced SIV infectivity, RT activity compromised | n.a. | PMID 18464893 |
| H82 (Zn finger) | H71 | n.a. | H71R/no DDB1-DCAF1 binding, no cell cycle arrest | PMID 17609381 |

LETTER

doi:10.1038/nature12771

# Efficient ethanol production from brown macroalgae sugars by a synthetic yeast platform

Maria Enquist-Newman[1]*, Ann Marie E. Faust[1]*, Daniel D. Bravo[1]*, Christine Nicole S. Santos[1]†, Ryan M. Raisner[1], Arthur Hanel[1], Preethi Sarvabhowman[1], Chi Le[1], Drew D. Regitsky[1]†, Susan R. Cooper[1], Lars Peereboom[2], Alana Clark[1], Yessica Martinez[1], Joshua Goldsmith[1], Min Y. Cho[1], Paul D. Donohoue[1], Lily Luo[1], Brigit Lamberson[1], Pramila Tamrakar[1], Edward J. Kim[1], Jeffrey L. Villari[1], Avinash Gill[1]†, Shital A. Tripathi[1]†, Padma Karamchedu[1], Carlos J. Paredes[1], Vineet Rajgarhia[1]†, Hans Kristian Kotlar[3], Richard B. Bailey[1], Dennis J. Miller[2], Nicholas L. Ohler[1], Candace Swimmer[1] & Yasuo Yoshikuni[1,4,5]

**The increasing demands placed on natural resources for fuel and food production require that we explore the use of efficient, sustainable feedstocks such as brown macroalgae. The full potential of brown macroalgae as feedstocks for commercial-scale fuel ethanol production, however, requires extensive re-engineering of the alginate and mannitol catabolic pathways[1–3] in the standard industrial microbe *Saccharomyces cerevisiae*. Here we present the discovery of an alginate monomer (4-deoxy-L-erythro-5-hexoseulose uronate, or DEHU) transporter from the alginolytic eukaryote *Asteromyces cruciatus*[4]. The genomic integration and overexpression of the gene encoding this transporter, together with the necessary bacterial alginate and deregulated native mannitol catabolism genes, conferred the ability of an *S. cerevisiae* strain to efficiently metabolize DEHU and mannitol. When this platform was further adapted to grow on mannitol and DEHU under anaerobic conditions, it was capable of ethanol fermentation from mannitol and DEHU, achieving titres of 4.6% (v/v) (36.2 g l$^{-1}$) and yields up to 83% of the maximum theoretical yield from consumed sugars. These results show that all major sugars in brown macroalgae can be used as feedstocks for biofuels and value-added renewable chemicals in a manner that is comparable to traditional arable-land-based feedstocks.**

The United Nations predicts that the world population will grow to 9.6 billion people by 2050 (ref. 5). According to the World Energy Outlook 2012 (ref. 6), the global demand for renewable energy production is anticipated to increase markedly; ethanol production is projected to increase 3.4 times by 2035. In 2010, approximately 40% of the US corn[7] and 55% of the Brazilian sugarcane[8] collected were already used to produce a majority (86%) of the world's total ethanol[9]. Meanwhile, the Food and Agriculture Organization projects that overall food production must increase by 70% between 2005 and 2050 (ref. 10). Because the arable land space is projected to increase by less than 5% by 2050, over 90% of the increase in crop production for both food and energy must be accomplished through yield improvements and increased farming intensity, causing significant stress on water resources[11] and fertiliser use[12]. Thus, more efficient and sustainable sources of biomass will be critical.

Brown macroalgae exhibit several features of an ideal feedstock that can complement the increased global demand on energy and food production. The cultivation of this biomass does not require arable land, fresh water or fertiliser, circumventing adverse impacts on food supplies and resource availability. Large-scale cultivation of brown macroalgae is already being practiced in several countries, yielding over 70 million metric tons per year in 2006 (ref. 3). Because brown macroalgae do not contain lignin, simple biorefinery processes such as milling, leaching and extraction can separate the sugars for conversion into biofuels and renewable chemicals[13]. Additionally, valuable materials, such as protein meal for animal feed and potash fertiliser for crop production, can be separated to support sustainable food production[13] (see Supplementary Discussion for a brown macroalgae biorefinery description).

The most abundant sugars in brown macroalgae are alginate, mannitol and glucan (present as laminarin and cellulose). Conventional industrial microbes can use mannitol and hydrolysed glucan[14]. However, the full potential of biofuel and renewable chemical production from brown macroalgae cannot be realized unless alginate is co-fermented[1]. Alginate composes 30–60% of the total sugars in brown macroalgae[13], so the inability of industrial microbes to catabolise alginate results in a substantial loss of product yield. Additionally, the excess reducing equivalents produced by ethanol fermentation from mannitol can be redox-balanced by alginate catabolism, which would otherwise require electron shunts such as oxygen[14] (Supplementary Fig. 1). Thus, enabling the co-fermentation of alginate and mannitol in an existing industrial microbe is a key criterion for the economic and efficient use of the sugars derived from brown macroalgae[1,2].

Alginate is a linear block copolymer of two uronates, β-D-mannuronate and α-L-guluronate, linked via a 1,4-glycosidic bond and arranged in varying sequences[13]. We previously characterized alginate metabolism in the marine bacterium *Vibrio splendidus* 12B01 (ref. 1). Reconstruction of this pathway in *Escherichia coli* enabled the co-fermentation of alginate, mannitol and glucan into ethanol at a yield of 0.281 g ethanol per g dry brown macroalgae[1,2]. Although this *E. coli* system provides a compelling proof-of-principle example that may be suitable for the production of higher value renewable chemicals such as 1,3-propanediol[15] and 1,4-butanediol[16], *Saccharomyces cerevisiae* is a more amenable host for commercial-scale fuel ethanol production and is the standard microbe in the bioethanol industry[17]. Enabling ethanol production from the brown macroalgae sugars through a *S. cerevisiae* platform requires the engineering of both the mannitol and alginate catabolic pathways (Supplementary Fig. 1). Although *S. cerevisiae* has a native mannitol metabolic pathway[18,19], control of its expression is cryptic and requires deregulation. Genes encoding the alginate metabolism pathway in bacteria[1] also need to be expressed in *S. cerevisiae*. A membrane transport system must be identified and isolated from an alginolytic eukaryote such as *A. cruciatus*[4]. Finally, the efficiencies of the mannitol and alginate metabolism pathways must be synchronised for redox control.

Our preliminary growth analyses and characterization of alginate/oligoalginate degradation profiles (Supplementary Figs 2 and 3) suggest that *A. cruciatus* is capable of DEHU uptake. Transporter activity can be efficiently evaluated by a strain that grows on the substrate of interest, so we constructed two *S. cerevisiae* screening strains with the

[1]Bio Architecture Lab Inc., 604 Bancroft Way, Suite A, Berkeley, California 94710, USA. [2]Department of Chemical Engineering and Materials Science, Michigan State University, 2527 Engineering Building, East Lansing, Michigan 48824-1226, USA. [3]Statoil ASA, Statoil Research Centre, Arkitekt Ebbells vei 10, Rotvoll, 7005 Trondheim, Norway. [4]BALChile S.A., Badajoz 100, Oficina 1404, Las Condes, Santiago 7550000, Chile. [5]BAL Biofuels S.A., Badajoz 100, Oficina 1404, Las Condes, Santiago 7550000, Chile. †Present addresses: Manus Biosynthesis Inc., 790 Memorial Drive, Suite 102, Cambridge, Massachusetts 02139 (C.N.S.S.); Calysta Energy, 1140 O'Brien Drive, Menlo Park, California 94025 (D.D.R.); Sutro Biopharma Inc., 310 Utah Avenue, Suite 150, South San Francisco, California 94080, USA (A.G.); Total New Energies USA, 5858 Horton Street, Emeryville, California 94560 (S.A.T.; V.R.).
*These authors contributed equally to this work.

ability to digest either DEHU (BAL2193) or oligoalginate (BAL2438; we did not preclude the possibility of identifying an oligoalginate transporter) by introducing the genes responsible for bacterial DEHU or oligoalginate catabolism, respectively (Supplementary Fig. 1 and Supplementary Table 1). However, the absence of a suitable positive control makes the screening process more complex. Negative results do not necessarily mean that the tested transporters are inactive; instead, the engineered pathways for DEHU and oligoalginate metabolism in the screening strains might not possess sufficient activity to support growth. To minimise this issue, we set as a criterion during strain construction that each enzyme must possess more than 0.1 μmol $min^{-1}$ $mg^{-1}$ specific activity in crude lysate. This activity is comparable to one of the slower steps of the glycolytic pathway in *S. cerevisiae*[20]. Each enzyme present in these screening strains was selected via *in vitro* testing (Supplementary Tables 3–7).

To identify an alginate transporter, an RNA-sequencing (RNA-seq)-based differential expression analysis of *A. cruciatus* grown on alginate versus glucose was performed using two assembly and expression analysis programs, Trinity[21] and Cufflinks[22]. The top 20 genes identified by each program are listed in Supplementary Tables 8 and 9, respectively. These results are similar and show that several transporters are highly induced when *A. cruciatus* is grown on alginate. Among them, the transcript comp8660 (XLOC_004822), encoding an MFS quinate transporter homologue, is the most highly expressed transcript. The list also contains transcripts that might be involved in uronate-containing polysaccharide metabolism, such as homologues of polysaccharide lyases (PL), alcohol dehydrogenase (ADH), dihydroxyacetone kinase (DAK) and dihydrodipicolinate synthase (aldolase)[23]. Genomic DNA sequencing revealed that some of these genes were clustered in close physical proximity to each other (Supplementary Fig. 4), including comp8660, implying that comp8660 plays an important role in alginate metabolism. A complementary DNA (cDNA) on the transcript comp8660 was cloned into a yeast expression plasmid (pDHT1-1) that was transformed into BAL2193 and BAL2438, and the transformed strains were tested for growth on DEHU and oligoalginate substrates, respectively. BAL2193 harbouring the plasmid showed improved growth in screening (S) media (Supplementary Table 10) containing DEHU compared to BAL2193 harbouring an empty vector (Fig. 1), whereas BAL2438 harbouring the same plasmid did not grow on oligoalginate. Thus, the cDNA derived from comp8660 encodes a DEHU transporter, named Ac_DHT1 (Supplementary Sequence).

In a complementary approach, we created a constitutively expressed cDNA library isolated from *A. cruciatus* that was actively growing on alginate. The cDNA library was then transformed into both BAL2193 and BAL2438 and tested for its ability to impart growth on DEHU and oligoalginate, respectively. The population was screened in triplicate by serially subculturing the strains in shake flasks over a span of 28 days to allow for the enrichment of actively growing strains. BAL2193 harbouring the cDNA library showed significant outgrowth after 13 days following the third round of subculture (Supplementary Fig. 5a). Additional rounds of subculture led to further increases in the final $D_{600\,nm}$ values at much shorter time intervals (48–72 h). By contrast, we observed no outgrowth in BAL2438 harbouring the cDNA library, even after 77 days and 5 rounds of subculture (Supplementary Fig. 5b). To determine the identities of the cDNA inserts, we isolated single colonies by plating BAL2193 cultures at the end of the third round of subculture. In total, 46 out of 63 tested colonies (21 from each flask) had three- to sixfold higher final $D_{600\,nm}$ values than BAL2193 harbouring an empty vector (Supplementary Fig. 6). Sequencing analysis revealed that all 46 strains contained plasmids (pDHT1-2) with the *Ac_DHT1* sequence. The colony with the best growth characteristics was re-grown in media containing DEHU (Fig. 1). When any of the genes in the DEHU metabolism pathway was removed from the host strain, the strain could not grow in media containing DEHU as the sole carbon source (Supplementary Fig. 7), indicating that growth in DEHU is pathway dependent.

Ac_DHT1 shows a high degree of homology to quinate transporters, which is unexpected because no clear correlation between quinate and

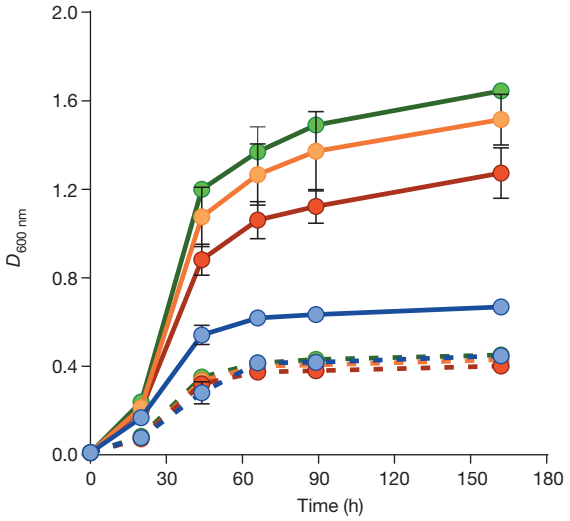



**Figure 1 | Functional validation of a DEHU transporter gene (*Ac_DHT1*) isolated via complementary strategies.** *Ac_DHT1* was isolated through two complementary strategies: direct cloning following an RNA-seq-based differential expression analysis and functional screening of a cDNA library derived from *A. cruciatus*. Growth validation for the isolated *Ac_DHT1* gene is shown. Blue, BAL2193 with an empty vector plasmid; orange, BAL2193 with pDHT1-1 (cDNA identified through the RNA-seq-based analysis, gene expression controlled by *pTDH3*); green, BAL2193 with pDHT1-2 (isolated through the cDNA library screen); and red, BAL2193 re-transformed with isolated pDHT1-2. The solid lines represent growth on media containing DEHU, and the dotted lines represent growth on media with no substrate. $D_{600\,nm}$ measurements were taken in a microplate reader. All data and error bars represent averages and standard deviations of 6–8 measurements.

DEHU can be inferred from their structures. To better understand their evolutionary correlation, we built a phylogenetic tree comprising the following set (303 protein sequences): Ac_DHT1, *Aspergillus nidulans* quinate transporter (QutD)[24], and their homologues along with the second most highly expressed MFS transporter in *A. cruciatus* (comp8944) (Fig. 2a). All 100 sequences obtained from a BLASTP search using *Aspergillus nidulans* QutD were classified into a single family (QutD family). DHT1 homologues are more highly diverged and were classified into three major groups (Fig. 2a). To address the question of why quinate and DEHU transporters are evolutionarily related, we analysed the structure of DEHU in solution using nuclear magnetic resonance (NMR) to determine whether it possesses the structure indicated previously using a chemical method[25] and routinely cited in other studies. The $^{1}$H-NMR, gradient correlation spectroscopy (gCOSY) and $^{13}$C-NMR analyses revealed that DEHU molecules are predominantly hydrated to form two cyclic hemiacetal stereoisomers, a behaviour that is similar to that of glutaraldehyde[26] and another unsaturated uronic acid[27] (Supplementary Figs 8–10 and Supplementary Tables 11 and 12). These DEHU derivatives are structurally similar to quinate (Fig. 2b), which could explain why QutD and DHT1 are evolutionarily related.

The functional regulation of mannitol catabolism in *S. cerevisiae* is cryptic[18]. However, it is understood that the expression of YEL070W/YNR073C (an $NAD^{+}$-dependent mannitol-2-dehydrogenase (M2DH)) along with a mannitol transporter is necessary to support *S. cerevisiae* growth on mannitol[19]. To address mannitol catabolism in the strains used in the present study, the gene expression levels of three strain backgrounds (Lalvin, Pasteur Red and SEY/Dip) that grew on mannitol (Supplementary Fig. 11) were analysed by microarray in three different growth substrates: glucose, raffinose, and mannitol. The analysis revealed that the top three genes induced in all three strains encode M2DH (YEL070W/YNR073C), previously uncharacterized putative MFS transporters (YEL069C(*HXT13*)/YNR072W(*HXT17*)), and an aldose-1-epimerase homologue (YNR071C). These genes were

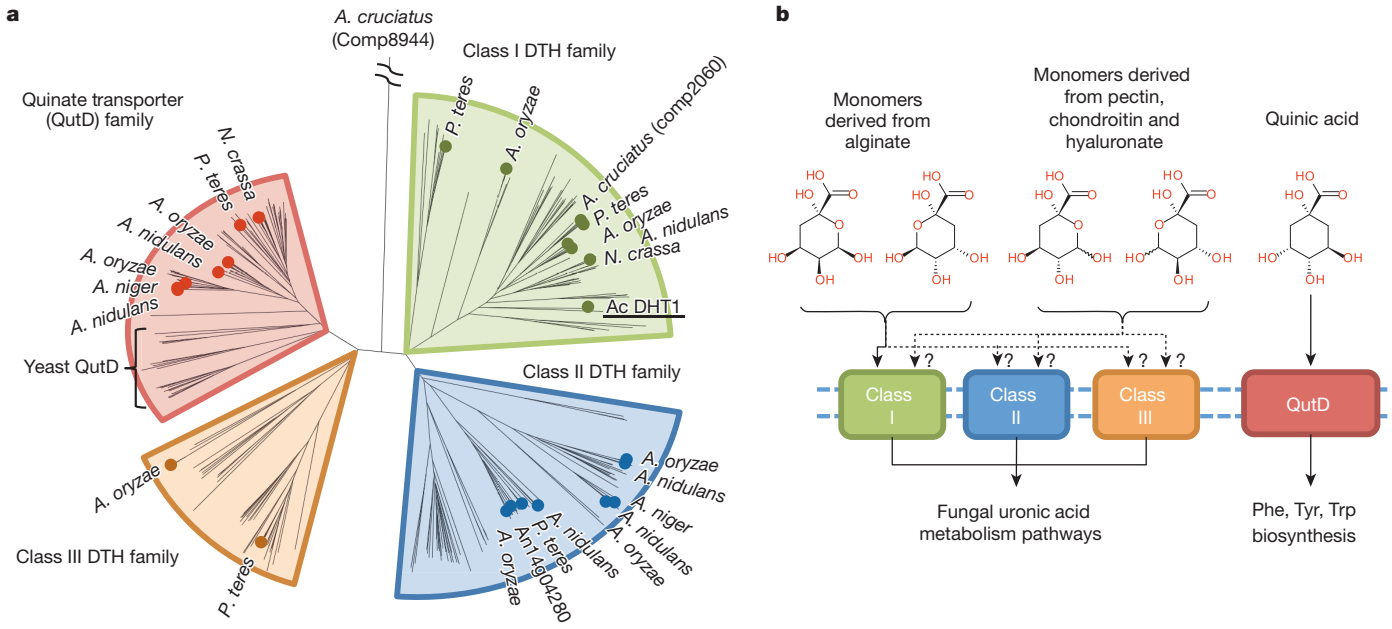



**Figure 2 | Evolutionary correlation among quinate (QutD), Ac_DHT1, and homologous transporters.** **a**, A phylogenetic tree of QutD, Ac_DHT1, and homologous transporters is generated. The tree indicates a strong evolutionary relationship between QutD and DHT1. Whereas QutD and its homologues are tightly clustered, DHT1 is diverged and clustered into three major groups (Class I, II, and III DHT1 homologue (DTH) families). *A. cruciatus* DHT1 is found in class I. *A. niger*, *Aspergillus niger*; *A. oryzae*, *Aspergillus oryzae*; *N. crassa*, *Neurospora crassa*; *P. teres*, *Pyrenophora teres*. **b**, A putative model showing how DTH families could be involved in uronic acid metabolism is shown. The $^{1}$H-NMR, its gCOZY, and $^{13}$C-NMR studies indicate that DEHU can be readily hydrated to form a cyclic hemiacetal structure in an aqueous solution, similar to the structure of quinate. Although further biochemical characterization of these transporters is required to validate the model, DTH families may be strongly diverged to allow the uptake of different stereoisomers of quinate-like hydrated, cyclic, hemiacetal forms of unsaturated uronates.

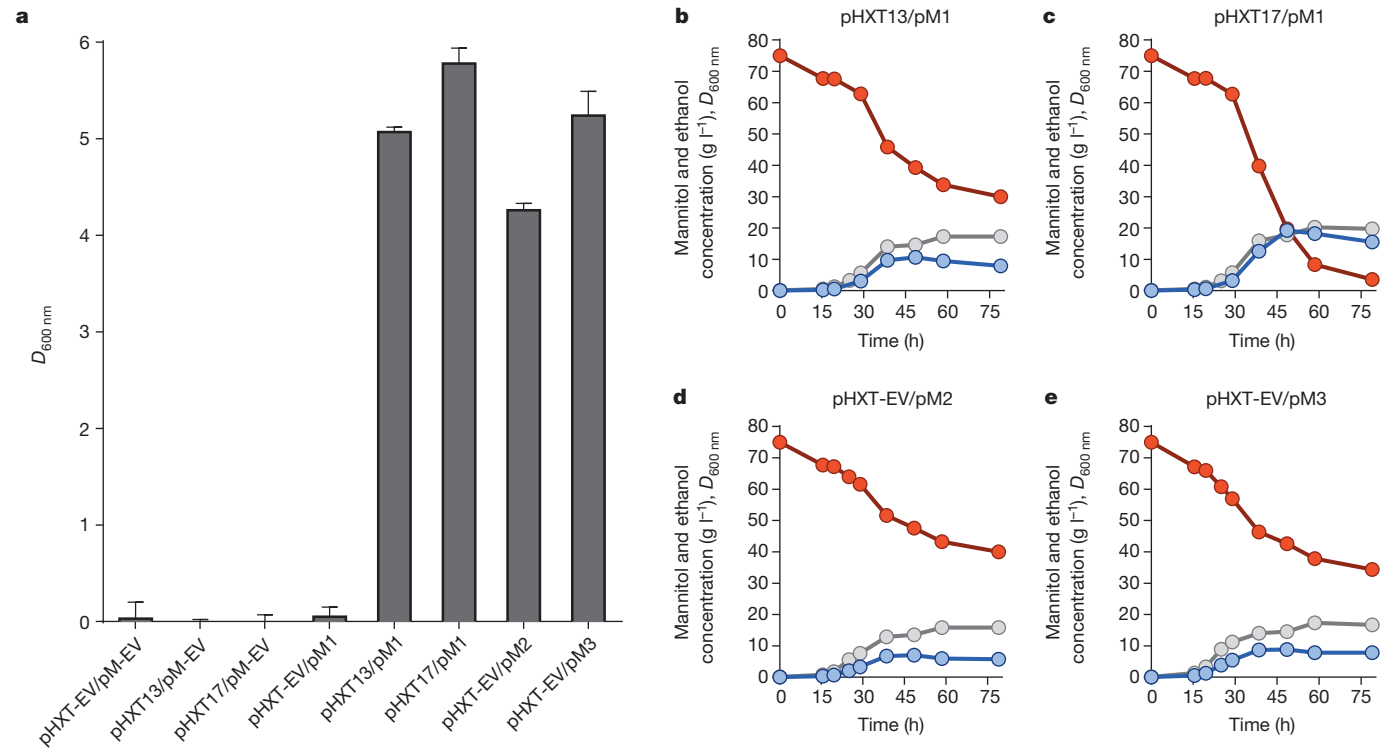



**Figure 3 | Engineering mannitol metabolism in *S. cerevisiae*.** **a**, Preliminary growth analysis of the *S. cerevisiae* strain BAL2970 harbouring the plasmids pM-EV (empty vector plasmid for pM1, pM2, and pM3), pM1 (YNR073C, YNR072W, and YNR071C present; only YNR073C overexpressed), pM2 (YNR073C and YNR072W present and overexpressed), pM3 (YNR073C, YNR072W, and YNR071C overexpressed), pHXT-EV (an empty vector for pHXT13 and pHXT17), pHXT13 (YEL069W) or pHXT17 (YNR072W) were tested. **b**–**e**, Ethanol production, mannitol consumption, and $D_{600\,nm}$ for BAL2970 harbouring pM1 and pHXT13 (**b**), pM1 and pHXT17 (**c**), pM2 and pHXT-EV (**d**), and pM3 and pHXT-EV (**e**), are shown. Overexpression of YNR072W resulted in better growth compared to overexpression of YEL069C (**b** and **c**), and overexpression of YNR071C slightly improved growth compared a construct lacking this gene (**d** and **e**). Red, mannitol concentration; blue, ethanol concentration; grey, $D_{600\,nm}$. All data and error bars represent averages and standard deviations of triplicate measurements.

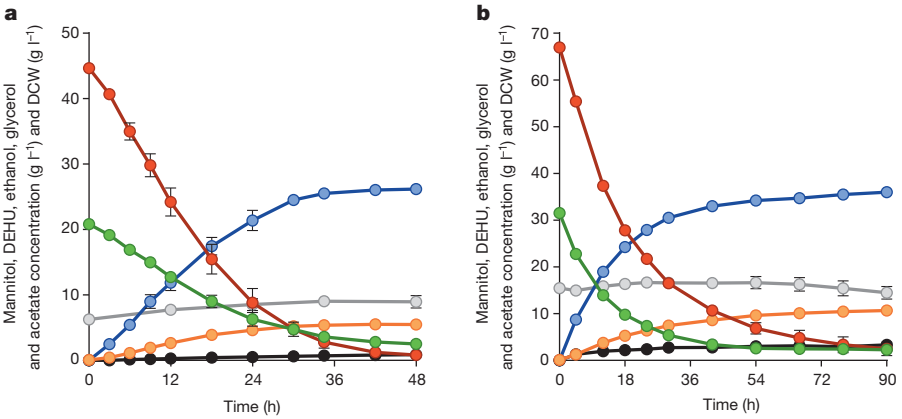



**Figure 4 | Ethanol fermentation from mannitol and DEHU.** Ethanol was fermented from a 1:2 molar ratio of DEHU:mannitol at 6.5% (w/v) (**a**) and 9.8% (w/v) (**b**) of total sugars using BAL3215. **a**, An ethanol titre of 3.3% (v/v) (26.2 g $l^{-1}$), maximum theoretical yield of 83% from consumed sugars (78% from fed sugars), and maximum ethanol productivity of 1.0 g $l^{-1}$ $h^{-1}$ were achieved. The ratio of mannitol:DEHU consumption was 2.4. **b**, An ethanol titre of 4.6% (v/v) (36.2 g $l^{-1}$), maximum theoretical yield of 75% from consumed sugars (71% from fed sugars), and maximum ethanol productivity of 1.9 g $l^{-1}$ $h^{-1}$ were achieved. The ratio of mannitol:DEHU consumption was 2.1. Red, mannitol concentration; green, DEHU concentration; blue, ethanol concentration; orange, glycerol concentration; black, acetate concentration; grey, dry cell weight (DCW). All data and error bars represent the averages and standard deviations of triplicate measurements.

combinatorially overexpressed in BAL2970 (Fig. 3a and Supplementary Table 2), which was then assessed for ethanol production, mannitol consumption and growth (Fig. 3b–e). As expected, the minimal genes that were required for *S. cerevisiae* growth on mannitol encoded M2DH and a mannitol transporter (see Supplementary Discussion).

To enable the co-metabolism of mannitol and alginate, all genes encoding the enzymes necessary for alginate and mannitol metabolism were chromosomally integrated, and the effects of the cofactor preferences of different DehR enzymes were evaluated using ethanol production as a guideline. Because *S. cerevisiae* M2DH generates excess NADH, it is anticipated that only an NADH-dependent DehR can counterbalance the excess reducing equivalents produced from mannitol consumption. Therefore, we chromosomally integrated genes encoding DehR, which preferentially use NADH and co-use NADH and NADPH (Supplementary Fig. 12), into the *S. cerevisiae* genome to construct BAL2759 and BAL2956, respectively (BAL2772, harbouring an NADPH-dependent DehR, was used as a control). Initial DEHU-dependent growth rates were poor in these strains, so we performed an adaptation experiment to improve the efficiency of DEHU catabolism. The strains were grown aerobically in shake flasks in media with DEHU as the only carbon source. The cultures were diluted every few days to maintain exponential growth over the course of three months. The strain doubling time decreased from more than 60 h to less than 5 h over the course of the experiment (Supplementary Fig. 13). A preliminary experiment suggested that BAL2956 showed an increased ability to produce ethanol from mannitol and DEHU (Supplementary Fig. 14). These results indicate that the ability of DehR to co-use NADH and NADPH efficiently is crucial (for example, the reservation of some NADH is necessary for basal cell growth and metabolic maintenance).

The adapted BAL2956 strain had limited capability for anaerobic ethanol fermentation from DEHU and mannitol, requiring microaerobicity for efficient ethanol production. Therefore, we adapted and selected the strain for growth on DEHU and mannitol under anaerobic conditions. Because we were unable to observe anaerobic growth of BAL2956 in liquid culture, these experiments were carried out on an agar plate with HC media containing DEHU and mannitol. Approximately 200 colonies appeared on the plate within 5 days after an actively growing aerobic culture of BAL2956 was plated. Several colonies were re-streaked on fresh plates with the same media conditions. Single colonies were isolated and grown aerobically in HC media containing DEHU, and the culture (BAL3215) with the fastest growth rate ($\mu > 0.12\,h^{-1}$) was chosen for further experiments. Ethanol fermentation experiments were performed in small-scale laboratory vessels with HC media containing an approximately 1:2 molar ratio of DEHU:mannitol at 6.5% (w/v) and 9.8% (w/v) total sugars (mimicking the sugars prepared via the brown macroalgae biorefinery approach, Supplementary Fig. 15). In both cases, ethanol was efficiently produced, achieving titres of 4.6% (v/v) (36.2 g $l^{-1}$), comparable to the benchmark titres for commercial cellulosic ethanol industries[28], and yields up to 83% of the maximum theoretical yield from consumed sugars (Fig. 4).

In conclusion, we have successfully developed the first *S. cerevisiae* synthetic biology platform that enables the use of unique sugars in brown macroalgae for high-efficiency ethanol fermentation. The results of the present study are of great significance because *S. cerevisiae* is the standard microbe for industrial-scale fuel alcohol production[17]. The use of this platform, however, is not limited to ethanol production; it can be used to produce many other biofuels and renewable chemicals with further genetic modifications[29,30]. Together with other synthetic biology platforms, we foresee that this *S. cerevisiae* platform technology could help to create a sustainable economy and society.

## METHODS SUMMARY

Enabling *S. cerevisiae* to ferment brown macroalgae sugars into ethanol required four major engineering modifications: (1) reconstruction of a bacterial alginate pathway, (2) identification and integration of a DEHU transporter, (3) deregulation of a native mannitol catabolic pathway, and (4) selection of a DehR with optimal cofactor preference for redox-balance. (1) Several genes encoding enzymes catalysing each step involved in DEHU catabolism were overexpressed in *S. cerevisiae*. The specific activity of each enzyme in a crude lysate was assayed. The genes encoding enzymes with high specific activities (>0.1 μmol $min^{-1}$ $mg^{-1}$) were chromosomally integrated into the DEHU transporter screening strain (BAL2193) and ethanol production strains (BAL2772, BAL2759, BAL2956 and BAL3215). (2) To isolate a gene encoding the DEHU transporter, two complementally methods were used: mRNA-seq-based analysis and cDNA library based on mRNA isolated from *A. cruciatus* actively growing on alginate. The mRNA-seq-based analysis was carried out using Trinity[21] and Cufflinks[22]. The transporter activity was assessed by determining whether overexpression of cDNAs derived from both methodologies enabled BAL2193 to grow on DEHU as the sole sugar substrate. (3) Differential gene expression analyses of three *S. cerevisiae* strains grown in glucose, raffinose and mannitol were performed using microarrays. The *S. cerevisiae* native M2DHs and putative mannitol transporters were overexpressed, and the gene combination supporting optimum growth of *S. cerevisiae* on mannitol was identified. (4) The genes involved in alginate and mannitol catabolism pathways were chromosomally integrated into a single *S. cerevisiae* host strain. The cofactor preference of DehR was investigated based on the ability to support ethanol production from brown macroalgae sugars (BAL2772, BAL2759, and BAL2956). Sequential passage of BAL2956 in DEHU media under aerobic conditions and in mannitol and DEHU media under anaerobic conditions allowed the strain (BAL3215) to adapt to enable efficient ethanol fermentation from brown macroalgae sugars.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 15 May; accepted 8 October 2013.**

**Published online 1 December 2013; corrected online 8 January 2014 (see full-text HTML version for details).**

1. Wargacki, A. J. *et al.* An engineered microbial platform for direct biofuel production from brown macroalgae. *Science* **335,** 308–313 (2012).
2. Santos, C. N., Regitsky, D. D. & Yoshikuni, Y. Implementation of stable and complex biological systems through recombinase-assisted genome engineering. *Nature Comm.* **4,** 2503 (2013).
3. Roesijadi, G., Jones, S. B., Snowden-Swan, L. J. & Zhu, Y. *PNNL-19944: Macroalgae as a Biomass Feedstock: A Preliminary Analysis* (Pacific Northwest National Laboratory, 2010).
4. Schaumann, K. & Weide, G. Efficiency of uronic acid uptake in marine alginate-degrading fungi. *Helgol. Meersunters.* **49,** 159–167 (1995).
5. World Population Prospects. *The 2012 Revision, Volume 1 Comprehensive Table* (United Nations, 2013).
6. International Energy Agency. World Energy Outlook 2012. (2012).
7. Interagency Agricultural Projections Committee. USDA Agricultural Projections to 2021. (2012).
8. Valdes, C. Can Brazil meet the world's growing need for ethanol? *AMBER WAVES* **9,** 38–45 (2011).
9. Renewable Fuels Association. http://www.ethanolrfa.org/pages/statistics#F (2013).
10. Food and Agriculture Organization of United Nations. How to Feed the World in 2050. (2009).
11. United Nations Educational, Scientific and Cultural Organization. The United Nations World Water Development Report 4: Managing Water under Uncertainty and Risk. (2012).
12. Stewart, W. M., Dibbb, D. W., Johnston, A. E. & Smyth, T. J. The contribution of commercial fertilizer nutrients to food production. *Agron. J.* **97,** 1–6 (2005).
13. Chapman, V. J. *Seaweeds and Their Uses*, 2nd edn (The Camelot Press, 1970).
14. Horn, S. J., Aasen, M. & Østgaard, K. Ethanol production from seaweed extract. *J. Ind. Microbiol. Biotechnol.* **25,** 249–254 (2000).
15. Nakamura, C. E. & Whited, G. M. Metabolic engineering for the microbial production of 1,3-propanediol. *Curr. Opin. Biotechnol.* **14,** 454–459 (2003).
16. Yim, H. *et al.* Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nature Chem. Biol.* **7,** 445–452 (2011).
17. Matrin Patel, M. C. *et al.* Medium and long-term opportunities and risks of the biotechnological production of bulk chemicals from renewable resources - the potential of white biotechnology. (The Brew Project, 2006).
18. Quain, D. E. & Boulton, C. A. Growth and metabolism of mannitol by strains of *Saccharomyces cerevisiae. J. Gen. Microbiol.* **133,** 1675–1684 (1987).
19. Noiraud, N., Maurousset, L. & Lemoine, R. Identification of a mannitol transporter, AgMaT1, in celery phloem. *Plant Cell* **13,** 695–705 (2001).
20. Mairta, P. K. & Lobo, Z. A kinetic study of glycolytic enzyme synthesis in yeast. *J. Biol. Chem.* **246,** 425–468 (1971).
21. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnol.* **29,** 644–652 (2011).
22. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7,** 562–578 (2012).
23. Hilditch, S. *Identification of the Fungal Catabolic D-galacturonate Pathway*. PhD thesis, Univ. Helsinki (2010).
24. Wheeler, K. A., Lamb, H. K. & Hawkins, A. R. Control of metabolic flux through the quinate pathway in *Aspergillus nidulans. Biochem. J.* **315,** 195–205 (1996).
25. Preiss, J. & Ashwell, G. Alginic acid metabolism in bacteria. II. The enzymatic reduction of 4-deoxy-L-erythro-5-hexoseulose uronic acid to 2-keto-3-deoxy-D-gluconic acid. *J. Biol. Chem.* **237,** 317–321 (1962).
26. Migneault, I., Dartiguenave, C., Bertrand, M. J. & Waldron, K. C. Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking. *Biotechniques* **37,** 790–796, 798–802 (2004).
27. Jongkees, S. A. & Withers, S. G. Glycoside cleavage by a new mechanism in unsaturated glucuronyl hydrolases. *J. Am. Chem. Soc.* **133,** 19334–19337 (2011).
28. Lau, M. W. & Dale, B. E. Cellulosic ethanol production from AFEX-treated corn stover using *Saccharomyces cerevisiae* 424A(LNH-ST). *Proc. Natl Acad. Sci. USA* **106,** 1368–1373 (2009).
29. Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B. & Keasling, J. D. Microbial engineering for the production of advanced biofuels. *Nature* **488,** 320–328 (2012).
30. Lee, J. W. *et al.* Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nature Chem. Biol.* **8,** 536–546 (2012).

**Acknowledgements** We thank R. Schekman, J. H. D. Cate and P. A. Silver for critical discussion and suggestions. This work is supported by the DOE under an Advanced Research Projects Agency–Energy (ARPA-E) award (DE-AR0000006), by CORFO INNOVA CHILE (código 09CTEI-6866), and by Statoil ASA. This report was prepared as an account of work sponsored by an agency of the US government. Neither the US government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favouring by the US government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the US government or any agency thereof.

**Author Contributions** Y.Y. conceived the overall project. C.S., Y.Y., M.E.-N., N.L.O., R.B.B., H.K.K. and V.R. supervised the overall project. R.M.R., S.R.C., P.D.D., Y.M. and A.G. characterised the *A. cruciatus* alginate degradation profiles. A.M.E.F., M.E.-N., P.K., L.L., M.Y.C. and S.A.T. constructed all *S. cerevisiae* host strains for screening. R.M.R. and P.D.D. prepared *A. cruciatus* genomic DNA and total RNA for the high-throughput sequencing analysis. D.D.R., R.M.R. and Y.Y. assembled and analysed the RNA-seq data. C.N.S.S., P.T., M.E.-N., J.G. and C.J.P. constructed and screened the *A. cruciatus* cDNA library. M.E.-N. and J.L.V. identified yeast strains that can grow on mannitol and performed microarray analyses to isolate mannitol catabolism genes. A.M.E.F., M.E.-N., L.L., M.Y.C., P.D.D. and J.G. constructed all *S. cerevisiae* host strains for fermentation experiments. Y.Y., D.D.B., M.E.-N. and A.C. carried out all adaptation and ethanol fermentation experiments. A.H. and P.S. developed and carried out biochemical assays for enzyme characterisation. S.R.C., P.S., Y.Y., D.D.B., B.L. and A.H. performed all necessary analytical experiments and analysed all fermentation samples. L.P., C.L., E.J.K., N.L.O. and D.J.M. conducted sample preparation and the NMR analysis of the DEHU structure. C.L., P.S., D.D.B. and Y.Y. prepared media for fermentation, Y.Y., M.E.-N., A.M.E.F., C.S., D.D.B., C.N.S.S. and L.P. wrote and revised the manuscript.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence should be addressed to Y.Y. (yoshikuni@ba-lab.com) or C.S. (candacesw@yahoo.com). Requests for materials should be addressed to Y.Y. (yoshikuni@ba-lab.com).

## METHODS

All reagents were purchased from Sigma-Aldrich unless stated otherwise. *Asteromyces cruciatus* ATCC 26324 was purchased from ATCC. *S. cerevisiae* strains were derived from the laboratory strain SEY6210/6211 (ATCC). Standard molecular biology techniques were used to engineer all *S. cerevisiae* strains. The genotypes of all *S. cerevisiae* screening and fermentation strains and the compositions of all plasmids that were used in this manuscript are listed in Supplementary Tables 1 and 2, respectively.

**Characterization of alginate degradation by *A. cruciatus* (in culture supernatant).** A strain of *A. cruciatus* was inoculated into 50 mM phosphate buffer at neutral pH supplemented with 10 mM nitrate, 2% sodium alginate, 5 μg ml$^{-1}$ ampicillin, and 20 μg ml$^{-1}$ kanamycin at a final volume of 200 ml in a shake flask. The culture was grown for 3 days in an orbital shaker at 200 r.p.m. at 25 °C. The culture supernatant was collected, filter-sterilized, and mixed with fresh 4% alginate solution. Ten days after the reaction was initiated, samples were filtered for high-performance liquid chromatography-mass spectrometry (HPLC-MS) analysis. The analysis was carried out using an HPLC Nexera system (Shimadzu) equipped with a Hypercarb (100 × 3 mm) column (Thermo Fisher Scientific) with a flow rate of 0.65 ml min$^{-1}$ and a temperature of 65 °C. A gradient of 0.2% trifluoroacetic acid in water against methanol from 30% to 90% over 15 min was programmed. Oligoalginate peaks were detected using an MS (Shimadzu) with selected ion monitoring detector.

**Characterization of oligoalginate degradation by *A. cruciatus* (growth-coupled degradation).** A strain of *A. cruciatus* was inoculated into YNB media supplemented with amino acids, 2% sodium alginate degraded by alginate lyase, 25 μg ml$^{-1}$ ampicillin and 25 μg ml$^{-1}$ kanamycin at a final volume of 25 ml in a shake flask. The culture was grown for 7 days in an orbital shaker at 200 r.p.m. at 25 °C. The culture supernatant was filtered for HPLC-ultraviolet analysis. The analysis was carried out using an HPLC Nexera system (Shimadzu) equipped with a TSKgel-amide 80 (2 mm × 15 cm) 5-μm Tosoh column (TOSHO) with a flow rate of 0.5 ml min$^{-1}$ and a temperature at 50 °C. A gradient of 25 mM ammonium formate in water against acetonitrile from 95% to 70% over 5 min, from 70% to 50% over 7 min, and 50% held constant for 2 min was programmed. Alginate oligomers were detected using an ultraviolet detector (Shimadzu) at 235 nm.

**Construction of *S. cerevisiae* tester strains to test DEHU metabolism enzyme activities.** Genes encoding DEHU reductase (DehR), 2-keto-3-deoxy-D-gluconate (KDG) kinase (KdgK), KDG-6-phosphate (KDGP) aldolase (KdgpA) and oligoalginate lyase from several organisms were codon optimised and/or synthesized (DNA2.0 Inc.). The synthesized genes were cloned into plasmids downstream of strong, constitutive *S. cerevisiae* promoters (*pTEF2* or *pFBA1*). DNA constructs containing these sequences for genomic integration were generated by high-fidelity PCR amplification using Phusion DNA polymerase (New England Biolabs) and integrated into the genome of the wild-type laboratory strain BAL2130 using a standard polyethylene glycol, lithium acetate, Tris-EDTA (PEG/LiAC/TE) yeast transformation protocol. Genomic integrations were confirmed by diagnostic PCR. The activity of each introduced enzyme in a BAL2130 crude lysate was measured with the appropriate enzyme activity assay described below. The enzyme activity results are listed in Supplementary Tables 3–7.

**Assay for oligoalginate lyase activity.** A spectrophotometric coupled enzyme assay for oligoalginate lyase was developed. The reaction mixture contained 50 mM HEPES (pH 7.4), 100 mM NaCl, 50 mM KCl, 5 mM $MgCl_2$, 0.005% (w/v) Triton X-100, 0.5% (w/v) oligoalginate, 4 mM NADPH, and purified DehR from *Agrobacterium tumefaciens* C58 in a final volume of 50 μl. Oligoalginate was prepared by incubating 0.10 mg ml$^{-1}$ alginate lyase with 50 g l$^{-1}$ sodium alginate in the presence of 2 mM dithiothreitol (DTT) for over 10 h at 37 °C, and this preparation was used directly in the assay. Excess partially purified DehR (sufficient to give a rate of ≥ 0.2 $A_{340\,nm}$ min$^{-1}$ with 10 mM DEHU and 2 mM NADPH) was present along with 2 mM NADPH. The reaction was initiated by adding 5 μl of *S. cerevisiae* crude lysate.

**Assay for DehR activity.** A spectrophotometric assay to identify DehR activity in *S. cerevisiae* lysates was developed. The assay directly measures the oxidation of NADPH when DEHU is reduced to KDG. The reaction mixture contained 50 mM HEPES (pH 7.4), 100 mM NaCl, 50 mM KCl, 5 mM $MgCl_2$, 0.005% (w/v) Triton X-100, 10 mM DEHU, and 4 mM NADPH in a final volume of 50 μl. The reaction was initiated by adding 5 μl of *S. cerevisiae* crude lysate.

**Assay for KdgK activity.** A spectrophotometric assay to identify KdgK activity in *S. cerevisiae* lysates was developed. KdgK activity is measured by NADH consumption coupled with KdgpA and lactate dehydrogenase activity. First, phosphorylated KDG is cleaved into pyruvate and glyceraldehyde-3-phosphate by KdgpA. Lactate dehydrogenase reduces pyruvate into lactate while simultaneously oxidising NADH. The reaction mixture contained 50 mM HEPES (pH 7.4), 100 mM NaCl, 50 mM KCl, 5 mM $MgCl_2$, 2 μM DTT, 0.005% (w/v) Triton X-100, 5 mM KDG, 10 mM ATP, 4 mM NADH, 2 μg ml$^{-1}$ partially purified KdgpA from *E. coli,* and 50 μg ml$^{-1}$ lactate dehydrogenase in a final volume of 50 μl. In this assay, KDG was prepared via the reaction of mannonate lactone (Carbsynth) and a purified mannonate dehydratase from *E. coli* (prepared in-house).

**Assay for KdgpA activity.** A spectrophotometric assay to identify KdgpA activity in *S. cerevisiae* lysates was developed. The reaction mixture contained 50 mM HEPES (pH 7.4), 100 mM NaCl, 50 mM KCl, 5 mM $MgCl_2$, 2 μM DTT, 0.005% (w/v) Triton X-100, 5 mM KDGP, 4 mM NADH, and 50 μg ml$^{-1}$ lactate dehydrogenase in a final volume of 50 μl. KDGP was synthesized by the reaction of KDG and ATP with KdgK.

**Construction of DEHU and oligoalginate transporter screening strains.** Genes encoding oligoalginate lyase from *Agrobacterium tumefaciens* C58 or *Oceanicola granulosus* HTCC2516, DehR from *Sphingomonas sp.* A1, KdgK from *Saccharophagus degradans* 2-40, and KdgpA from *Vibrio splendidus* 12B01 were selected based on the results of enzyme activity assays (with the exception of the oligoalginate lyase from *Agrobacterium tumefaciens* C58, which was integrated before enzyme screening; Supplementary Tables 3–6). The sequences were cloned downstream of constitutive *S. cerevisiae* promoters (*pTEF2*, *pFBA1* or *pTDH3*) into plasmids. DNA constructs containing these sequences for genomic integration were generated by high-fidelity PCR amplification and were integrated into the *S. cerevisiae* genome using a standard PEG/LiAC/TE yeast transformation protocol to create the host strains BAL2193 and BAL2438. Genomic integrations were confirmed by diagnostic PCR. The genotypes of these strains are listed in Supplementary Table 1. The activity of each introduced enzyme in crude lysates was measured with the appropriate enzyme activity assay.

**Identification of DEHU transporter (Ac_DHT1) through RNA-seq-based analysis.** *A. cruciatus* was grown aerobically at 30 °C in YP media with either 4% sodium alginate or 2% glucose as the primary carbon source. Cells were collected from exponentially growing cultures, and DNA and total RNA were isolated. The genomic DNA was sequenced by Eureka Genomics using 51-cycle paired-end reads on an Illumina Genome Analyzer II instrument (Illumina). These sequence reads were assembled using the Velvet software package[31] to generate a library of putative genomic DNA fragments with a minimum scaffold size of 200 bases. The RNA-seq analysis was carried out by Prognosys Biosciences using 76-cycle single-end reads on an Illumina Genome Analyzer 5 instrument (Illumina). The differential expression of mRNAs in *A. cruciatus* grown on alginate versus glucose was analysed using the Trinity[21] and Cufflinks[22,32] program packages. A cDNA on the transcript comp8660 was amplified by PCR using total RNA isolated from *A. cruciatus* as a template. The amplified DNA fragment was cloned into an *S. cerevisiae* CEN/ARS-based expression vector under the constitutive *S. cerevisiae TDH3* promoter.

**DEHU-dependent growth screening in deep-well plates.** DEHU was prepared from a 5% sodium alginate dissolved in water that was treated with 0.025 mg ml$^{-1}$ alginate lyase and 0.2 mg ml$^{-1}$ oligoalginate lyase (prepared in house)[1]. This mixture was subjected to high-speed centrifugation followed by filtration through a 0.22-μm sterile filter (When needed, this DEHU solution can be concentrated using a conventional rotary evaporator). To minimise background growth, yeast minimal media consisting of YNB supplemented with CSM lacking uracil (Sunrise Science Products) was used at a concentration recommended by the manufacturer (Screening (S) media in Supplementary Table 10). DEHU was added to each experimental well at a concentration of 8.8 g l$^{-1}$. Growth assays were carried out in 96-deep-well plates at 29 °C with 90% humidity and shaking at 950 r.p.m. Growth was assessed by measuring the $D_{600\,nm}$ of the culture using an Eon Microplate Spectrophotometer (BioTek).

**Identification of Ac_DHT1 using an *A. cruciatus* cDNA library.** Total RNA was isolated from 1 l cultures of *A. cruciatus* that were actively growing on alginate using the RNeasy Plant Mini Kit (Qiagen). The mRNA was purified with the Dynabead mRNA purification kit (Invitrogen) and used as an input for cDNA library construction with the Superscript Full-Length cDNA Library Construction Kit (Invitrogen). cDNA inserts were cloned into a Gateway cloning-compatible yeast expression vector under the control of the *S. cerevisiae TDH3* promoter. The plasmid contained a blasticidin resistance gene for plasmid selection.

The cDNA library was introduced into the host strains BAL2193 and BAL2438. These strains were grown at 30 °C in S media (Supplementary Table 10) containing 13.2 g l$^{-1}$ DEHU or 25 g l$^{-1}$ oligoalginate, respectively. Cultures were periodically set back to an $D_{600\,nm}$ of 0.1 (indicated by the dashed lines in Supplementary Fig. 5 for BAL2193). After the third round of subculture, BAL2193 showed significant outgrowth, and cells from this culture were plated onto YPD plates supplemented with blasticidin to isolate single colonies. A total of 21 colonies from each flask were tested for growth on S media with DEHU and S media with no substrate in a deep-well plate format (600 μl) to isolate the strain harbouring *Ac_DHT1*. The positive clones that were identified via secondary screening were then sequenced to identify the cDNA insert.

**Construction of strains to study the pathway dependence of DEHU consumption.** Strains derived from BAL2193 to test pathway dependence during growth in

DEHU media were generated by sporulating BAL2193 and isolating the haploid strain BAL2267. BAL2267 was mated to the wild-type haploid strain BAL2130, and the resulting heterozygous diploid was sporulated. Following sporulation, haploids that were missing each individual DEHU consumption pathway gene were isolated (BAL2295, BAL2296, BAL2297 and BAL2298). The genotypes of all pathway dependence strains are listed in Supplementary Table 1.

**Construction of a phylogenetic tree for Ac_DHT1.** A BLASTP search against the non-redundant protein sequence database was carried out using the protein sequences Ac_DHT1, an Ac_DHT1 homologue (a cDNA on the transcript comp2060), *Aspergillus niger* An14g04280, and *Aspergillus nidulans* QutD as queries. For each sequence, the protein sequences for the top 100 BLAST hits were retrieved. These protein sequences were pooled (400 sequences). The putative protein sequence for an MFS transporter (comp8944 and XLOC_008636), which is highly expressed in *A. cruciatus* grown in media containing alginate versus glucose, was used as an anonymous root. A multiple sequence alignment (MSA) for these sequences was generated using MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/). All redundant sequences were eliminated from the MSA, and a distance matrix for the remaining 289 aligned protein sequences was generated based on neighbour-joining algorithms. The phylogenetic tree was visualized using Geneious Pro software (Geneious).

This first tree identified that Ac_DHT1, its homologue from *A. cruciatus*, and An14g04280 from *Aspergillus niger* are classified into two different subfamilies (class I and II DHT1 homologue families) and suggested the presence of an additional subfamily of DHT1 homologues that is more strongly related to the QutD family. To refine the tree, a new MSA was rebuilt based on a pool that included 100 protein sequences that were identified through a BLASTP search using gi189198523 (a representative protein in this family) as a query in addition to the 289 protein sequences in the original set. All redundant sequences were eliminated from the MSA, and a new tree was constructed using the remaining 303 aligned protein sequences.

**Sample preparation of DEHU for NMR analysis.** DEHU was prepared from a 5% sodium alginate solution dissolved in water that was treated with 0.025 mg $ml^{-1}$ alginate lyase and 0.2 mg $ml^{-1}$ oligoalginate lyase[1]. The mixture was then subjected to high-speed centrifugation, followed by filtration through a 0.22-μm sterile filter. This DEHU solution was concentrated on a rotary evaporator that maintained a bath temperature below 40 °C, followed by drying under high vacuum conditions (<100 mTorr) for 15–20 h, resulting in a yellow crystalline solid. Phosphate-buffered (approx. 0.2 M) NMR samples of DEHU (approx. 0.1 M) were prepared by dissolving dry DEHU in $D_2O$ with either t-butanol (0.1 M) or methanol (0.1 M) as an internal standard. Phosphate was used to buffer the sample to pH 7.0.

**Standard $^1$H- and $^{13}$C-NMR spectra.** NMR spectra were measured at ambient temperature on a Varian UNITYplus 500 spectrometer equipped with a 5 mm pulsed-field-gradient (PFG) switchable broadband probe and operating at 499.74 MHz ($^1$H) and 125.67 MHz ($^{13}$C). One-dimensional (1D) $^1$H-and $^{13}$C-NMR spectra were acquired under standard conditions and referenced to the solvent residual peak. $^1$H-NMR spectra were obtained using a spectral width of 4,529 Hz over 64,000 data points and a relaxation delay of 2 s. $^{13}$C-NMR spectra were obtained using a relaxation delay of 3 s and spectral width of 26,595.74 Hz over 32,768 data points and multiplied by an exponential function corresponding to a 0.50 Hz line broadening before Fourier transformation.

**2D $^1$H-NMR spectra.** Two-dimensional (2D) $^1$H-NMR spectra were obtained using gradient pulses on a 5 mm PFG switchable broadband probe without sample spinning. Phase-sensitive spectra were acquired using the hyper-complex States method. Threefold linear prediction was applied to the *F*1 dimension as implemented by standard Varian software.

**gCOSY.** Absolute value gCOSY spectra were obtained using a spectral width of 4,529 Hz in both dimensions. The pre-acquisition delay was set to 1.0 s, and 400 increments with 4 transients of 1,024 data points were acquired. Both *F*2 and *F*1 were multiplied by unshifted sine-bell weighting functions that were matched to the acquisition or evolution time. Prior to 2D Fourier transformation, *F*2 and *F*1 were zero filled to 2,048 and 2,048 data points, respectively.

**Construction of an *S. cerevisiae* host strain capable of mannitol metabolism.** An *S. cerevisiae* genomic region containing YNR073C (mannitol-2-dehydrogenase (M2DH)), YNR072W (*HXT17*), and YNR071C (gene encoding a putative aldose-1-epimerase) was amplified by high-fidelity PCR and cloned into a yeast CEN/ARS plasmid downstream of the *S. cerevisiae TDH3* promoter (*pTDH3* driving the expression of YNR073C only; the other two ORFs retained their native promoter regions). This plasmid was named pM1. A second plasmid, containing YNR073C and YNR072W under the control of the *S. cerevisiae TDH3* and *TEF2* promoters, respectively, was also generated (pM2). To increase the expression levels of all three genes in this genomic region, the DNA composing the *S. cerevisiae FBA1* and *TEF2* promoters were subcloned into pM1 to drive the expression of YNR071C and YNR072W, respectively. This plasmid was named pM3. These plasmids, along with plasmids containing only YEL069C (pHXT13) or YNR072W (pHXT17) under the control of the *S. cerevisiae TDH3* promoter, were transformed into the strain BAL2970, and the transformed strains were tested for the ability to grow in HC media containing 7.5% mannitol as the sole carbon source.

**Construction of *S. cerevisiae* host strains with integrated mannitol and DEHU metabolism pathways.** Genes encoding DehR from different sources (*Vibrio splendidus* 12B01, *Agrobacterium tumefaciens* C58 or *Vibrio harveyi*), KdgK from *E. coli*, KdgpA from *Vibrio splendidus* 12B01, and *Ac_DHT1* were codon-optimised for *S. cerevisiae* expression (with the exception of *Agrobacterium tumefaciens* DehR) and synthesized (DNA2.0 Inc.). The synthesized sequences were cloned into plasmids downstream of strong, constitutive *S. cerevisiae* promoters (*pFBA1* and *pTDH3*). These sequences, along with the genes and associated promoters from pM3, were subcloned into plasmids that contained wild-type biosynthetic genes and their associated 5′ and 3′ untranslated regions (UTRs). DNA constructs for genomic integration that contained these genes were generated by restriction enzyme digestion of plasmids and then integrated into the *S. cerevisiae* genome at the appropriate auxotrophic loci to create the prototrophic host strains BAL2759 (containing *V. splendidus* DehR), BAL2772 (containing *Agrobacterium tumefaciens* DehR), and BAL2956 (containing *V. harveyi* DehR). The genotypes of these strains are listed in Supplementary Table 1.

**Media and growth conditions.** All adaptation and ethanol production experiments were carried out using a defined synthetic medium that was supplemented with a mixture of amino acids as a nitrogen source. In general, experiments in which the media contained less than 4% carbon were carried out using a low-carbon (LC) medium (Supplementary Table 10). Experiments with media containing greater than 4% carbon were conducted using the high carbon (HC) medium (Supplementary Table 10). The pH levels of all media were adjusted to 5.5 with 13 N NaOH before filter sterilization. The pH levels of all shake flask and screw-cap bottle cultures were maintained throughout the experiments by supplementation with 20–50 mM MES buffer. All fermentation experiments were supplemented with Tween 80 (0.42 g $l^{-1}$) and ergosterol (10 g $l^{-1}$).

**Adaptation of engineered *S. cerevisiae* strains under aerobic conditions.** BAL2759, BAL2772 and BAL2956 strains with high specific growth rates in DEHU media were selected by serial transfer in 125 ml unbaffled shake flasks. Cultures were inoculated in 25 ml LC media with 8.8 g $l^{-1}$ DEHU as the sole carbon source. Growth was monitored by $D_{600\,nm}$ throughout the experiment. The cells were diluted into fresh media before the culture density reached 50% of the maximum cell mass supported by the medium. During serial transfers, the cultures were diluted at least 100-fold, and the process was repeated as needed for a period of 4–6 months. The DEHU concentration was increased gradually to 17.6 g $l^{-1}$ in all cultures and increased again to 35 g $l^{-1}$ in the case of BAL2956. All serial transfer flasks were grown on a rotary shaker and maintained at 29 °C with constant shaking and controlled humidity (250 r.p.m. and 90% humidity).

**Microaerobic ethanol production from DEHU and mannitol.** Yeast cultures were grown in HC medium containing 3% DEHU and 6% mannitol to prepare inoculums for ethanol production experiments. The inoculums for these starter cultures were taken from exponentially growing mid-log cultures from the adaptation experiments after at least 150 generations of selective pressure for DEHU-dependent growth (Supplementary Fig. 13). Cells were acclimated through two rounds of culturing in the aforementioned media. The second-round starter cultures were grown to mid-log phase and collected by centrifugation. These cell pellets were used to inoculate the ethanol production experiments.

Ethanol production was first carried out in 125 ml screw-cap bottles with silicone septa, magnetic stir bars, and working volumes of 30 ml each. The cells were co-fed with 3.4% DEHU and 4.4% mannitol in HC medium to select strains for follow-up experiments. Upon inoculation, cultures were placed in a water bath at 29 °C on top of a magnetic stir plate set to 500 r.p.m. Samples were taken periodically via syringe sampling through the septa for HPLC analysis, DEHU quantification, and $D_{600\,nm}$ measurements.

**Selection of a BAL2956 strain variant that is capable of growth under anaerobic conditions.** The adapted BAL2956 strain with a high specific growth rate in HC media containing DEHU as the substrate was unable to produce ethanol efficiently without maintaining microaerobic conditions. Therefore, cells that were capable of growing on agar plates with HC media containing 2% DEHU and 4% mannitol under anaerobic conditions were selected. A total of 50 μl of BAL2956 aerobic culture was plated onto several agar plates. The plates were enclosed in a tightly sealed container with AnaeroPack (Mitsubishi Gas Chemical) and incubated at 30 °C for several days. Within the first 5 days, approximately 200 colonies appeared on each plate. To isolate single colonies, several of these colonies were re-streaked onto agar plates containing fresh HC media supplemented with 2% DEHU and 4% mannitol. These colonies were grown on HC media supplemented with 3% DEHU aerobically, and the colony that showed the fastest growth rate in this medium was chosen for further experiments (BAL3215).

**Ethanol fermentation from DEHU and mannitol.** Seed yeast cultures of the BAL3215 were aerobically grown in HC medium containing approximately 1:2 molar ratio of DEHU:mannitol at 6.5% (w/v) and 2% DEHU to prepare inoculums for ethanol production experiments. Ethanol fermentation experiments were carried out in 125 ml screw-cap bottles with silicone septa, magnetic stir bars and working volumes of 55 ml each (the head space was flushed with nitrogen gas). The cells were co-fed with an approximately 1:2 molar ratio of DEHU and mannitol at 6.5% (w/v) and 9.8% (w/v) of total sugars. Upon inoculation (the culture grown in DEHU/mannitol media was inoculated into 6.5% (w/v) fermentation media, while the culture grown in DEHU media was inoculated into 9.8% (w/v) fermentation media), the cultures were placed in a water bath at 29 °C on top of a magnetic stir plate set to 350 r.p.m. Samples were taken periodically via syringe sampling through the septa for HPLC analysis and dry cell weight quantification.

**Quantification of mannitol, DEHU, ethanol, glycerol and acetate in culture media.** The metabolite analyses were carried out using the HPLC Prominence system (Shimadzu) equipped with a Rezex ROA organic acid H+ 300 × 7.8 mm Phenomenex column and run with a flow rate of 0.6 ml $min^{-1}$ and a temperature of 60 °C. The method is isocratic with the mobile phase of 5 mM sulphuric acid and was run over 30 min. Analytes were detected using a refractive index detector (Shimadzu).

**Quantification of DEHU.** The DEHU standard curve for HPLC analysis was generated using At_DehR (prepared in house). The reaction mixture contained 50 mM HEPES at pH 7.4, 100 mM NaCl, 50 mM KCl, 5 mM $MgCl_2$, 0.005% (w/v) Triton X-100, 4 mM NADPH, and 0.132 μg $\mu l^{-1}$ DehR in a 50 μl total assay volume. DEHU was quantified based on the consumption of NADPH, measured by monitoring the *A* at 340 nm using a plate reader.

31. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18,** 821–829 (2008).
32. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnol.* **31,** 46–53 (2013).

LETTER

doi:10.1038/nature12773

# *De novo* protein crystal structure determination from X-ray free-electron laser data

Thomas R. M. Barends[1], Lutz Foucar[1], Sabine Botha[1], R. Bruce Doak[1,2], Robert L. Shoeman[1], Karol Nass[1], Jason E. Koglin[3], Garth J. Williams[3], Sébastien Boutet[3], Marc Messerschmidt[3] & Ilme Schlichting[1]

**The determination of protein crystal structures is hampered by the need for macroscopic crystals. X-ray free-electron lasers (FELs) provide extremely intense pulses of femtosecond duration, which allow data collection from nanometre- to micrometre-sized crystals[1–4] in a 'diffraction-before-destruction' approach. So far, all protein structure determinations carried out using FELs have been based on previous knowledge of related, known structures[1–5]. Here we show that X-ray FEL data can be used for *de novo* protein structure determination, that is, without previous knowledge about the structure. Using the emerging technique of serial femtosecond crystallography[1–4,6], we performed single-wavelength anomalous scattering measurements on microcrystals of the well-established model system lysozyme, in complex with a lanthanide compound. Using Monte-Carlo integration[6,7], we obtained high-quality diffraction intensities from which experimental phases could be determined, resulting in an experimental electron density map good enough for automated building of the protein structure. This demonstrates the feasibility of determining novel protein structures using FELs. We anticipate that serial femtosecond crystallography will become an important tool for the structure determination of proteins that are difficult to crystallize, such as membrane proteins[1,2,8].**

Detailed knowledge of protein structures provides essential insight into their function at the molecular level, helping to elucidate basic biological processes and guiding the design of new drugs for medical applications. The vast majority of protein structures are determined by X-ray crystallography. Since its beginning almost 100 years ago, progress in crystallography has closely reflected advances in X-ray sources and other instrumentation, allowing high-throughput approaches and the analysis of ever smaller crystals. The latter is important because many macromolecules, particularly membrane proteins, are difficult to crystallize, often yielding only very small crystals. The analysis of these, however, is complicated by radiation damage.

This problem is now being solved through the use of X-ray free-electron lasers (FELs). These provide femtosecond, ultrabright X-ray pulses which are so brief that useful diffraction data can be collected before the sample is destroyed by radiation damage. Since the recent initial demonstration of structural-biological studies using X-ray FELs[1–5], these devices have already expanded the possibilities of protein crystallography, as shown most recently by X-ray data collection of the undamaged electronic structure of the highly radiation-sensitive $Mn_4CaO_5$ cluster of photosystem II at room temperature[8]. Although FEL data collection is fundamentally different from conventional crystallography and data processing techniques are still being developed, it is already evident that high-resolution data can be collected from micron-sized crystals at ambient temperature. For example, FEL data on lysozyme as a model system agree well with low-dose synchrotron data despite a dose of 30 MGy per crystal[3], the typical dose limit for data collection at cryogenic temperature[9]. The FEL analysis of micron-sized crystals of the trypanosomal protease cathepsin B revealed new high-resolution structural features that showed how the native protease is inhibited[4]. These findings, and the demonstration that X-ray diffraction data from nanometre- to micrometre-sized crystals of large membrane protein complexes[1,2,5] can be collected, illustrate the potential of FEL-based protein crystallography as a new tool for the analysis of the large group of proteins that are difficult to crystallize.

Crystallographic structure determination requires the retrieval of information about the phases of the diffracted radiation which is lost during the measurement of intensities. So far, the data for all FEL-determined protein structures have been phased by molecular replacement[1–5] using phases from known, related structures. However, to establish FEL-based crystallography as a true stand-alone tool for macromolecular structure determination, it is essential to show that the data can be phased *de novo*. FEL-specific phasing approaches have been proposed, such as oversampling-based methods[10], as well as a variation on phasing induced by radiation damage[11,12] using the extremely high intensity of FEL pulses to change the anomalous scattering behaviour of specific atoms[13]. However, there is no apparent reason why conventional phasing methods cannot be used: these include the use of heavy atoms in multiple or single isomorphous replacement approaches, with or without multi- or single-wavelength anomalous diffraction (MAD or SAD) measurements which exploit element-specific scattering at X-ray absorption edges.

Protein crystallography at X-ray FEL sources typically uses the emerging method of serial femtosecond crystallography (SFX). Given the extremely high intensity pulses generated by FELs, every X-ray exposure results in the complete destruction of the sample. Hence every diffraction image requires a new crystal, and these are introduced into the beam in a thin column of liquid[14,15]. Moreover, during the femtosecond-timescale exposure, no rotation of the crystal can be performed as in conventional crystallography, so that only still images containing partially recorded reflections are obtained. Variations in crystal size and quality and the lack of control over the crystal orientations complicate SFX data processing. Furthermore, the spectrum, intensity and beam profile of an FEL source typically vary significantly from shot to shot, further complicating the analysis. In SFX this is solved by averaging over a large number of exposures (Monte-Carlo integration), which then yields fully integrated diffraction intensities[6,7,16].

Given an adequate number of exposures, SFX-derived data are sufficiently accurate to reveal even small features such as the differences in electron density between different side chains[3] and the very weak anomalous diffraction from endogenous sulphur atoms in a protein[17]. However, these experiments relied on molecular replacement to obtain phases. *De novo* phasing from a heavy atom derivative, on the other hand, requires an accuracy well beyond that needed to find the heavy atom(s) in phased difference Fourier maps.

To test whether sufficient accuracy can be attained for *de novo* phasing by SFX, we collected a highly redundant, 2.1 Å resolution SFX data set for a lysozyme heavy atom derivative that gives a strong anomalous signal from two gadolinium atoms per asymmetric unit[18]. Lysozyme microcrystals were soaked in gadoteridol, an organic gadolinium complex,

[1]Max-Planck Institute for Medical Research, Jahnstrasse 29, D-69120 Heidelberg, Germany. [2]Department of Physics, Arizona State University, PO Box 871504, Tempe, Arizona 85287-1504, USA. [3]SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA.

and injected in their soaking solution into the vacuum chamber of the coherent X-ray imaging instrument (CXI)[19] of the LCLS at the SLAC National Accelerator Laboratory using a liquid microjet[14] essentially as described previously[3]. X-ray pulses of nominally 50 fs duration and 8.5 keV photon energy of 2.6 mJ average power were used to collect 2,402,199 diffraction patterns. Of these 191,060 crystal diffraction patterns (8%) were identified using CASS[20], 59,667 of which were indexed (31%) and integrated using CrystFEL[7].

As expected from the quality of the anomalous difference Patterson map calculated using ~60,000 diffraction patterns (Fig. 1), heavy atom binding sites could be determined using automated methods (SHELXD[21]) and refined using SHARP[22]. The resulting refined sites were used for likelihood-based SAD phasing using PHASER for Experimental Phasing[23] followed by density modification using DM[24]. Subsequently, automatic building by ARP/wARP[25] resulted in correct main-chain tracing for over 60% of the structure. Iterative cycles of rebuilding and refinement resulted in a model at 2.1 Å resolution with good statistics (see Extended Data Table 1). The quality of the phases at the various stages of the phasing process is shown in Fig. 2, a sequence of electron density maps from the various stages of the phasing process is shown in Fig. 3. Figure 3 also indicates $CC_{map}$ at each stage, that is, the correlation of the electron density map with the final electron density phased by the refined model.

As expected for SFX data[6,7,16], the data quality depends strongly on the number of integrated patterns (Fig. 4). In SFX, an internal quality measure called $R_{split}$ is used[7], which is calculated by splitting the set of diffraction images into two halves, normally even- and odd-numbered images, integrating each half to obtain two sets of intensities and calculating the $R$-factor

$$R_{split} = \frac{1}{\sqrt{2}} \cdot \frac{\sum_{hkl} \left| I_{hkl}^{even\_images} - I_{hkl}^{odd\_images} \right|}{\sum_{hkl} \frac{1}{2} \left( I_{hkl}^{even\_images} + I_{hkl}^{odd\_images} \right)}$$

between them. $R_{split}$ is closely related to $R_{merged\text{-}I}$[26] used in conventional crystallography, apart from the factor $\sqrt{2}$ in the denominator, which is introduced to account for the reduction in multiplicity caused by splitting the data set in two. $R_{merged\text{-}I}$, in turn, can be cast in an analytical form known as $R_{p.i.m.}$[27], the precision-indicating merging $R$-factor. All these $R$-factors account for the increase in data quality (signal-to-noise ratio) afforded by high measurement multiplicities.

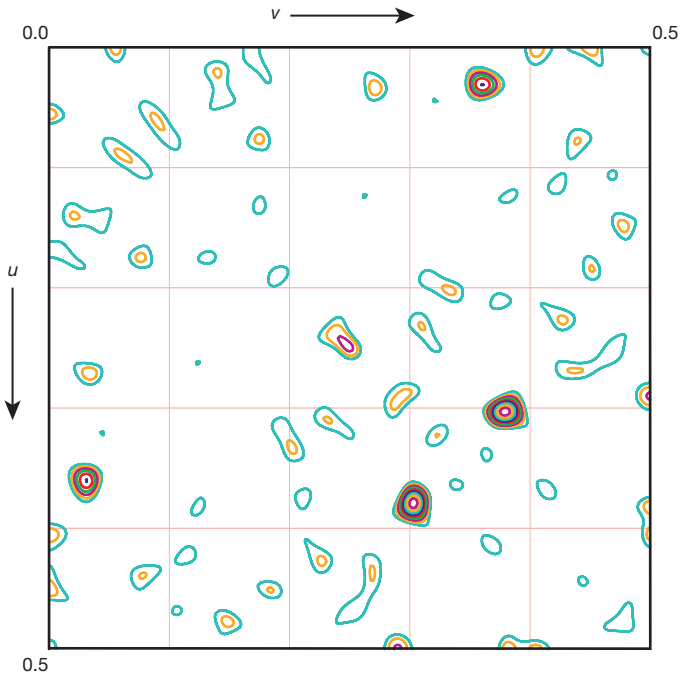



**Figure 1 | $w = 0.5$ section of the origin-removed, super-sharpened anomalous difference Patterson map of the SFX lysozyme gadolinium data, using ~60,000 images.** Clear peaks are observed from the anomalous scattering of the gadolinium atoms. This figure was prepared using XPREP.

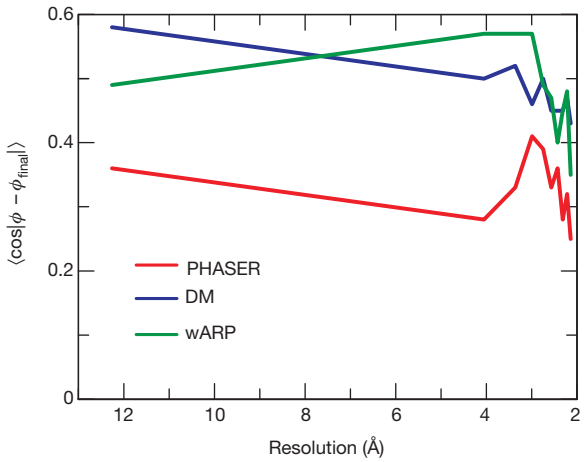



**Figure 2 | Quality of the phases at the various stages of the phasing process.** As another, resolution-dependent measure of the progress of the phasing process through its various stages, we calculated the average cosine of the difference between the phases at a certain stage and the final, refined phases, in analogy to the classical FOM measure. The quality of the phases as measured by $\langle \cos|\phi - \phi_{final}| \rangle$ after SAD phasing improved markedly after solvent flattening over the whole resolution range, after which wARP provided further improvement at intermediate resolution.

Importantly, it has been suggested that $R_{p.i.m.}$ can be used to predict whether the data are of sufficient quality for successful structure solution by comparing it to $R_{ano}$, an $R$-factor based on anomalous intensity differences $\frac{\sum_{hkl} |I_{hkl} - I_{-h-k-l}|}{\sum_{hkl} \frac{1}{2}(I_{hkl} + I_{-h-k-l})}$. Whereas as a measure of anomalous signal strength $R_{ano}$ is compromised by measurement errors[28], the ratio $R_{ano}/R_{p.i.m.}$ is a useful predictor of the feasibility of substructure solution and phasing[27–29].

To see whether $R_{split}$ has a similar predictive value, we calculated $R_{ano}/R_{split}$ for data sets calculated using 60,000, 30,000, 15,000 and 7,500 images and found it to be 1.8, 1.4, 1.2 and 1.1, respectively. When all ~60,000 patterns (red line, Fig. 4) were used, excellent values of $R_{split}$ were obtained, as well as a high anomalous correlation coefficient $CC_{ano}$ of 0.48. Using a lower number of images (green, blue and purple lines, Fig. 4) resulted in significantly increased $R_{split}$, but more importantly in a marked decrease in the anomalous correlation coefficient $CC_{ano}$.

In the anomalous difference Patterson map (Fig. 1), too, a strong dependence of peak height on the number of integrated patterns was observed. The Patterson peak at $(u,v,w) = (0.38, 0.30, 0.50)$ provides a good example, its height being 6.7, 5.7, 4.0 and 3.1 σ using 60,000, 30,000, 15,000 and 7,500 patterns, respectively. Using SHELXD[21], correct heavy atom sites could be found in all these cases, but only with 60,000 images did phasing result in an interpretable map. Thus, in this particular case at least, there appears to be a correlation between successful phasing and $R_{ano}/R_{split}$. Extended Data Fig. 1 shows $R_{ano}$ and $R_{split}$ as a function of resolution.

Lanthanide LIII edges give a very strong anomalous signal, with large values of $f''$ close to the $CuK_\alpha$ wavelength, making lanthanide derivatives attractive for both in-house and synchrotron measurements[18]. We were therefore surprised by the relatively low $CC_{ano}$ and $R_{ano}/R_{split}$ of the SFX data and collected a low-dose, room-temperature data set from a large crystal mounted in a capillary using a rotating anode

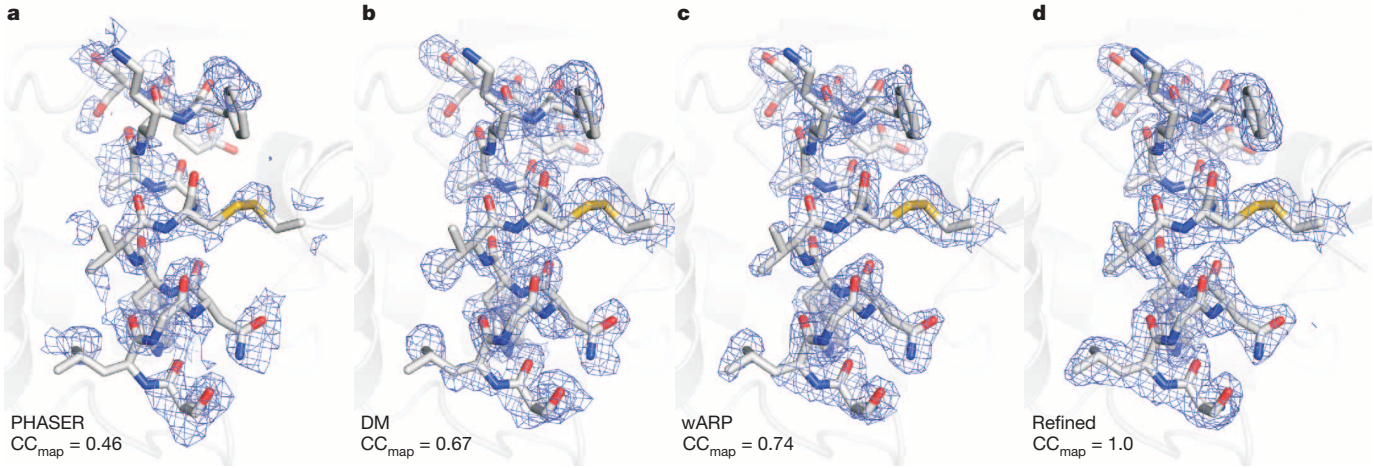



**Figure 3 | Progression of the phasing process.** Electron density maps are shown. **a**, SAD phasing with PHASER. **b**, Solvent flattening with DM. **c**, Automatic building using wARP. **d**, Final map after refinement. The correlation between the respective maps and the final, refined $2mF_o - DF_c$ electron density (**d**) is indicated. All maps are contoured at 1.0 $\sigma$.

source. As expected, the anomalous signal was very strong ($CC_{ano} = 0.92$, $R_{ano}/R_{p.i.m.}$, see Extended Data Table 1 and Extended Data Fig. 1) and phasing was possible using a fully automated approach. Although these data are not completely isomorphous with our SFX data no large systematic differences are apparent. The lower $CC_{ano}$ of the SFX data is therefore probably owing to the fact that the SFX data have not yet fully converged, because $CC_{ano}$ and $R_{ano}/R_{split}$ are still increasing with the number of patterns used (see Fig. 4 and above). This is, however, not a fundamental hindrance, because the number of collected diffraction images was limited by the available beam time. The anticipated higher repetition rates of both upgraded and new FEL facilities will greatly speed up the acquisition of high-multiplicity data sets. In addition, the use of seeded FEL pulses[30] with their expected higher reproducibility and 'cleaner' spectrum may result in faster convergence of the merged reflection intensities. Moreover, we expect that improvements in data processing, such as the inclusion of profile fitting and post-refinement, will reduce the number of patterns required for convergence. Lastly, continued development of the current, first-generation fast detectors will improve the hitherto limited accuracy of diffraction measurements and thus of the derived anomalous (and isomorphous) differences. All of these improvements are expected to reduce the number of patterns and thus sample quantity required. This will be especially important if the weak anomalous signal from endogenous sulphur is to be used[17] (see Extended Data Fig. 2).

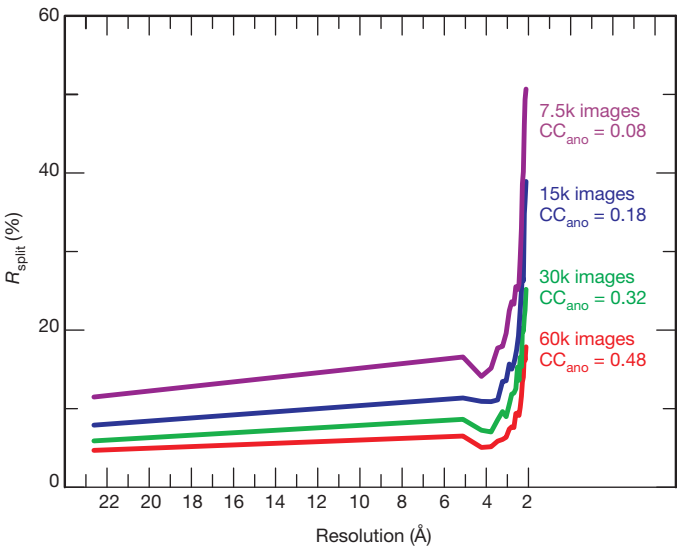



**Figure 4 | Data quality as a function of resolution and number of indexed patterns used to derive integrated intensities as shown by $R_{split}$.** The anomalous correlation coefficient $CC_{ano}$ for the whole resolution range is indicated as well.

Despite the current limitations of the experimental setup, we have shown here that X-ray free-electron lasers provide diffraction data from small crystals that are sufficiently accurate for *de novo* phasing. This proof-of-principle demonstration opens a unique and novel path to the determination of structures that elude traditional analysis. Such targets include radiation-sensitive samples such as the very small crystals that are often the only ones available for many macromolecules that are difficult to crystallize, in particular membrane proteins and their complexes.

## METHODS SUMMARY

Microcrystals ($\leq 1 \times \leq 1 \times \leq 2\ \mu m^3$) of hen egg-white lysozyme were grown as described earlier[3] and transferred to 8% NaCl, 0.1 M sodium acetate buffer, pH 4.0. 100 mM gadoteridol[18] ($Gd^{3+}$:10-(2-hydroxypropyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triacetic acid) was added to this storage solution and the crystals left to incubate at room temperature for at least 30 min before data collection. SFX diffraction data were collected using X-ray pulses of 50 fs duration (electron bunch length) and 8.5 keV photon energy of 2.6 mJ average power, essentially as described previously[3]. A suspension of 30% (v/v) of gadoteridol-treated lysozyme crystals in their soaking solution was injected into the 1 µm focus chamber of the CXI instrument[19] at LCLS using a liquid jet[14] of 3–4 µm diameter running at 30 µl $min^{-1}$. Single-shot diffraction patterns were collected using a CSPAD detector at 120 Hz. All frames were saved, and protein diffraction patterns were identified using CASS[20] and indexed and integrated using CrystFEL[7]. Using data to 1.8 Å resolution, heavy atom binding sites were determined using SHELXD[21] and refined using SHARP[22]. The resulting refined sites were used for likelihood-based SAD phasing using PHASER for Experimental Phasing[23] followed by density modification using DM[24]. Subsequently, automatic building by ARP/wARP resulted in correct main-chain tracing for over 60% of the structure[25]. Iterative cycles of rebuilding and refinement resulted in a model at 2.1 Å resolution with good statistics. A comparison data set using a single large lysozyme crystal soaked with gadoteridol was collected using a rotating anode, similarly as described previously[3]. A full description of the methods is available in Supplementary Information.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 23 July; accepted 14 October 2013.**
**Published online 24 November 2013.**

1. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470,** 73–77 (2011).

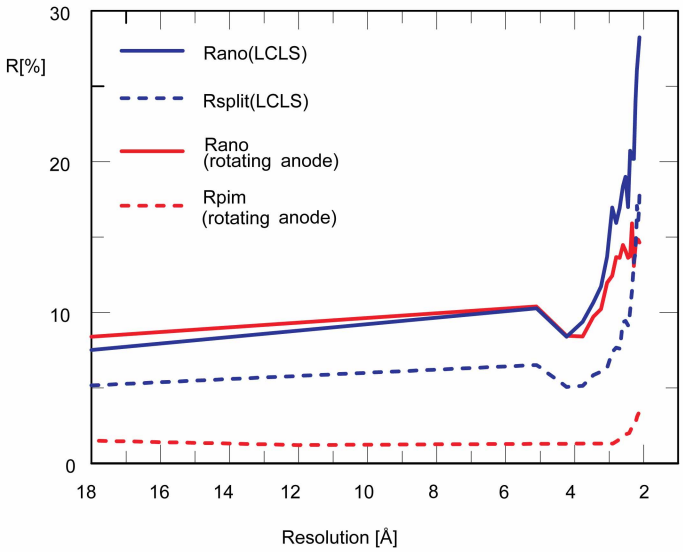
2. Johansson, L. C. *et al.* Lipidic phase membrane protein serial femtosecond crystallography. *Nature Methods* **9,** 263–265 (2012).
3. Boutet, S. *et al.* High-resolution protein structure determination by serial femtosecond crystallography. *Science* **337,** 362–364 (2012).
4. Redecke, L. *et al.* Natively inhibited *Trypanosoma brucei* cathepsin B structure determined by using an X-ray laser. *Science* **339,** 227–230 (2013).
5. Kern, J. *et al.* Room temperature femtosecond X-ray diffraction of photosystem II microcrystals. *Proc. Natl Acad. Sci. USA* **109,** 9721–9726 (2012).
6. Kirian, R. A. *et al.* Femtosecond protein nanocrystallography-data analysis methods. *Opt. Express* **18,** 5713–5723 (2010).
7. White, T. A. *et al.* CrystFEL: a software suite for snapshot serial crystallography. *J. Appl. Cryst.* **45,** 335–341 (2012).
8. Kern, J. *et al.* Simultaneous femtosecond X-ray spectroscopy and diffraction of photosystem II at room temperature. *Science* **340,** 491–495 (2013).
9. Owen, R. L., Rudino-Pinera, E. & Garman, E. F. Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proc. Natl Acad. Sci. USA* **103,** 4912–4917 (2006).
10. Spence, J. C. H. *et al.* Phasing of coherent femtosecond X-ray diffraction from size-varying nanocrystals. *Opt. Express* **19,** 2866–2873 (2011).
11. Banumathi, S., Zwart, P. H., Ramagopal, U. A., Dauter, M. & Dauter, Z. Structural effects of radiation damage and its potential for phasing. *Acta Crystallogr. D* **60,** 1085–1093 (2004).
12. Ravelli, R. B. G., Leiros, H.-K. S., Pan, B., Caffrey, M. & McSweeney, S. Specific radiation damage can be used to solve macromolecular crystal structures. *Structure* **11,** 217–224 (2003).
13. Son, S.-K., Chapman, H. N. & Santra, R. Multiwavelength anomalous diffraction at high X-ray intensity. *Phys. Rev. Lett.* **107,** 218102 (2011).
14. Weierstall, U., Spence, J. C. H. & Doak, R. B. Injector for scattering measurements on fully solvated biospecies. *Rev. Sci. Instrum.* **83,** 035108 (2012).
15. Sierra, R. G. *et al.* Nanoflow electrospinning serial femtosecond crystallography. *Acta Crystallogr. D* **68,** 1584–1587 (2012).
16. Kirian, R. A. *et al.* Structure-factor analysis of femtosecond micro-diffraction patterns from protein nanocrystals. *Acta Crystallogr. A* **67,** 131–140 (2011).
17. Barends, T. R. M. *et al.* Anomalous signal from S atoms in protein crystallographic data from an X-ray free-electron laser. *Acta Crystallogr. D* **69,** 838–842 (2013).
18. Girard, E., Chantalat, L., Vicat, J. & Kahn, R. Gd-HPDO3A, a complex to obtain high-phasing-power heavy-atom derivatives for SAD and MAD experiments: results with tetragonal hen egg-white lysozyme. *Acta Crystallogr. D* **58,** 1–9 (2002).
19. Boutet, S. & Williams, G. J. The Coherent X-ray Imaging (CXI) instrument at the Linac Coherent Light Source (LCLS). *New J. Phys.* **12,** 035024 (2010).
20. Foucar, L. *et al.* CASS-CFEL-ASG software suite. *Comput. Phys. Commun.* **183,** 2207–2213 (2012).
21. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D* **66,** 479–485 (2010).
22. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364,** 215–230 (2007).
23. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
24. Cowtan, K. 'dm': An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31,** 34–38 (1994).
25. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protocols* **3,** 1171–1179 (2008).
26. Diederichs, K. & Karplus, P. A. Improved *R*-factors for diffraction data analysis in macromolecular crystallography. *Nature Struct. Biol.* **4,** 269–275 (1997).
27. Weiss, M. S. Global indicators of X-ray data quality. *J. Appl. Crystallogr.* **34,** 130–135 (2001).
28. Weiss, M. S., Sicker, T. & Hilgenfeld, R. Soft X-rays, high redundancy, and proper scaling: a new procedure for automated protein structure determination via SAS. *Structure* **9,** 771–777 (2001).
29. Panjikar, S. & Tucker, P. A. Phasing possibilities using different wavelengths with a xenon derivative. *J. Appl. Crystallogr.* **35,** 261–266 (2002).
30. Amann, J. *et al.* Demonstration of self-seeding in a hard-X-ray free-electron laser. *Nature Photon.* **6,** 693–698 (2012).


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

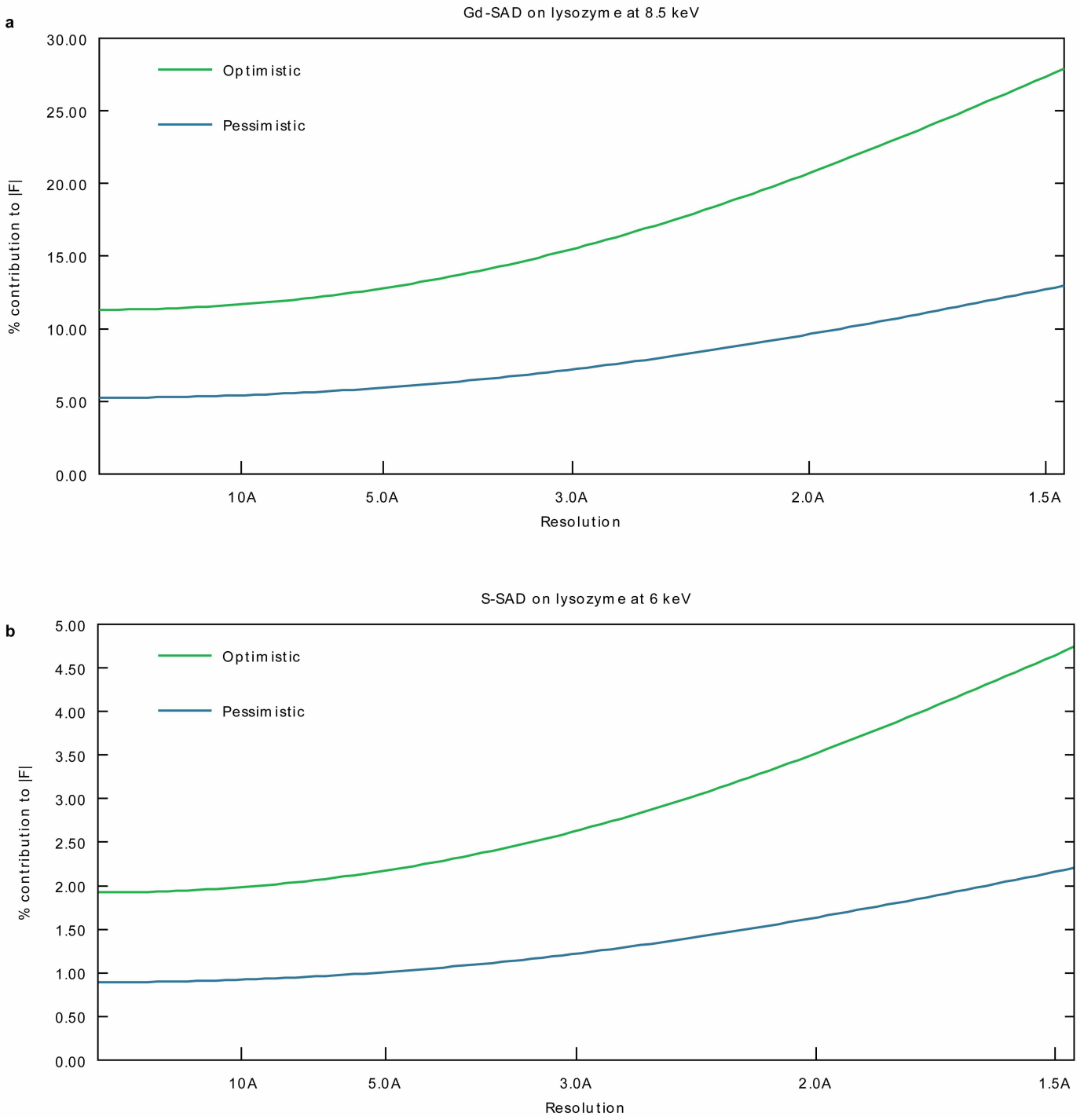
**Acknowledgements** Portions of this research were carried out at the Linac Coherent Light Source, a National User Facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. The CXI instrument was funded by the LCLS Ultrafast Science Instruments (LUSI) project funded by the US Department of Energy, Office of Basic Energy Sciences. We acknowledge support from the Max Planck Society and from the EU for an Incoming Scientist Award to R.B.D. We thank the staff at the LCLS for their support and are grateful to S. Pesch and R. van Gessel (Bracco Imaging Konstanz and Singen, Germany) for the gift of the sample of gadoteridol. We thank H. Zimmermann for suggestions, W. Kabsch for discussions and J. Wray for critically reading the manuscript. In addition, we acknowledge L. Hammon and C. Patty for laboratory support, and the MCC staff for the beam they provided. We are indebted to C. Roome and F. Koeck for computing support.

**Author Contributions** T.R.M.B. and I.S. conceived the research, I.S. prepared crystals, Sa.B., R.B.D. and R.L.S. performed sample injection. Sé.B., G.J.W., J.E.K. and M.M. performed data collection, T.R.M.B., L.F. and K.N. performed data processing and analysis. T.R.M.B. and I.S. wrote the manuscript with input from all authors.

**Author Information** Structure factor amplitudes and anomalous differences have been deposited in the Protein Data Bank along with the refined structure with accession code 4N5R, and diffraction patterns of crystal hits will be deposited at http://cxidb.org/. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.R.M.B. (thomas.barends@mpimf-heidelberg.mpg.de) or I.S. (ilme.schlichting@mpimf-heidelberg.mpg.de).




**Extended Data Figure 1 | Anomalous signal strength of the SFX data (blue lines) as well as the rotating anode data (red lines) as measured by $R_{ano}$ on intensities (solid lines).** The noise in the data is indicated in terms of $R_{split}$ for the SFX data and $R_{p.i.m.}$ for the rotating anode data (dashed lines).




**Extended Data Figure 2 | Expected anomalous signal strength for a SAD experiment on lysozyme.** Expected anomalous signal strength for a SAD experiment on lysozyme with 2 gadolinium atoms per protein molecule at 8.5 keV (top panel) and for a sulphur-SAD experiment on lysozyme with 10 sulphur atoms per protein molecule at 6.0 keV (bottom panel). In each case, an optimistic scenario with all anomalous scatterers ordered is shown (green line) as well as a pessimistic scenario in which 60% of the anomalous scatterers are ordered (blue line). This figure was prepared using the anomalous scattering web server at http://skuld.bmsc.washington.edu/scatter/AS_signal.html.

**Extended Data Table 1 | Data collection, phasing and refinement statistics**

| | LCLS SFX data | Rotating anode data |
|---|---|---|
| **Data collection** | | |
| Space group | $P4_32_12$ | $P4_32_12$ |
| Cell dimensions[a] | | |
| *a*, *b*, *c* (Å) | 79, 79, 39 | 78.1, 78.1, 39.2 |
| *α*,*β*,*γ* (º) | 90, 90, 90 | 90, 90, 90 |
| Wavelength (Å) | 1.45 | 1.5418 |
| Pulse energy/fluence at sample | 60 μJ/ $8 \cdot 10^{23}$ photons/s [b] | $10^7$ photons/s |
| Dose (MGy) | 22 MGy/crystal | 1 kGy/dataset |
| Number of collected diffraction patterns | 2,402,199 | 360 (high-res run) |
| | | 140 (low-res run) |
| Number of crystal hits | 191,060 | 500 |
| Number of indexed images patterns | 59,667 | 500 |
| Resolution (Å) [c] | 40-2.1 (2.14-2.10) | 40-1.8 (1.83-1.8) |
| $R_{split} = \frac{1}{\sqrt{2}} \cdot \frac{\sum_{hkl} \left\| I_{hkl}^{even\_images} - I_{hkl}^{odd\_images} \right\|}{\sum_{hkl} 1/2 \left( I_{hkl}^{even\_images} + I_{hkl}^{odd\_images} \right)}$ | 0.061 (0.179) | n.a. |
| $R_{merge}$ | n.a.[d] | 0.030 (0.068) |
| CC* [25] | 1.0 (0.99) | 1.0 (0.99) |
| $CC_{1/2}$ | 0.99 (0.96) | 1.0 (0.95) |
| $CC_{ano}$ | 0.48 | 0.92 |
| $I/\sigma I$ | 11.9 (4.7) | 39.7(7.3) |
| Completeness (%) | 100 (100) | 96.6 (93.9) |
| SFX multiplicity of observations[e] | 1383.5 (1310.0) | n.a. |
| Redundancy[e] | n.a. | 16.1 (11.7) |
| **Refinement** | | |
| Resolution (Å) | 40-2.1 | |
| No. reflections | 7287 | |
| $R_{work}$/ $R_{free}$ | 0.230/0.259 | |
| No. atoms | | |
| Protein | 992 | |
| Ligand/ion | 58 (2 gadoteridol) | |
| Water | 70 | |
| B-factors | | |
| Protein | 29.9 | |
| Ligand/ion | 39.5 | |
| Water | 43.7 | |
| R.m.s deviations | | |
| Bond lengths (Å) | 0.008 | |
| Bond angles (º) | 1.17 | |

[a]In SFX, only cell dimensions averaged over all crystals used can be reported, causing the lower accuracy reported here. Also, the difference with the cell dimensions of the macroscopic crystal used for the rotating anode data set points to limited nonisomorphism.
[b]Beamline transmission 50%; the resulting 1.3 mJ pulse energy was attenuated to 4% transmission.
[c]Highest resolution shell is shown in parenthesis.
[d]In SFX, all reflections are partials and no separate measurements of full reflections are recorded; Therefore, $R_{merge}$ defined as the normalized sum of the absolute differences between single observations of fully integrated reflections and their mean value, is meaningless.
[e]For the SFX data, 59,667 crystals where used, whereas for the rotating anode data, a single crystal was used.

CORRIGENDUM

doi:10.1038/nature12813

# Corrigendum: The zebrafish reference genome sequence and its relationship to the human genome

Kerstin Howe, Matthew D. Clark, Carlos F. Torroja, James Torrance, Camille Berthelot, Matthieu Muffato, John E. Collins, Sean Humphray, Karen McLaren, Lucy Matthews, Stuart McLaren, Ian Sealy, Mario Caccamo, Carol Churcher, Carol Scott, Jeffrey C. Barrett, Romke Koch, Gerd-Jörg Rauch, Simon White, William Chow, Britt Kilian, Leonor T. Quintais, José A. Guerra-Assunção, Yi Zhou, Yong Gu, Jennifer Yen, Jan-Hinnerk Vogel, Tina Eyre, Ruby Banerjee, Jianxiang Chi, Beiyuan Fu, Elizabeth Langley, Sean F. Maguire, Gavin Laird, David Lloyd, Emma Kenyon, Sarah Donaldson, Harminder Sehra, Jeff Almeida-King, Jane Loveland, Stephen Trevanion, Matt Jones, Mike Quail, Dave Willey, Adrienne Hunt, John Burton, Sarah Sims, Kirsten McLay, Bob Plumb, Joy Davis, Chris Clee, Karen Oliver, Richard Clark, Clare Riddle, David Elliott, Glen Threadgold, Glenn Harden, Darren Ware, Sharmin Begum, Beverley Mortimore, Giselle Kerry, Paul Heath, Benjamin Phillimore, Alan Tracey, Nicole Corby, Matthew Dunn, Christopher Johnson, Jonathan Wood, Susan Clark, Sarah Pelan, Guy Griffiths, Michelle Smith, Rebecca Glithero, Philip Howden, Nicholas Barker, Christine Lloyd, Christopher Stevens, Joanna Harley, Karen Holt, Georgios Panagiotidis, Jamieson Lovell, Helen Beasley, Carl Henderson, Daria Gordon, Katherine Auger, Deborah Wright, Joanna Collins, Claire Raisen, Lauren Dyer, Kenric Leung, Lauren Robertson, Kirsty Ambridge, Daniel Leongamornlert, Sarah McGuire, Ruth Gilderthorp, Coline Griffiths, Deepa Manthravadi, Sarah Nichol, Gary Barker, Siobhan Whitehead, Michael Kay, Jacqueline Brown, Clare Murnane, Emma Gray, Matthew Humphries, Neil Sycamore, Darren Barker, David Saunders, Justene Wallis, Anne Babbage, Sian Hammond, Maryam Mashreghi-Mohammadi, Lucy Barr, Sancha Martin, Paul Wray, Andrew Ellington, Nicholas Matthews, Matthew Ellwood, Rebecca Woodmansey, Graham Clark, James D. Cooper, Anthony Tromans, Darren Grafham, Carl Skuce, Richard Pandian, Robert Andrews, Elliot Harrison, Andrew Kimberley, Jane Garnett, Nigel Fosker, Rebekah Hall, Patrick Garner, Daniel Kelly, Christine Bird, Sophie Palmer, Ines Gehring, Andrea Berger, Christopher Dooley, Zübeyde Ersan-Ürün, Cigdem Eser, Horst Geiger, Maria Geisler, Lena Karotki, Anette Kirn, Judith Konantz, Martina Konantz, Martina Oberländer, Silke Rudolph-Geiger, Mathias Teucke, Christa Lanz, Günter Raddatz, Kazutoyo Osoegawa, Baoli Zhu, Amanda Rapp, Sara Widaa, Cordelia Langford, Fengtang Yang, Stephan C. Schuster, Nigel P. Carter, Jennifer Harrow, Zemin Ning, Javier Herrero, Steve M. J. Searle, Anton Enright, Robert Geisler, Ronald H. A. Plasterk, Charles Lee, Monte Westerfield, Pieter J. de Jong, Leonard I. Zon, John H. Postlethwait, Christiane Nüsslein-Volhard, Tim J. P. Hubbard, Hugues Roest Crollius, Jane Rogers & Derek L. Stemple

*Nature* **496,** 498–503 (2013); doi:10.1038/nature12111

In this Letter, five authors were inadvertently omitted: Sharmin Begum and Christine Lloyd from the Wellcome Trust Sanger Institute, and Christa Lanz, Günter Raddatz and Stephan C. Schuster from the Max Planck Institute for Developmental Biology. David Elliot was incorrectly listed as David Eliot, Beverley Mortimore was incorrectly listed as Beverly Mortimer, and James D. Cooper was incorrectly listed as James Cooper. In addition, the acknowledgements section should state that author S.C.S. was supported by the German Research Foundation (DFG Grant NU 22/5). These errors, along with corresponding minor changes to the Author Contributions section, have been corrected in the HTML and PDF versions of the original manuscript.

**TURNING POINT** Helping people motivates bioengineering PhD student **p.251**

**@NATUREJOBS** Follow us on Twitter for the latest news and features **go.nature.com/e492gf**

**NATUREJOBS** For the latest career listings and advice **www.naturejobs.com**

OIVIND HOVLAND/IKON IMAGES/CORBIS

EQUALITY

# Standing out

*Welcoming lab environments and networking organizations help lesbian, gay, bisexual and transgender scientists to excel.*

BY CAMERON WALKER

Looking up from a courtyard at the College of New Jersey in Ewing, passers-by can see a rainbow pride flag in the window of chemist Benny Chan. He has not always been so open. Chan came out as gay to just a few people in graduate school, and although he did not hide his sexual orientation once he started working at the College of New Jersey, he decided not to be vocal about it until he met the requirements for tenure. His past advisers and the administration at his new job were supportive, he says, but "there's always that little bit of doubt in my head" — one uncomfortable or discriminatory colleague could cause problems.

Scientists do not always share their personal sides in the lab. Deciding whether to be open about one's identity can be an acute issue for lesbian, gay, bisexual and transgender (LGBT) researchers. Unlike some other minorities, LGBT people "have the ability to conform, because it's not always a visible trait", Chan says. But hiding something as basic as sexual orientation or gender identity can be detrimental to mental health and work. "You need to spend a lot of extra energy if you feel like you need to hide a part of your life," says Chan.

Researchers may have trouble finding colleagues who share their experiences — which can be anything from overt or subtle discrimination to complete comfort in the workplace. Many want to know how best to support younger LGBT scientists, who might not know where to turn for mentoring.

As broader awareness of LGBT scientists grows and more of the science community starts to appreciate the issues that affect them — including same-sex marriage and anti-discrimination laws — groups are convening to foster a sense of community and, in some cases, to develop best-practice guidelines. These organizations aim to ensure that LGBT researchers get the support they need so that isolation does not keep them from being effective scientists.

## GATHERING DATA

There is a growing body of research on women and ethnic minorities in science, but the number and experience of LGBT researchers has been less widely studied. To address this, Jeremy Yoder, an evolutionary-biology postdoc at the University of Minnesota, Twin Cities, and Allison Mattheis, an educational researcher at California State University, Los Angeles, gathered more than 1,400 responses to the 'Queer in STEM' survey, which examined sexual diversity among people working in science, technology, engineering and maths (STEM) and how their identities might affect their careers. Most were from the United States, but there were also responses from Canada, the United Kingdom and India, among other countries.

Preliminary results suggest that participants who rated their workplaces as safe and welcoming and whose employers supported LGBT-specific needs — such as health-care benefits for same-sex partners in the United States — were more likely to be open with their colleagues about their identities. However, Yoder and Mattheis found that where respondents lived made no difference to how 'out' they were to colleagues or students, even if the ▶

researchers were in big cities or in regions thought to be LGBT-friendly.

Yoder and Mattheis hope that their survey results, which they are in the process of writing up for submission, will make other scientists more aware of and welcoming to LGBT researchers. The pair contend that when heterosexual researchers know about their colleagues' identities, they are more likely to support policies such as partner benefits and expanding equal-opportunity employment to cover sexual orientation and gender identity. And the authors expect that more information about the community will encourage LGBT people to enter STEM. "By making queer folk working in STEM more visible, we can help prompt STEM workplaces, professional societies and university departments to take LGBT-specific needs into consideration in policy," Yoder says.

## CREATING SAFE SPACES

LGBT researchers can turn to a growing number of support and networking groups (see 'Safe meeting spaces'). Some groups are working on best-practice guidelines to help academic departments to deal better with LGBT issues. Elena Long, a postdoc in nuclear physics at the University of New Hampshire in Durham who started the LGBT+ Physicists group in 2009, has worked with colleagues to create a guide for physics departments.

These guidelines range from changes that can be made quickly — such as using gender-neutral language in the classroom and lab or inviting LGBT speakers to campus — to those that require long-term, department-wide efforts, such as adding non-discrimination statements to job announcements and making diversity training available to faculty members and staff.

***"You need to spend a lot of extra energy, if you feel like you need to hide a part of your life."***
Benny Chan

Many institutions offer on-campus training about LGBT issues (often called 'safe zone' training). This usually consists of a several-hour session in which participants learn about resources for LGBT students and the community itself. They may receive stickers that they can place on their office doors to identify safe spaces in which people are welcome to discuss LGBT issues. Some institutions have diversity offices that run these programmes. Independent organizations such as the Diversity Trust, based in the United Kingdom, also offer training.

Although a sticker may seem like a small effort, a study of the Safe Zone programme at Iowa State University in Ames in 2002 suggests that these programmes can improve the climate on campus by visibly affirming that the needs of LGBT students are valid, and increasing heterosexuals' awareness of both the LGBT community and their own biases (N. J Evans *J. College Student Dev.* **43,** 522–539; 2002).

The best-practices guide from LGBT+ Physicists also offers some measures to ease the path for transgender researchers. On a departmental level, simplifying the process of name changes on campus records — and indicating that changes will not affect someone's job, tenure or award applications — can be particularly meaningful for transgender people, who may have elected to transition to or have started identifying as their preferred gender during graduate studies. The CV can be a minefield: many transgender people "face an extremely difficult choice when applying to a new position", says Long. "Either risk discrimination by outing yourself as trans, or risk discrimination by leaving out a significant chunk of your past work under a different name."

GREG PARKINSON, HTTP://GREGPARKINSONPHOTO.COM

## FINDING COLLEAGUES

Meeting researchers with similar backgrounds and concerns is becoming easier. LGBT researchers have been convening an informal networking dinner at meetings of the American Astronomical Society (AAS) for more than 20 years, but "you had to know it existed", says Jane Rigby, an astrophysicist at NASA.

After several members of this group wrote a charter, the AAS Council created an official working group on LGBTIQ Equality (the I stands for 'intersex' and the Q for 'queer' or 'questioning'). The group's networking and other events now appear in the AAS conference programme. The working group is also collaborating with LGBT+ Physicists on joint best-practice guidelines.

There are also online physics and astronomy 'out' lists, to which LGBT researchers have voluntarily added their names and, in many cases, contact information so that they can be helpful to others. Both lists also include non-LGBT researchers who support the community. Some institutions, such as the University of California, San Francisco, have their own out lists.

Many institutions have LGBT networks. At CERN, Europe's particle-physics lab near Geneva, Switzerland, an LGBT group hosts social events and weekly lunches in the cafeteria to promote visibility, which is potentially helpful for LGBT visitors.

Young LGBT scientists can find both community and professional networking through mentoring. The US-based non-profit National Organization of Gay and Lesbian Scientists and Technical Professionals (NOGLSTP), which provided funding for the Queer in STEM survey, offers eight-month mentoring programmes for members. Through a partnership with MentorNet, an online STEM mentoring network, it matches undergraduate and graduate students, postdocs and other early-career professionals with mid- or later-career scientists in academia or industry. The goal

## LGBT EVENTS

### *Safe meeting spaces*

Several organizations worldwide hold conferences and events specifically aimed at lesbian, gay, bisexual and transgender (LGBT) scientists.

- Since 2010, the US National Organization of Gay and Lesbian Scientists and Technical Professionals (NOGLSTP) has put on Out to Innovate, a biennial two-day career summit that includes a career fair, workshops and speakers. In 2014 it will be in Atlanta, Georgia, and will be co-hosted with Out in STEM, a national society supporting LGBT students. The summit will facilitate mentoring and include industry representatives and tours of local companies.
- The non-profit organization Out for Work in Washington DC, which assists LGBT students with career development, runs annual conferences.
- Ecologists have been attending an informal networking LGBT lunch at the Ecological Society of America meeting since the late 1990s.
- A group of geoscientists holds an independent dinner for LGBT researchers during the annual American Geophysical Union meeting in San Francisco, California.
- The UK Gay and Lesbian Association of Doctors and Dentists has an annual conference for students and holds educational events. It also facilitates networking.
- The Australian Lesbian Medical Association in South West Rocks supports lesbian doctors and medical students, and their partners. It offers social events, mentoring and an annual meeting.
- Workplace Pride in Amsterdam holds an annual conference aimed at improving the workplace for LGBT people.
- Sticks & Stones, a diversity-focused career fair that bills itself as Europe's largest for LGBT and straight people, will be held in Berlin in 2014. Last year, several pharmaceutical and technology companies attended. C.W.

is to keep LGBT people in STEM careers and to provide someone for students to talk to if they feel that they cannot discuss their personal lives with their advisers, says Rochelle Diamond, who is the chair of the NOGLSTP's board of directors and manages two labs at the California Institute of Technology (Caltech) in Pasadena.

### LGBT MOBILITY

Changes to marriage laws in some countries may influence acceptance of LGBT people in society at large, and improve the prospects of scientists looking for the right department fit (see *Nature* **454,** 132–133; 2008). In the United States, for example, there is still a patchwork of state laws that forbid same-sex marriage. But last June, the US Supreme Court declared that the section of the Defense of Marriage Act that prohibited federal recognition of same-sex marriage was unconstitutional. That may boost the immigration of LGBT scientists, who can now sponsor foreign-born spouses for permanent-resident status. It can also help US-based researchers and their spouses. Rigby and her wife and child are now on the same insurance plan; combined with other benefits that are now permitted, they may save several thousand dollars this year.

At conferences, Carolyn Brinkworth, an astronomer at the Infrared Processing and Analysis Center at Caltech, wears a rainbow sticker with the words 'Safe Space', or a badge from an LGBT youth organization for which she volunteers. Young scientists have approached her to say that they have not felt comfortable being out at work. "It's rare that they tell me the climate is hostile," she says. More often, she says, these researchers do not want to think about introducing a potential new source of work stress by coming out, or are not sure how their advisers or peers will react to their identity.

But Chan has found that being out proved better not only on a personal level, but also on a professional one. A volunteer for the American Chemical Society (ACS), Chan discussed being gay in an ACS publication after his tenure decision. Later he received multiple e-mails from colleagues whom he knew from ACS meetings. Most were e-mails of support, but one colleague also asked him about his single-crystal X-ray diffractometer. The two have now collaborated on multiple papers.

And at an LGBT reception at an ACS meeting, Chan also met a researcher who may host his sabbatical. "Being out has really helped me," he says. "It frees you up to think of your research, and your scholarship." ■

**Cameron Walker** *is a freelance writer based in Santa Barbara, California.*

# TURNING POINT
# Eleni Antoniadou

*PhD student Eleni Antoniadou co-founded the London-based start-up Transplants Without Donors in 2009 to develop tissue-engineered organs. Antoniadou, who also blogs for* The Huffington Post*, was shortlisted in September in the science category of the 2013 Women of the Future Awards, Britain's industry-funded search for successful early-career women.*

**What led you to tissue engineering?**
I was working at a hospital as an undergraduate and saw that prosthetics had limitations. I wanted to do research that could give patients something better. I found regenerative medicine and tissue engineering to be promising fields.

**What was your first tissue-engineering project?**
While studying for a master's in nanotechnology and regenerative medicine at University College London, I worked on neural generation — testing biomaterials that could become artificial nerves. I also got involved in developing a business plan for an artificial trachea. I felt overwhelmed when it was successfully received by a patient. It was proof that tissue engineering could be applied in clinical practice.

**So you launched the start-up soon afterwards?**
While in London, I joined several physicians and scientists to co-found Transplants Without Donors so that we could work on tissue-engineering scaffolds for several different organs. In launching this company, I came to appreciate the complexity of the science behind tissue engineering. In 2010, after receiving a scholarship from the Fulbright Program and the Institute of International Education, I came to the University of Illinois at Urbana-Champaign to get a master's in bioengineering, with a focus on developing artificial skin. This is challenging, yet is a product that many patients need.

**What has been the start-up's main challenge?**
Securing financial support. But it was also challenging to find people with the appropriate multidisciplinary background. We had to learn how to design experiments so that all the scientists on our 25-member team could contribute to and understand them. We are hoping that the products we launch next year — mostly tissue-engineering scaffolds and bioreactors for different organs — will be used by other researchers. Sharing products throughout labs could really help to move the field forward.

**You spent time at NASA recently. How did that influence your research?**
I was beginning a PhD at the University of Illinois when the European Space Agency and NASA selected me to work at the biosciences division of NASA's centre for nanotechology for several months. That was a turning point in my career: it was the most innovative place I'd ever been. I saw the importance of tackling big, risky projects.

**How did you start blogging for *The Huffington Post*?**
After being nominated for the award, I was invited to write for the blog to raise awareness of the future of technology and of women in science. So far, I've written about the future of tissue-engineered organs and the importance of space exploration. Thanks to my posts, I've had scientists approach me to collaborate on projects and heard from people who are curious about tissue engineering.

**Name a pivotal moment in your career.**
In the past few years, I've been to Peru and Costa Rica to volunteer with the Foundation for the International Medical Relief of Children, a non-profit organization based in Philadelphia, Pennsylvania, that sends out teams to perform operations or offer health care. We gave vaccinations and pharmaceuticals to sick kids, including those victimized by the illegal organ trade. It was really fulfilling and has helped to drive everything we do in the lab.

**What do you plan to do after you get your PhD?**
I would like to do research in the lab, working full-time at Transplants Without Donors to bring products to market. We need to develop a legislative framework for tissue-engineering products — one that will be universal. ■

INTERVIEW BY VIRGINIA GEWIN

# OKAMI

*A matter of honour.*

BY GRACE TANG

"*Okaasan*, look!" The boy tugged on his mother's arm, pointing at the balding man splayed across the pavement. "That man is asleep on the street!"

The stench of alcohol wafted up from the unconscious man's body. While the boy's mother politely chose to ignore the man's transgression, Yuka openly stared. She gingerly picked her way over him and into the club from which he had emerged.

She scanned the tables for her client as her eyes adjusted to the dark. Hiro was sitting in a private booth, a bottle of whisky in the centre of the table.

"Hi." She slid into the booth to sit facing him.

"I already told your boss, I'm not interested in hostesses tonight."

"I'm Yuka."

He looked up from his glass, one eyebrow raised.

"You don't look the part."

"And that's precisely why I'm so good at what I do."

The left corner of his mouth raised slightly in a half smile. Given what her boss had told her of her client's reputation, Yuka considered this a minor victory. He reached into his jacket pocket and produced a phone, sliding it across the table.

"In the photo album you will find geo-tagged, time-stamped pictures of your target. I trust that is all you need."

She pocketed the phone and rose from the booth. "Consider it done."

The days were short this time of year, meaning the drinking started early, fuelled by the Western holiday season. Yuka tried to block out the drunken chorus of businessmen in the karaoke bar next door as she studied her target, looking for all the world like a teenage girl texting her friends.

Her target appeared young, probably in his mid-twenties, somewhat good-looking. For most people, the deductions would stop there, but Yuka picked up on things most people would miss, like the characteristic bump of a concealed sidearm.

The geo-tags were erratic, but mostly centred in downtown Tokyo, disturbingly near areas where certain high-ranking Yakuza members had recently been found dead. Judging from the time stamps on the photos, he was active mostly at night.

She was liking this job less and less.

She flipped through the rest of the pictures quickly, until one image made her do a double take.

She pressed the back button and zoomed in on her target's neck. She squinted long and hard until she had convinced herself without a shadow of a doubt that her eyes were not playing tricks on her.

"Son of a b—"

She was going to need more than a handgun for this job.

JACEY

"Hiro," Yuka struggled to control herself as she growled into the phone he had given her. "I want the fee doubled."

"We had a deal."

"When we struck that deal, I thought I was dealing with a *human*."

"He is technically still *part* human."

"*Technically*, he can smell my goddamn gun from two miles away!"

"So, get a gun that can shoot farther than two miles. I'm sure your boss can spare you one."

Yuka was sweating bullets. She could handle any human, but this was the first time she had had to take down an *Okami*.

She methodically set up the sniper on the ledge of the roof. It was half an hour to the new year, and the cheerful sounds of party-goers echoed from the apartments and streets below her. According to her boss's intel, her target would be exiting the opposite building about now, after completing his own hit.

Minutes passed like hours. Yuka took her eyes off the sight for a second to glance at her wristwatch. Where was he?

"Looking for me, I presume."

Yuka jumped, drawing her handgun from her waist at the same time as she spun round to face him.

He was more impressive in the flesh, his movements graceful, his voice commanding. But most surprising of all was his scent — mostly canine, a hint of human, powerful and yet profoundly … sad.

"You're not one of my hits. I won't harm you."

He seemed almost nonchalant about the fact that she was pointing her handgun at him, finger on the trigger.

Yuka lowered her weapon slightly.

"Is this a trick?"

"I wish." He laughed, unsmiling. "Perhaps you are too young, or your masters treat you well. But I am tired of being used."

Her grip on her weapon tightened. "I have no masters. The people I work for take a cut of my fee, but what I earn is my own."

"Boss, master, what's the difference? At the end of the day, we are all their dogs. Isn't that why we wear these collars?"

For the first time tonight, Yuka was aware of the cool metal of the pendant she wore around her neck, one that identified people of her species, specifically engineered for her profession.

"As I was saying, I am tired. You may complete your job."

She was shaking now, blinking furiously.

"*Sumimasen, Oniisan*. I don't want to, but … I have my obligations …"

"Yes, the loyalty of the wolf is why it was chosen to construct us. The only honourable way out of service to our masters is death."

He bowed. "*Sayonara*."

He disappeared down the stairwell. Wiping her eyes, Yuka readjusted her sniper, gazing through the night-vision-enabled sight.

The clock struck midnight, and fireworks lit up the sky, accompanied by thunderous booms and the stench of gunpowder.

"*Okaasan*! Look!" The boy tugged at his mother's sleeve and pointed. She smiled and held him tighter in her arms, their faces lit up by the showers of shimmering sparks.

No one paid any attention to the young, well-dressed man, lying prone in the middle of the pavement, as they picked their feet carefully around him. ■

NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

**Grace Tang** *is a graduate student in psychology at Stanford University.*