

# Science at scale

Highlights 2017/18



# 25 years of pushing the boundaries of breakthrough science

Through our big and bold ideas, scientific independence and cutting-edge infrastructure, we engage in long-term exploratory projects that influence science and impact people's lives on a global scale.



## Contents

### 02 What we do

---

03 Director's Introduction

04 Wellcome Sanger Institute at a Glance

### 08 Our work

---



10 Cancer, Ageing and Somatic Mutation



18 Cellular Genetics



22 Human Genetics



28 Infection Genomics



34 Malaria

### 40 Our approach

---



42 Scale



44 Innovation



46 Culture



48 Influence



50 Connections

### 52 Other information

---

52 Image Credits

53 Institute Information

Wherever you see this **dark blue text** or **yellow**, click to find more information.

What we do

Our work

Our approach

Other information



# What we do



## Director's Introduction

This year the Wellcome Sanger Institute celebrates its 25th Anniversary. We have come a long way in that time: from helping to deliver the first reference human genome to asking bold questions about the genomics underlying health and disease. We are proud of our past, but excited about our future.

The influence of genomics has progressed in the same way, with society and the NHS embracing genomic technology to transform medicine and lifestyle. While the interrogation and interpretation of genomes, once the field solely of data specialists, is now being carried out by schools across the UK. This remarkable shift is encapsulated in *Genome Decoders: Whipworm* a pioneering collaboration between Connecting Science, Sanger researchers, the Institute for Research in Schools and school students.

Our researchers are now able to interrogate an adult genome to understand its entire mutational history all the way back to the 4-8 cell stage of the embryo and discern the injuries and insults suffered by the genome on the way. They survey and understand the landscape of diversity and mutation within species spread across countries and continents to discern the rise of antibiotic and pesticide resistance. They develop new techniques to create new models of disease that more closely mirror the human condition.

Our scientists are setting new challenges and goals that seem as impossible now as sequencing the Human Genome Project did just 25 years ago. In scale: from the *Human Cell Atlas* – that aims to unpick the entire human body cell by cell – to sequencing the genomes of all life on earth. In new scientific fields: such as synthesising genomes. In health: from developing early warning systems for infectious disease outbreaks to penetrating malaria's ever-changing defensive coat to produce effective vaccines and treatments.

As we look to the next 25 years, the Sanger will continue to evolve its science focus to navigate and lead in these emerging fields.

**Professor Sir Mike Stratton, Director**  
Wellcome Sanger Institute

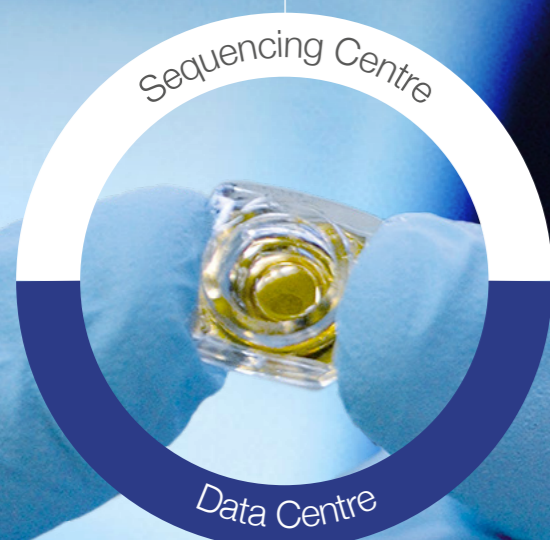
What we do

Our work

Our approach

Other information

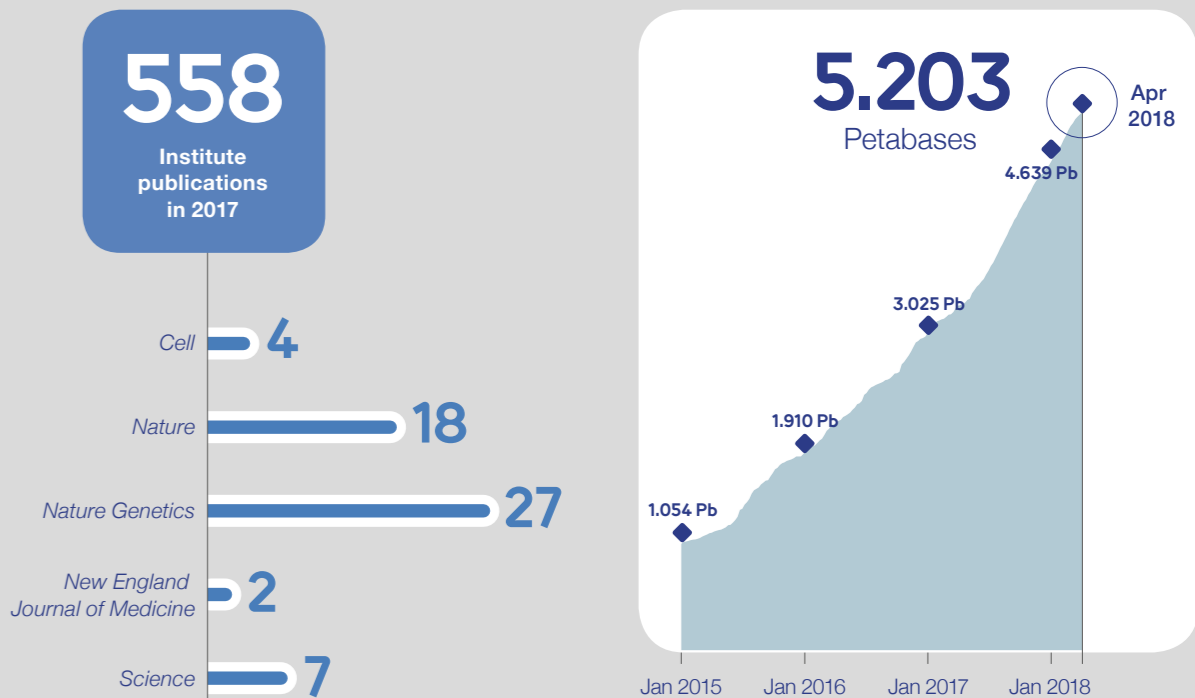
# At a Glance



# 2017 timeline



# Year in numbers



\* In December 2017

# Our work

With secured funding from Wellcome, we are able to strategically focus our work in five key research fields



## 10 Cancer, Ageing and Somatic Mutation

Provides leadership in data aggregation and informatics innovation, develops high-throughput cellular models of cancer for genome-wide functional screens and drug testing, and explores somatic mutation's role in clonal evolution, ageing and development.



## 18 Cellular Genetics

Explores human gene function by studying the impact of genome variation on cell biology. Large-scale systematic screens are used to discover the impact of naturally-occurring and engineered genome mutations in human iPS cells, their differentiated derivatives, and other cell types.



## 22 Human Genetics

Applies genomics to population-scale studies to identify the causal variants and pathways involved in human disease and their effects on cell biology. It also models developmental disorders to explore which physical aspects might be reversible.



## 28 Infection Genomics

Investigates the common underpinning mechanisms of evolution, infection and resistance to therapy in bacteria and parasites. It also explores the genetics of host response to infection and the role of the microbiota in health and disease.



## 34 Malaria

Integrates genomic, genetic and proteomic approaches to develop and enhance high-throughput tools and technologies to study specific biological problems relevant for malaria control and to understand the fundamental science of the human host, the mosquito vector and the *Plasmodium* pathogen.



What we do

Our work

Our approach

Other information



# Cancer, Ageing and Somatic Mutation



**16%**  
of patients had mutations affecting the PI3K signalling pathway

## 1 Bone cancer drug leads

**Applying two of the Sanger Institute's key strengths – long-term collaboration and whole-genome analysis – to rare bone cancers has revealed how existing drugs may offer new treatment options.**

Chordoma and osteosarcoma have been the latest two cancer types to benefit from Sanger scientists' long-standing relationship with University College London Cancer Institute and the Royal National Orthopaedic Hospital NHS Trust. In two studies published in *Nature Communications*, the researchers demonstrated how combining clinical samples and knowledge with whole-genome sequencing and analysis delivers insights into the molecular biology of bone cancer that could inform future treatments.<sup>1,2</sup>

In 2013, this approach uncovered cancer-driver mutations shared by almost all forms of chondroblastoma and giant cell bone tumours.<sup>3</sup> In 2017, applying the same method to more than 100 chordomas – rare cancers of the skull and spine – revealed that 16 per cent of patients had mutations affecting the PI3K signalling pathway.<sup>1</sup> PI3K inhibitors are already being used to treat a range of cancers, and represent a possible new option for chordoma patients carrying PI3K pathway mutations.

Approximately one quarter of cases had an additional copy of the *brachyury* gene, previously implicated in inherited forms of chordoma. In addition, some 10 per cent of cases were linked to mutation of a novel cancer gene, *LYST*. In the longer term, these genes could be valuable leads for new drug development.

Similarly, the largest whole-genome sequencing study yet undertaken in osteosarcoma – the most common form of bone cancer in children and young adults – could also deliver rapid patient benefits. Of the 112 sequenced tumours investigated, 7 per cent had mutations in genes involved in insulin-like growth factor signalling, which is likely to play a key role in the control of bone growth.<sup>2</sup>

Crucially, drugs targeting this pathway, IGFR1 inhibitors, already exist. Although trials of IGFR1 inhibitors in osteosarcoma have shown limited benefits, further studies concentrating on patients with mutations affecting insulin-like growth factor signalling could enhance the chances of success, paving the way for personalised treatments.



### References

1. Tarpey PS *et al.* The driver landscape of sporadic chordoma. *Nat Commun.* 2017; **8**: 890.
2. Behjati S *et al.* Recurrent mutation of IGF signalling genes and distinct patterns of genomic rearrangement in osteosarcoma. *Nat Commun.* 2017; **8**: 15936.
3. Behjati S *et al.* Distinct *H3F3A* and *H3F3B* driver mutations define chondroblastoma and giant cell tumor of bone. *Nat Genet.* 2013; **45**: 1479–82.

## 2 How cancer evolves

Innovative methods adapted from evolutionary biology have provided a new way of viewing cancer development. The findings have wide-ranging implications for our understanding of cell ageing, the hunt for cancer driver mutations, and precision oncology.

Cancers can be seen as models of evolution in action, with cancer cells evolving as they accumulate genetic changes that provide a selective growth advantage. Now that abundant cancer genome sequence data are available, Sanger scientists have seized the opportunity to study populations of cancer cells with the same approaches developed to explore the evolution of populations of organisms.

In a landmark study published in *Cell*, Sanger researchers interrogated data from more than 7,500 cancers across 29 cancer types.<sup>1</sup> Remarkably, species and cancers evolve in diametrically opposed ways. Species evolution is typically characterised by negative selection – the loss of mutations because they lower fitness. Yet this is hardly ever seen in cancer. Instead, cancer development is dominated by positive selection – preserving the handful of driver mutations that give a cell a competitive growth advantage.

The results have profound implications. Since somatic cells have such a remarkable tolerance of mutations, the findings highlight the important role of individual mutations in cellular and organismal ageing.

The team discovered that out of the many accumulated mutations acquired over a person's lifetime just one to ten key alterations in a cell are required to cause cancer, depending on the tumour type. The results also reveal that approximately half of positively selected driver mutations are not in genes previously implicated in cancer, suggesting that many cancer-driving changes remain to be discovered.

In addition, these methods provide a way to assess whether or not specific mutations are truly driving cancer – an essential step in generating the robust knowledge base that will be needed to apply genomics to clinical decision making and deliver an era of precision oncology.



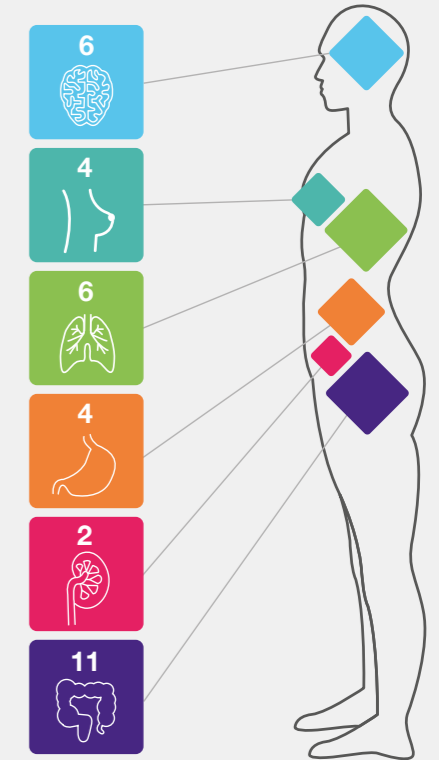
### Reference

1. Martincorena I *et al.* Universal patterns of selection in cancer and somatic tissues. *Cell.* 2017; **171**: 1029–1041.e21.

Data from more than 7,500 cancers from 29 tumour types were analysed



### Number of mutations needed to cause cancer

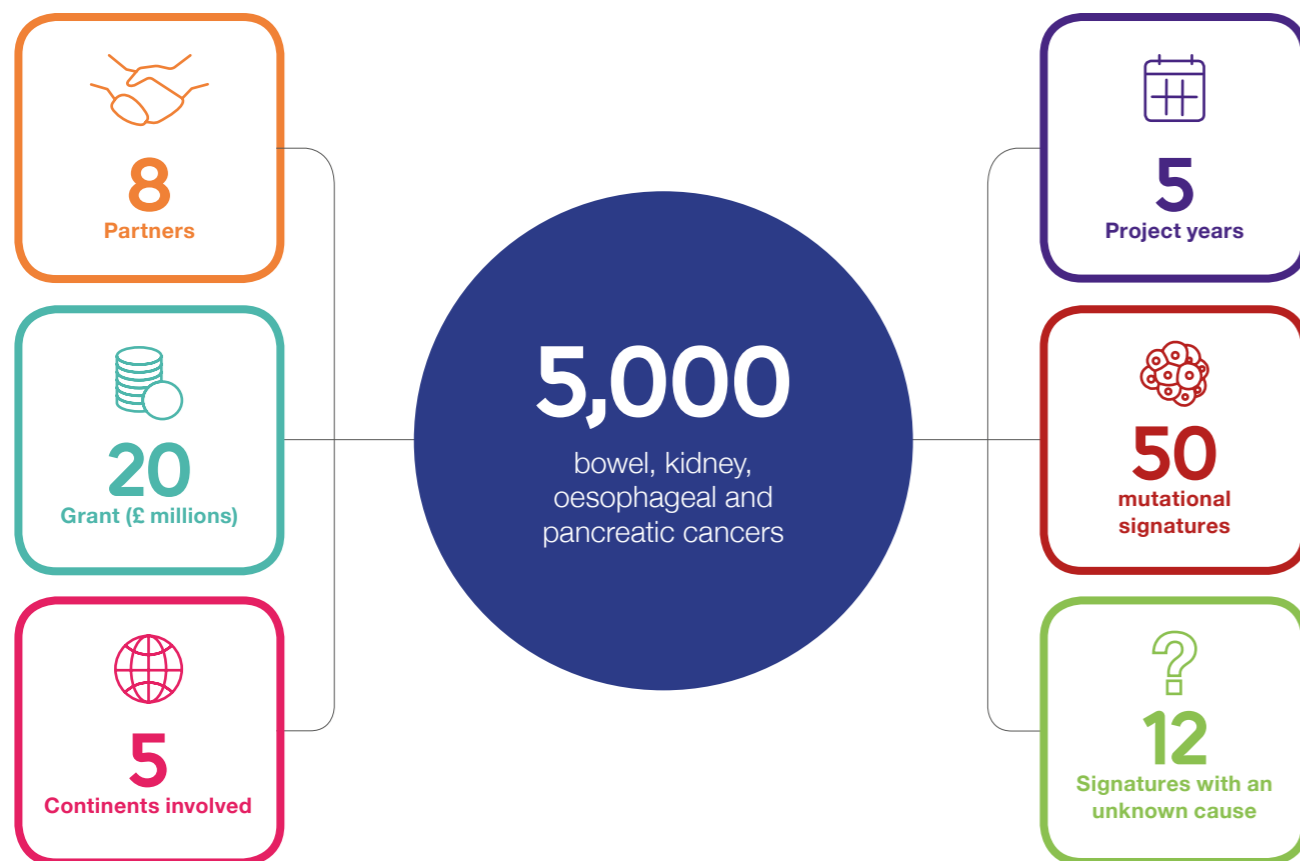


“Across cancer types a relatively consistent small number of mutated genes is required to convert a single normal cell into a cancer cell, but the specific genes differ according to cancer type. This increasingly precise understanding provides the foundation for the discovery and use of targeted therapies.”

**Professor Sir Mike Stratton**  
An author of the study and Director of the Wellcome Sanger Institute

### In this section

- 1 Bone cancer drugs leads
- 2 How cancer evolves
- 3 Sanger takes on mutational signatures Grand Challenge
- 4 First mutations in human life discovered
- 5 Key genes in cancer suppression found
- 6 Mutational signatures pinpoint best breast cancer treatments
- 7 Key role of epitranscriptomics in leukaemia
- 8 Knowledge bank for precision oncology
- 9 Antique childhood tumours give new understanding
- 10 How organoids reveal cancer's true diversity



### 3 Sanger takes on mutational signatures Grand Challenge

Sanger researchers are leading an international quest to identify the causes of cancer-driving mutations, funded by a £20 million Grand Challenge award from Cancer Research UK.

It has recently become clear that cancer cell genomes contain distinct patterns of DNA damage – mutational signatures – that give important clues to the origins of cancer. DNA-damaging and cancer-causing insults such as tobacco smoke and ultraviolet radiation all leave distinctive scars in the genome.

Cancer genome sequencing and innovative computational tools have revealed dozens of mutational signatures.<sup>1</sup> Some have been linked to external exposures or to mechanistic processes, such as the breakdown of mechanisms that repair DNA and maintain the integrity of the genome. However, the underlying causes of many mutational signatures remain a mystery.

In the Grand Challenge initiative – one of just four supported in the first round of funding – an international team will study 5,000 pancreatic, kidney, oesophageal and bowel cancers from five continents. The study will generate exceptionally high-resolution data on the mutational processes operating in these cancers, of potential value in diagnosis and predicting prognosis. Furthermore, data on exposures could reveal the key factors triggering DNA damage – opening up new opportunities for cancer prevention.



**Reference**  
1. Alexandrov LB *et al.* Signatures of mutational processes in human cancer. *Nature*. 2013; **500**: 415–21.



### 4 First mutations in human life discovered

In *Nature*, Sanger researchers reported how they traced, in adult cells, mutations that arose when an embryo was just hours old.

As cells divide, mutations inevitably arise, albeit very rarely. As adults we are, therefore, genetic mosaics, our cells carrying different combinations of mutations depending on their evolutionary journey from a single fertilised egg. Moreover, a mutation that arose early in life will be carried by many cells in the body.

Researchers in the Cancer, Ageing and Somatic Mutation programme have exploited this fact to develop statistical tools that can 'date' when mutations arose. They generated whole-genome sequences using blood samples from 279 individuals, and identified 163 mutations that arose very early in development – remarkably, as far

back as the two-, four- and eight-cell stage of development.<sup>1</sup>

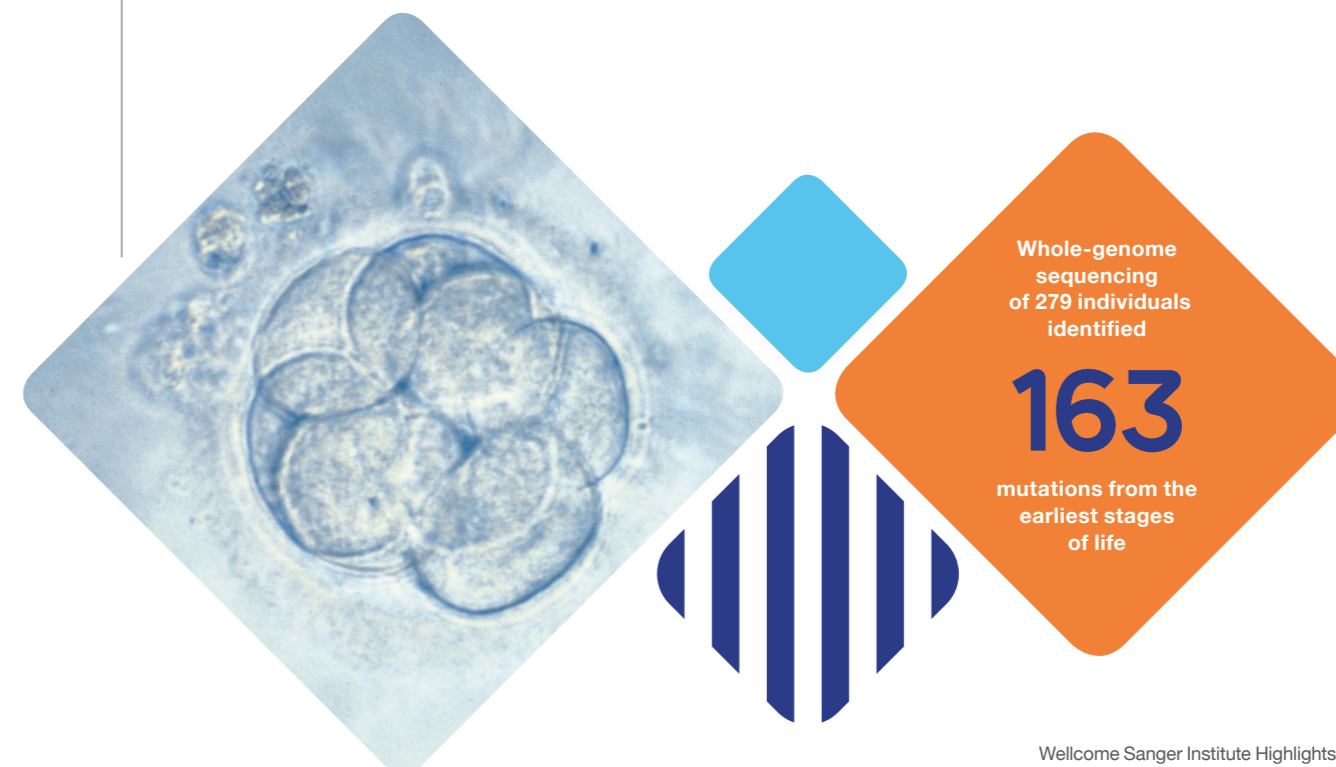
Moreover, these mutations could be used as markers to track the descendants of these primordial cells. Strikingly, this revealed that the two cells in a two-cell embryo are not equivalent – one typically gives rise to 70 per cent of adult body tissues. This skewing was also seen at later rounds of cell division.

The work also enabled the team to determine that, on average, three mutations occur at each round of division, more than previously thought. In addition, by applying their understanding of mutational

signatures to characterise mechanisms of DNA damage, the researchers found that two particular mutational processes predominate: those responsible for signatures 1 and 5.<sup>2</sup> Ultimately, the techniques will enable studies of adult cells to provide new insights into embryonic development.



**References**  
1. Ju YS *et al.* Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature*. 2017; **543**: 714–718.  
2. Alexandrov LB *et al.* Signatures of mutational processes in human cancer. *Nature*. 2013; **500**: 415–21.





5

## Key genes in cancer suppression found

Large-scale studies in mice have identified genes that slow the emergence of cancer and prevent its deadly spread.

Mouse models are an important complement to human studies for identifying genes involved in cancer – either driving tumour formation or protecting against it – and for exploring their function. In particular, genetic manipulation enables studies to be carried out in mice that would be impossible in humans.

This approach has been used to conduct a large-scale screen for genes that inhibit the development of cancer – so-called tumour suppressor genes – which are frequently lost in cancer cells.

It is increasingly clear that genes involved in cancer do not act in isolation; their impact often depends on other mutated genes present in a cell. For example, mutations in the *PTEN* tumour suppressor – the second most commonly mutated gene in human cancers – typically act in combination with other mutated genes.

To identify other tumour suppressors that cooperate with *Pten*, Sanger researchers developed an innovative transposable

element which, when mobilised, disrupts *Pten* as well as any gene into which it inserts. An analysis of 278 prostate, breast and skin cancers that developed after activation of the transposable element, published in *Nature Genetics*, revealed hundreds of mutations that cooperated with *Pten*.<sup>1</sup>

Focusing on five of the most promising leads, the team showed that disrupting these genes in human cell lines drove cancerous changes. The work has therefore identified a wealth of new tumour suppressors acting in concert with *Pten* to prevent cancer developing, opening up new avenues of therapy development.

Studies in mice have also generated important new insights into metastasis, the spread of cancer around the body. Metastasis is driven by genetic changes in cancer cells, but also depends on how receptive body tissues are to colonisation by cancer cells. By systematically assessing 800 mouse strains lacking specific genes,

Sanger researchers identified 23 genes whose loss influenced colonisation, either accelerating or inhibiting metastasis.<sup>2</sup>

The biggest impact was linked to loss of the *Spns2* gene, previously implicated in immune cell function but not cancer biology. The results, published in *Nature*, highlight the exciting possibility of targeting *Spns2* to influence immune system function and limit the metastatic spread of cancer cells.



### References

- de la Rosa J *et al.* A single-copy *Sleeping Beauty* transposon mutagenesis screen identifies new *PTEN*-cooperating tumor suppressor genes. *Nat Genet.* 2017; **49**: 730–741.
- van der Weyden L *et al.* Genome-wide *in vivo* screen identifies novel host regulators of metastatic colonization. *Nature.* 2017; **541**: 233–236.

Studying cancer in mouse models has allowed Sanger scientists to discover five genes that help to suppress tumour development

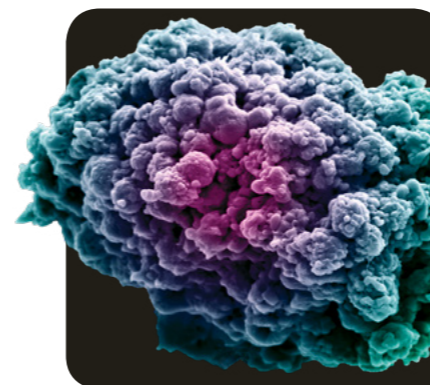
23

genes identified by Sanger researchers whose loss influenced colonisation, either accelerating or inhibiting metastasis



6

## Mutational signatures pinpoint best breast cancer treatments



Sanger researchers have gained new insights into the mutational processes underlying breast cancer – knowledge that could be used to identify patients likely to benefit from particular anti-cancer drugs.

Sanger research has revealed that mutagenic agents and processes can damage genomes in distinctive ways, creating ‘mutational signatures’ that may shed light on the causes of DNA damage and the origins of cancer. Furthermore, drugs developed for one type of cancer could potentially be active against other types sharing the same mutational signature.

In breast cancer, for example, so-called PARP inhibitors have been developed to treat patients with cancers caused by *BRCA1* and *BRCA2* mutations, which disrupt DNA repair and are associated with a specific mutational signature. While *BRCA1* and *BRCA2* mutations are associated with 1–5 per cent of breast cancer cases, a Sanger-led study published in *Nature Medicine* found that up to 20 per cent of tumours show a *BRCA*-like mutational signature.<sup>1</sup> PARP inhibitors could therefore potentially be of benefit to many more women.

Similarly, a further study of 640 breast cancer genomes found that 11 had a mutational signature associated with an abnormality in another DNA repair mechanism, mismatch repair, not previously seen before in breast cancer.<sup>2</sup> A class of drugs known as checkpoint inhibitors has been used to treat colorectal cancers with defective mismatch repair, and could potentially provide another option for women whose breast cancers show the same patterns of DNA damage.

In addition, using mutational signatures to identify damage to mismatch repair genes was more accurate than using genetic and epigenetic sequencing alone. Standard genetic analysis identified only six out of the 11 people affected, in contrast mutational signatures highlighted all 11. This suggests that patients could be stratified into different categories for treatment based on their mutational signatures.



### References

- Davies H *et al.* HRDetect is a predictor of *BRCA1* and *BRCA2* deficiency based on mutational signatures. *Nat Med.* 2017; **23**: 517–525.
- Davies H *et al.* Whole-genome sequencing reveals breast cancers with mismatch repair deficiency. *Cancer Res.* 2017; **77**: 4755–4762.

7

## Key role of epitranscriptomics in leukaemia

Gene editing screens have identified an RNA-modifying enzyme that plays a critical role in acute myeloid leukaemia (AML), and could be an important target for new therapeutics.

Enzymes that modify RNA – for example by adding methyl groups to RNA bases – are turning out to be important regulators of gene activity. To explore their contribution to cancer, Sanger researchers and their colleagues used CRISPR-Cas9 gene editing in systematic screens for genes necessary for the growth of mouse leukaemia cells. This screening, published in *Nature*, revealed 46 genes coding for RNA-modifying enzymes critical to leukaemia cell growth.<sup>1</sup>

One of the strongest effects was seen after knockout of *METTL3*, which methylates A residues in RNA transcripts. The team went on to show that loss of *METTL3* slowed the proliferation of cultured mouse and human leukaemia cells, and made cells less likely to seed new leukaemias when injected into mice. Crucially, loss of *METTL3* had no detrimental impact on normal cells.

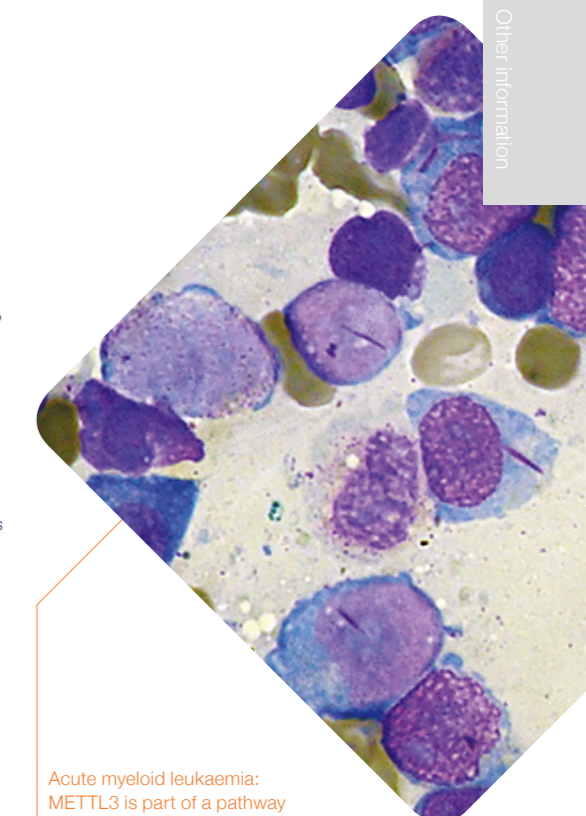
*METTL3* was found to bind to the promoter region of a suite of genes, in the presence of a protein known as CEBPZ. At these genes, *METTL3* methylated the protein-coding section of RNA transcripts, leading to more efficient translation and increased synthesis of the corresponding proteins. Among these proteins were several known to drive proliferation of AML cells.

The study identified *METTL3* as part of a pathway critical to the development of AML, making the protein an exciting new target for a cancer that kills two out of every three people affected.



### Reference

- Barbieri I *et al.* Promoter-bound *METTL3* maintains myeloid leukaemia by m<sup>6</sup>A-dependent translation control. *Nature.* 2017; **552**: 126–131.



Acute myeloid leukaemia: *METTL3* is part of a pathway that plays an important role in AML development, making the protein an exciting drug target

What we do

Our work

Our approach

Other information





8

## Knowledge bank for precision oncology

In *Nature Genetics*, Sanger researchers and clinical colleagues described how they developed a prototype 'knowledge bank' to show how genomic information could guide clinical decision-making in acute myeloid leukaemia.<sup>1</sup> The database incorporated genomic and clinical data from more than 1,500 patients, with statistical tools providing insight into likely prognosis and preferred treatment options for individual patients. Building this knowledge bank was an important milestone on the route to precision medicine in oncology, and has pinpointed the challenges that need to be overcome to deliver the vision in medical practice.



### Reference

1. Gerstung M *et al.* Precision oncology for acute myeloid leukemia using a knowledge bank approach. *Nat Genet.* 2017; **49**: 332–340.

More than  
**1,500**  
patients incorporated  
onto the database

9

Some cancers  
only appear  
3 or 4 times in  
**50**  
years

## Antique childhood tumours give new understanding

Until now, some cancers have been so rare that only limited insight into their genetic basis could be obtained. But now a joint project between Sanger researchers and pathologists at Great Ormond Street Hospital for Children is opening up a century's worth of hospital sample collections for genomic scrutiny.<sup>1</sup> Technological advances now enable sequence information to be obtained from tissue samples fixed with formalin and embedded in paraffin wax. In *Lancet Oncology* the scientists detail likely driver mutations in three rare childhood cancers dating back to the 1920s. The team hopes to apply the same approach to more of the hospital's archive to explore other exceptionally rare cancers.



### Reference

1. Virasami A *et al.* Molecular diagnoses of century-old childhood tumours. *Lancet Oncol.* 2017; **18**: e237.



## How organoids reveal cancer's true diversity

All tumours are genetically different, even if they occur in the same organ. And **each tumour contains genetically different tribes of cancerous cells** competing for dominance.<sup>1</sup> Now Sanger cancer researchers, working closely with researchers at the Hubrecht Institute, have shown that every cell within every tumour is genetically different to each other.

Writing in *Nature*, the team describes the first study to **combine cutting-edge organoid techniques with single-cell approaches to interrogate individual cells** within the same colorectal cancer tumour.<sup>2</sup> Their work, which exploits an innovative methodology to overcome many of the problems inherent in current

single-cell genomic techniques, has the potential to be applied at scale to investigate the evolution of numerous tumour types.

The researchers isolated individual cells from cancerous and healthy portions of a person's colon, and then grew them into organoids – 3D clusters of cells that mimic the gut – to stably amplify the numbers of cells for study. In this way the team was able to interrogate the genetic, epigenetic, transcriptomic and functional differences between neighbouring healthy and tumour cells from the same tissue.

They found that the tumour cells had many more mutations than healthy cells and that each cell in a tumour was

genetically different to each other. In addition, the mutational processes at work in the cancer cells were markedly distinct from those seen in normal cells. These cancer-specific processes offer opportunities to further understand how cancers develop and provide novel targets for therapeutics.



### References

1. Fischer A *et al.* High-definition reconstruction of clonal composition in cancer. *Cell Reports* 2014; **7**: 1740–1752.
2. Roerink S *et al.* Intra-tumour diversification in colorectal cancer at the single cell level. *Nature.* 2018; Apr 11. doi: 10.1038/s41586-018-0024-3.

Gut organoids for single-cell insights – using mini 3D clusters of cells can overcome the problems with current single-cell techniques



What we do  
Our work  
Our approach  
Other information



# Cellular Genetics



## 1 One million cells down...

Just one year after its launch, the Human Cell Atlas, an ambitious international partnership to map the human body cell by cell has sequenced, analysed and characterised more than one million cells.

“The data will provide an entry point for deeper study of cells’ functions and interactions, both within their home tissues and throughout the body. Such knowledge will, over time, have a transformative effect on the understanding of human biology and health.”

**Dr Sarah Teichmann**  
Head of Cellular Genetics at the Sanger Institute and co-Chair of the Human Cell Atlas Organising Committee

Innovative technologies for analysing individual cells are opening up new opportunities to define the myriad cell types of the human body, and to explore how their functions relate to the genes active within them. An atlas mapping every cell type in the body would be a tremendous boon for biomedical research, and is the goal of a global partnership led by an organising committee, co-chaired by Sanger and Broad Institute researchers, encompassing 10 countries. Writing in *Nature*, the Human Cell Atlas team has set out the goals and challenges facing this potentially transformative initiative.<sup>1</sup>

For phase one, the collaboration seeks to collect and study between 30-100 million individual cells from select organs and tissues. The researchers use massively-parallel single-cell RNA sequencing (a suite of genomic techniques capable of identifying gene expression profiles in thousands of individual cells at a time), related technologies to characterise other molecules, and spatial methods to map cells’ locations and interactions. The results will give a firm foundation of understanding of how different cells work with each other both in their home tissues and throughout the body.

Employing these techniques, the Human Cell Atlas has investigated more than one million cells collected from bone marrow and cord blood from healthy human donors. The single-cell RNA expression data of these immune cells form the foundation of an openly available resource that will allow the global scientific community to further their research. Over time, as the initiative’s data are made freely available, an open and globally accessible reference map of the human body will be built, in much the same way that the Human Genome Project unlocked the human genome.

**Reference**  
1. Rozenblatt-Rosen O, Stubbington MJT, Regev A, Teichmann SA. The Human Cell Atlas: from vision to reality. *Nature*. 2017; **550**: 451–453.



### In this section

- 1 One million cells down...
- 2 High-speed, high-volume cells on demand
- 3 Growing replacement organs
- 4 New, ultra-flexible stem cells recreate earliest steps of life
- 5 HipSci delivers new national stem cell resource
- 6 Drug resistance: some mutations need a little genetic help

## 2 High-speed, high-volume cells on demand

An innovative new method is enabling researchers to generate large numbers of precisely defined cell types in days instead of weeks.

To explore the properties of specialised cells, researchers are increasingly culturing and driving the differentiation of human pluripotent stem cells (hPSCs), embryonic stem cells and induced pluripotent stem cells. However, differentiation of cultured hPSCs is typically slow and inefficient.

Now, Sanger scientists have developed a new method that is both quicker and generates larger numbers of differentiated cells.<sup>1</sup> The technique is based on the introduction of two gene cassettes into ‘genomic safe harbours’ – sites where integration of a new genetic element will have no adverse impact on the cell.

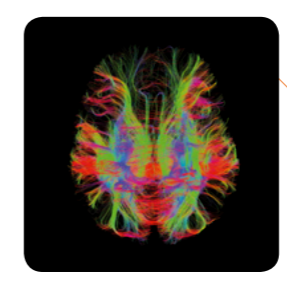
The first cassette is permanently active and produces a transcription factor that is activated only in the presence of an antibiotic, doxycycline; the second cassette contains transgenes required to reprogramme the hPSC into a specialised

adult cell, which are switched on by the doxycycline-dependent transcription factor. Addition of doxycycline therefore drives rapid and efficient cellular reprogramming.

Using this method, known as OPTi-OX, the Sanger team successfully generated large numbers of neurons and skeletal muscle cells within days, and developed a protocol for production of human oligodendrocytes. The platform could theoretically be adapted to generate many other cell types – potentially including novel cell types identified by the Human Cell Atlas project.

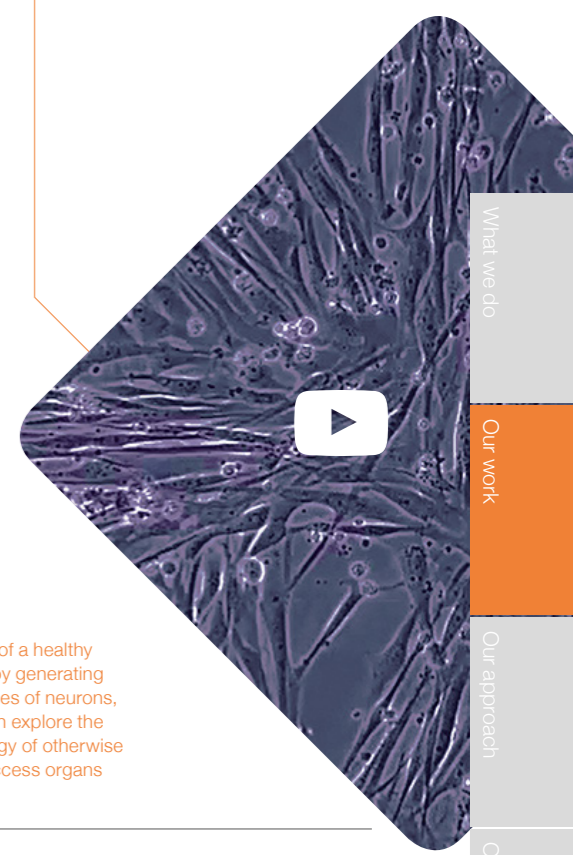


**Reference**  
1. Pawlowski M *et al.* Inducible and deterministic forward programming of human pluripotent stem cells into neurons, skeletal myocytes, and oligodendrocytes. *Stem Cell Reports*. 2017; **8**: 803–812.



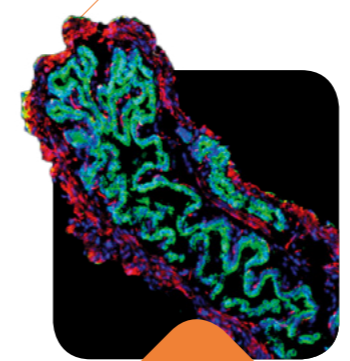
Nerve fibres of a healthy adult brain: by generating large quantities of neurons, scientists can explore the cellular biology of otherwise difficult-to-access organs

OPTi-OX is a new method that can generate neuron and skeletal cells in days instead of weeks



## 3

Bile duct cells: organoid technology has been successfully used to grow new bile ducts for transplant surgery in mice



Bile duct disorders are responsible for **70%** of paediatric liver transplants

## Growing replacement organs

A multidisciplinary team including Sanger researchers has used tissue engineering to generate artificial bile ducts from cultured cells, and shown that they function effectively in mice.

Bile ducts carry bile from the liver to the intestine. Bile duct disorders are currently difficult to treat, and are responsible for 70 per cent of paediatric liver transplants.

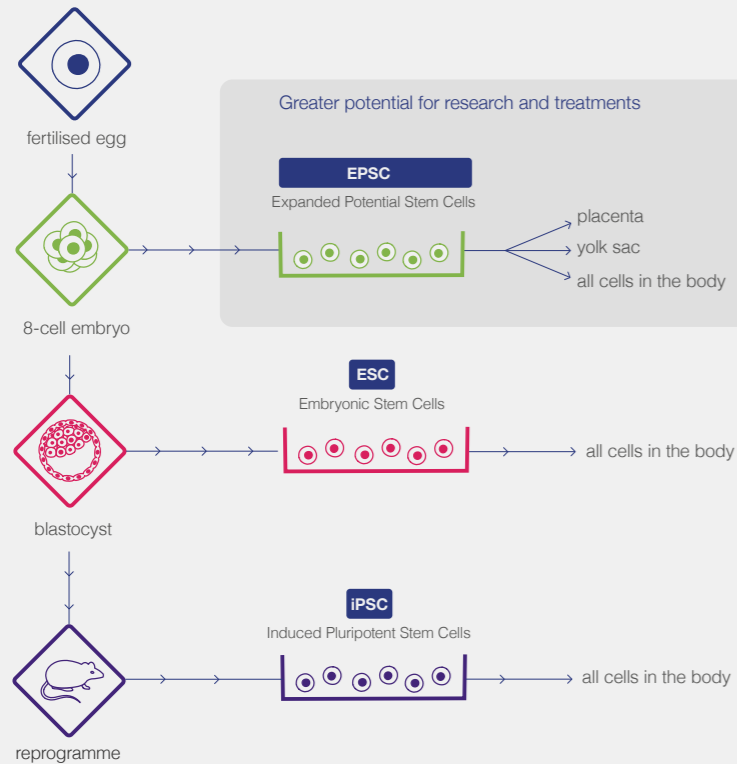
Artificial bile ducts created by tissue engineering could offer an alternative approach to treatment. In pursuit of this objective, a Cambridge-based team has been able to extract human cholangiocytes – the cells that line the walls of bile ducts – and grow them up in large numbers in cultures, maintaining their functional properties.<sup>1</sup> The cells spontaneously formed duct-like structures – organoids – and were also able to populate biodegradable scaffolds. As reported in *Nature Medicine*, these structures could rescue a mouse model of bile duct injury.

The isolation of cholangiocytes from accessible sites in the body, rapid production of large numbers of functioning cells in culture, and successful seeding of scaffolds are significant steps towards the development of replacement tissue for use in regenerative medicine.



**Reference**  
1. Sampaziotis F *et al.* Reconstruction of the mouse extrahepatic biliary tree using primary human extrahepatic cholangiocyte organoids. *Nat Med*. 2017; **23**: 954–963.

### The different types of stem cell



## 4 New, ultra-flexible stem cells recreate earliest steps of life

Sanger researchers have created primordial 'expanded potential stem cells' that can generate the placenta and yolk sac, as well as the embryo itself.

Embryonic stem cells and induced pluripotent stem (iPS) cells are highly flexible, able to generate all the different cell types of an adult organism. However, to date, they have not been able to create the extra-embryonic tissues (such as the placenta and yolk sac) that support an embryo's development. Sanger scientists have developed a technique that completes the full picture and will allow experimental genomic exploration of miscarriage and developmental disorders.

The team inhibited the expression of proteins thought to drive differentiation of extra-embryonic cell lineages (mitogen-activated protein kinases, Src and Wnt/Hippo/TNKS1/2), to extract and propagate cells from even earlier in development – eight-cell mouse embryos.<sup>1</sup> As well as embryonic tissues, these 'expanded

potential stem cells' (EPSCs) can be induced to generate the other two types of blastocyst stem cells that produce the placenta and yolk sac. Writing in *Nature*, the team has also shown that later-stage embryonic stem cells and iPS cells can be converted into EPSCs.

EPSCs promise to be a valuable tool for understanding the earliest stages of embryonic development and placenta formation. Furthermore, the methods could also be applied to generate EPSCs for mammalian species where attempts to generate embryonic stem cells and iPS cells have not yet succeeded.



**Reference**  
1. Yang J *et al.* Establishment of mouse expanded potential stem cells. *Nature*. 2017; **550**: 393–397.

## 5 HipSci delivers new national stem cell resource

A national consortium including Sanger researchers has unveiled a resource of comprehensively characterised induced pluripotent (iPS) stem cell lines. The Human Induced Pluripotent Stem Cell Initiative (HipSci) has generated and characterised 711 cell lines from 301 healthy donors, publishing their initial analyses in *Nature*.<sup>1</sup> The work has revealed that common genetic variation has a significant impact on the properties of iPS cell lines, and has shed light on gene regulatory regions and the impact of DNA rearrangements commonly seen in such cells. The detailed characterisation provides crucial information for those using the cell lines in their research.

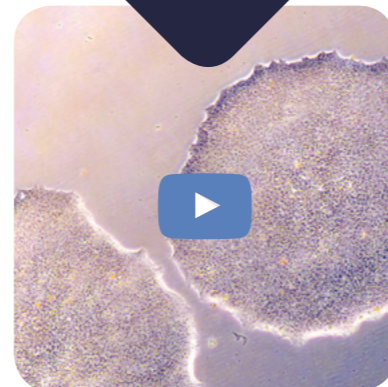


**Reference**  
1. Kilpinen H *et al.* Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature*. 2017; **546**: 370–375.

Human Induced Pluripotent Stem Cell Initiative (HipSci) has generated and characterised

711

cell lines from 301 healthy donors



## Drug resistance: some mutations need a little genetic help

High genetic diversity in bacterial and cancer cells results in poorer outcomes for people treated with antibiotics or chemotherapy drugs. A massive study in yeast conducted by Sanger scientists has discovered why.

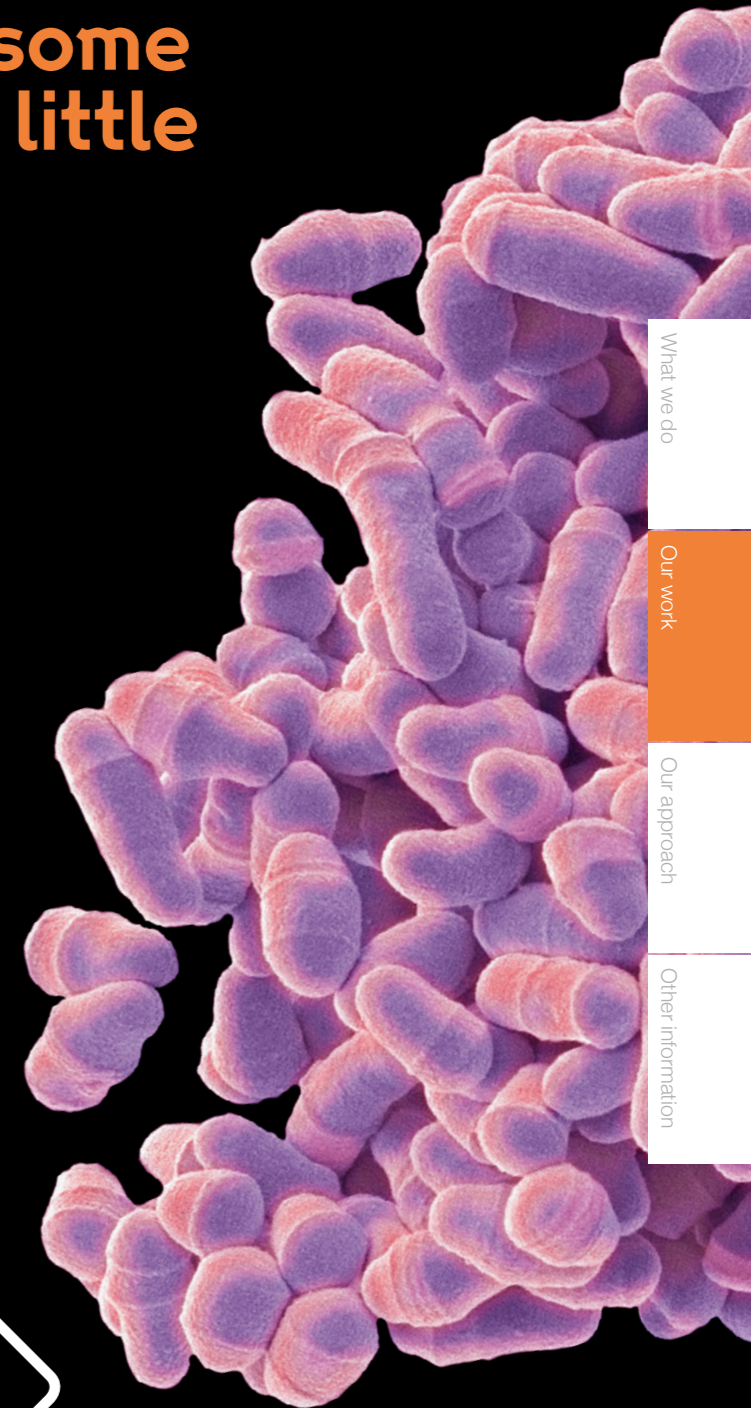
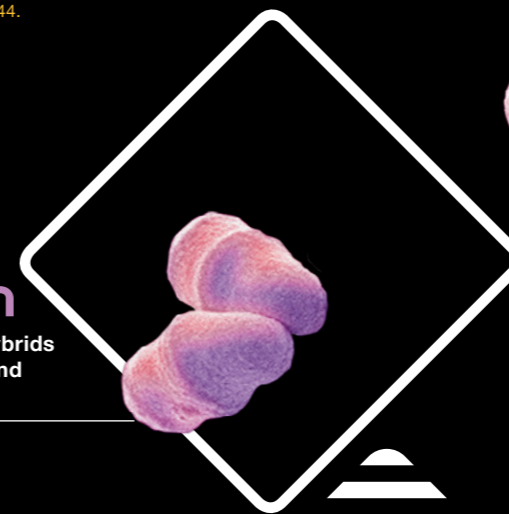
The team cross-bred two distantly related species of yeast, from Africa and the Americas, to create more than 10 million genetically distinct hybrids and followed their evolution in response to chemotherapeutic drugs<sup>1</sup>. This approach revealed which mutations drive resistance, and how they are affected by other mutations already present in the genome.

The results pointed to two types of driver mutations: strong drivers whose presence always produce resistance, and weaker ones that require the assistance of pre-existing background mutations to be effective. Further study will help to identify which genetic profiles are more conducive to resistance than others.



**Reference**  
1. Vázquez-García I *et al.* Clonal heterogeneity influences the fate of new adaptive mutations. *Cell Rep*. 2017; **21**: 732–744.

10 million genetically distinct hybrids of yeast from Africa and the Americas



What we do

Our work

Our approach

Other information



# Human Genetics

## In this section

- 1 Sequencing reveals new causes of developmental disorders
- 2 Pinpointing the causes of IBD
- 3 Baby's DNA influences mother's pre-eclampsia risk
- 4 How losing a gene could help your heart
- 5 Why HbA1c diabetes test may fail African Americans
- 6 Greeks bearing (welcome) genetic gifts
- 7 DNA analyses reveal secrets of human history
- 8 UK Biobank helps unlock the biology of osteoarthritis

# 1 Sequencing reveals new causes of developmental disorders

## A Sanger-led study of children with undiagnosed developmental disorders has provided hundreds of families with a firm diagnosis.

In the Deciphering Developmental Disorders (DDD) initiative, Sanger researchers and clinical geneticists from across the British Isles are using genome sequencing to identify possible causes of unexplained developmental disorders, and to determine how genome-based approaches could be integrated into routine care.

Focusing on *de novo* mutations – those not present in either parent but arising spontaneously during cell division after conception – the DDD team sequenced the coding region of the genomes of more than 4,000 affected families (children and their parents). The results, published in *Nature*, estimated new mutations in 42 per cent of children studied.<sup>1</sup> Around half of mutations were predicted to lead to complete loss of gene function and half to altered function. Almost one quarter of cases could be assigned to existing syndromes, while *de novo* mutations were identified in 14 genes that had not been previously linked to developmental disorders.

The results provide unprecedented insight into the spectrum of developmental disorders across the UK. They suggest that one in every 300 children born in the UK has a rare developmental disorder caused by a new mutation – approximately 2,000 children a year. Collectively, *de novo* mutations account for more disorders than the common chromosomal disorders caused by the duplication of either chromosome 13, 18 or 21. The researchers estimate that 400,000 babies are likely to be born with a developmental disorder caused by a *de novo* mutation every year across the globe.

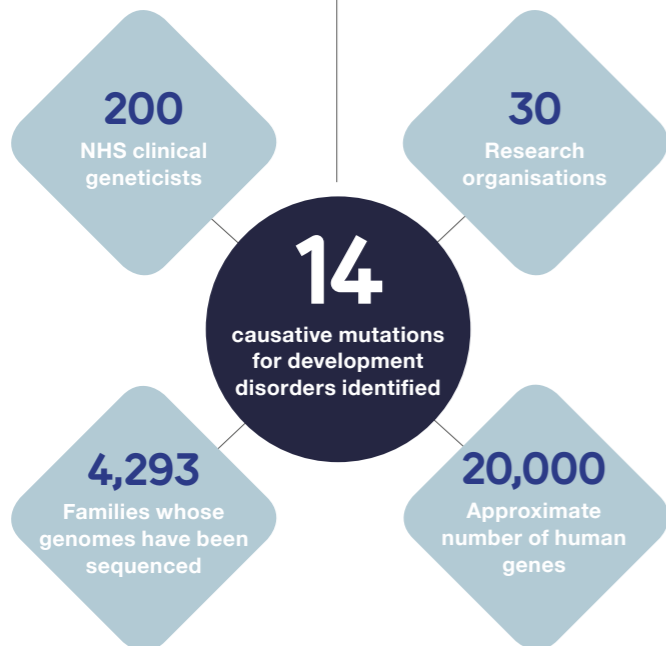
In a follow-on study, also published in *Nature*, the DDD team explored whether *de novo* mutations in regulatory elements controlling gene expression in the foetal brain could account for cases where no coding sequence mutations were identified. Such mutations would not affect the function of a protein but might disrupt carefully regulated developmental programmes by affecting when, where, or how much of a protein is produced. Sequencing of known control regions in nearly 8,000 cases – by far the largest such study ever undertaken – revealed *de novo* mutations in 1–3 per cent of patients with unexplained neurodevelopmental disorders.

The study proved the hypothesis that noncoding regions of the genome are able to contribute to neurodevelopmental disorders, but also offered cause for optimism: such DNA variations only rarely have such an effect. In fact, the researchers found that less than 1 per cent of noncoding changes were involved in neurodevelopmental disorders.<sup>2</sup>



### References

1. Deciphering Developmental Disorders Study. Prevalence and architecture of *de novo* mutations in developmental disorders. *Nature*. 2017; **542**: 433–438.
2. Short PJ *et al.* *De novo* mutations in regulatory elements in neurodevelopmental disorders. *Nature*. 2018; **555**: 611–619.



immune system function. Notably, integrins are part of pathways already being targeted therapeutically, suggesting that new associations at common variants can provide therapeutically important leads.

Finally, Sanger-led research published in *Nature* has made important progress in the fine-mapping of susceptibility loci to identify the precise genetic changes underlying increased risk of disease.<sup>3</sup> Using high-density genotyping tools specific for loci implicated in immune-related disease, the team was able to nail down 18 specific genetic changes with a high degree of confidence and a further 27 with a reasonable degree of certainty (>50 per cent). The results provide 45 specific leads for exploring mechanisms of disease and identifying possible therapeutic targets.



### References

1. Luo Y *et al.* Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at *ADCY7*. *Nat Genet*. 2017; **49**: 186–192.
2. de Lange KM *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet*. 2017; **49**: 256–261.
3. Huang H *et al.* Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature*. 2017; **547**: 173–178.

## 2 Pinpointing the causes of IBD

Sanger-led analyses are generating a short list of genetic prime suspects that contribute to inflammatory bowel disease.

Inflammatory bowel disease (IBD) is an exemplar of how genetic analyses can provide insight into the causes and mechanisms of a complex disease. So far, more than 200 loci that affect the risk of IBD have been identified using genome-wide association studies (GWAS), and additional locations continue to be discovered. Three studies published in 2017 by Sanger scientists are providing the mathematical tools to narrow down this pool to the most valuable leads for therapeutic intervention.

One Sanger-led study published in *Nature Genetics* explored the potential of low-coverage whole-genome sequencing to identify low-frequency variants affecting IBD risk.<sup>1</sup> Sequencing of more than 4,000 cases revealed one significant finding – variation at the *ADCY7* gene that doubled the risk of ulcerative colitis – but suggested that this type of variation makes little contribution to Crohn's disease risk. Overall, the study demonstrated that future projects should involve larger-scale and deeper sequencing, to identify very rare variants of larger effect, and employing bigger GWAS were likely to elucidate greater insights.

In fact a GWAS and meta-analysis of nearly 60,000 subjects, also published in *Nature Genetics*, revealed 25 new IBD loci.<sup>2</sup> Among the most notable new discoveries were disease-linked variants in three integrin genes, which have wide-ranging roles in cell adhesion, cell signalling and

## 3 Baby's DNA influences mother's pre-eclampsia risk

Why does pre-eclampsia – raised blood pressure that can lead to complications during pregnancy – occur? Is there a genetic component? InterPregGen – an international study involving Sanger scientists, and geneticists, midwives and obstetricians from Finland, Iceland, Kazakhstan, Norway, the UK and Uzbekistan – has sought to find the answers.

Pregnancy is a two-way biochemical conversation between mother and baby. Yet genome-wide association studies (GWAS) of maternal genomes have failed to conclusively identify any maternal factors that increase the risk of pre-eclampsia.

For this study, published in *Nature Genetics*, the researchers examined the other side of the coin: how foetal growth impacts maternal biology. The team carried out the first GWAS of the genomes of more than 4,300 affected mothers' offspring to search for genetic variants that might have contributed to the development of pre-eclampsia.<sup>1</sup>

The scientists discovered that genetic variants close to the fetal *FLT1* gene increased the mother's risk of pre-eclampsia. The *FLT1* gene codes for a protein involved in placental growth and function, enhancing understanding of the mechanisms influencing pre-eclampsia risk.



### Reference

1. McGinnis R *et al.* Variants in the fetal genome near *FLT1* are associated with risk of preeclampsia. *Nat Genet*. 2017; **49**: 1255–1260.





Sanger researchers helped identify more than **1,300** genes whose function could be completely lost without any obvious impact on health



## 4 How losing a gene could help your heart

**Gene loss is not always harmful – sometimes it can even be beneficial.**

One way to understand how a gene works is to turn it off and observe what happens. Normally this is only possible in biological models, such as cell lines and organoids, or in model organisms like mouse or zebrafish. Such models are informative, but can never completely replicate the effect of gene loss in a fully functioning human body.

Sometimes, however, nature provides. For a variety of reasons, some people are born with both their copies of a gene not working. The results can be surprising. Although we have only around 20,000 genes, a study published in *Nature* shows that a surprisingly large number may be inessential.

Working with a population of 10,503 people in Pakistan, an international consortium that includes Sanger researchers identified more than 1,300 genes whose function could be completely lost without any obvious impact on health.<sup>1</sup> Most of the people affected had lost just one gene, but one individual had lost six.

Notably, because data had been collected on multiple aspects of participants' cardiometabolic health, the team was able to explore whether these natural gene knockouts had any impact on human physiology. At least seven of the disrupted genes showed an association with cardiometabolic traits.

One of the most striking discoveries was an association between loss of the *APOC3* gene and clearance rates of dietary fats from the bloodstream. Family members lacking *APOC3* cleared fats quicker than their relatives who had a functional *APOC3* gene, and are likely to have a reduced risk of developing cardiovascular disease. The findings suggest that targeting *APOC3* could be a therapeutic strategy.



**Reference**

1. Saleheen D *et al.* Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature*. 2017; **544**: 235–239.



**“We’re now entering a new era in genetics where we can systematically examine what it means for humans when parts of the blueprint are missing.”**

**Professor John Danesh**  
Director of the MRC/BHF Cardiovascular Epidemiology Unit at Cambridge University and Associate Faculty at the Sanger Institute



## 5 Why HbA1c diabetes test may fail African Americans

A genetic variant discovered in 11 per cent of African Americans could be responsible for large numbers of missed diagnoses of diabetes.

Type 2 diabetes is commonly diagnosed and assessed by monitoring levels of glycated haemoglobin (HbA1c) in the bloodstream. As blood sugar levels increase, the glucose molecules attach to haemoglobin, forming HbA1c. Therefore raised HbA1c levels indicate the presence of diabetes.

However, levels of HbA1c in the body are not solely affected by the amount of glucose in the blood. Genome-wide association studies (GWAS) have identified 18 loci that influence HbA1c levels, some by affecting blood glucose control, others by altering red blood cell function. Could such genetic factors affect HbA1c levels so strongly as to mask a person's diabetes status?

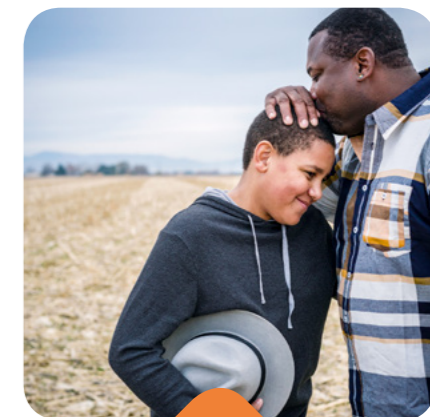
It was this question that spurred 200 scientists, including Sanger researchers, to carry out the largest international meta-analysis of GWAS data on HbA1c levels, published in *PLOS Medicine*.<sup>1</sup> They discovered that 11 per cent of African Americans carry an X chromosome variant in the *G6PD* gene that leads to more rapid recycling of haemoglobin, generating an artificially low HbA1c reading.

Relying solely on the HbA1c test to diagnose diabetes could mean that 2 per cent of African Americans – approximately 650,000 people – could remain undiagnosed. The results emphasise the importance of alternative tests for assessing type 2 diabetes in African Americans or of combining HbA1c monitoring with *G6PD* genotyping.



**Reference**

1. Wheeler E *et al.* Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLOS Med*. 2017; **14**: e1002383.



Over **65,000** African Americans could be misdiagnosed using HbA1c diabetes test



## 6 Greeks bearing (welcome) genetic gifts

A Sanger-led whole-genome sequencing study of 250 residents of the Greek island of Crete may have revealed why they remain anomalously healthy despite a diet rich in animal fat. Members of the isolated population of Mylopotamos had high levels of a genetic variant – rs145556679 – associated with low bloodstream levels of triglycerides and very-low-density lipoprotein, both important risk factors for heart disease.<sup>1</sup> This variant appears to be almost unique to the Mylopotamos population. The study illustrates how remote and isolated populations may harbour unusual genetic variants that have a significant impact on human health and disease.



**Reference**

1. Southam L *et al.* Whole genome sequencing and imputation in isolated populations identify genetic associations with medically-relevant complex traits. *Nature Commun*. 2017; **8**: 15606

**250** residents of Crete may have revealed the secret to a healthy long life



## 7 DNA analyses reveal secrets of human history

Sequencing projects have shone a fascinating light on extreme isolationism, genocide and assimilation around the world.



**381**  
Papua New Guinean people from  
**85**  
language groups were DNA sequenced



The first detailed genetic study of the people of Papua New Guinea, published in *Science*, has provided a new glimpse into ancient populations that have survived almost untouched by the modern world.<sup>1</sup>

Genotyping of 381 Papua New Guinean people from 85 different language groups, by an international collaboration including Sanger researchers, has revealed that the country's populations have been genetically independent from Europe and Asia for most of the past 50,000 years.

Within Papua New Guinea, groups speaking different languages were also genetically highly distinct from one another, demonstrating little mixing between tribes. Remarkably, despite being geographically quite close, groups living in the mountainous highlands and lowland populations are genetically separated by some 10–20,000 years.

A further notable feature is the apparent absence in Papua New Guinea of the genetic 'homogenisation' of populations

seen in Europe and elsewhere following the emergence of agriculture and a switch from a hunter-gatherer lifestyle. Potentially, later Bronze Age and Iron Age technological developments, which did not take hold in Papua New Guinea, may have been more important drivers of homogenisation.

In the Middle East, a comparison of five genomes from 4,000-year-old Canaanite skeletons and 99 modern-day Lebanese genomes suggests that the enigmatic Canaanites have left a substantial modern-day genetic legacy. Bronze Age Canaanites developed an extensive civilisation but have largely vanished from history, in part because they left few written records. According to Biblical accounts, the Canaanites were annihilated, but the genetic evidence points to their survival – 90 per cent of Lebanese ancestry could be traced back to the Canaanites.<sup>2</sup>

Long-term survival is also the theme of a separate study examining the genetic heritage of the Parsi, a culturally distinct South Asian population. Followers of the ancient religion of Zoroastrianism,

the Parsi are believed to have migrated to what is now India from Persia in the 7–10th century. An international collaboration has confirmed that the Parsi are genetically distinct, more related to ancient Neolithic Iranian populations than local Indian and Pakistani populations.<sup>3</sup> The genetic findings are consistent with Parsi historical accounts of assimilation into the Indian sub-continent populations in the seventh century, described as 'Like sugar in milk'.



### References

- Bergström A *et al.* A Neolithic expansion, but strong genetic structure, in the independent history of New Guinea. *Science*. 2017; **357**: 1160–1163.
- Haber M *et al.* Continuity and admixture in the last five millennia of Levantine history from ancient Canaanite and present-day Lebanese genome sequences. *Am J Hum Genet*. 2017; **101**: 274–282.
- Chaubey G *et al.* "Like sugar in milk": reconstructing the genetic history of the Parsi population. *Genome Biol*. 2017; **18**: 110.

## UK Biobank helps unlock the biology of osteoarthritis

Sanger scientists, with collaborators, have developed an integrated approach that exploits the power of the UK Biobank resource and DECODE database in Iceland to discover nine previously unknown regions of the genome associated with osteoarthritis.<sup>1</sup> By combining these findings with further genomic techniques, wet lab experiments and epidemiology the team was able to identify five promising targets for drug development and uncover biological pathways involved in the disease.

This unified approach, deploying the skills of specialists across a range of disciplines and data sources, could help to unlock other, equally opaque complex diseases.

Osteoarthritis is a debilitating disease that affects up to 40 per cent of people over 70 in the UK. No drug treatments are available, with pain relief and surgery the only therapeutic option.

It has a strong genetic component: up to 50 per cent of the variation in a person's risk of developing the disease is due to genetic factors. Yet the disease has proved remarkably resistant to genomic investigation, yielding only 21 disease-associated locations in the genome.

To unlock this complex condition, Sanger researchers conducted the largest genomic study of osteoarthritis to date. They studied 16.5 million DNA variations provided by the first release of data from the UK Biobank and identified 173 candidate genetic variants. After using the DECODE data to refine and confirm their findings in an independent population, nine previously unidentified disease-associated genomic regions remained.

To discover the biological pathways involved and offer new targets for drug development, the team conducted functional genomic experiments on diseased and healthy cartilage cells.

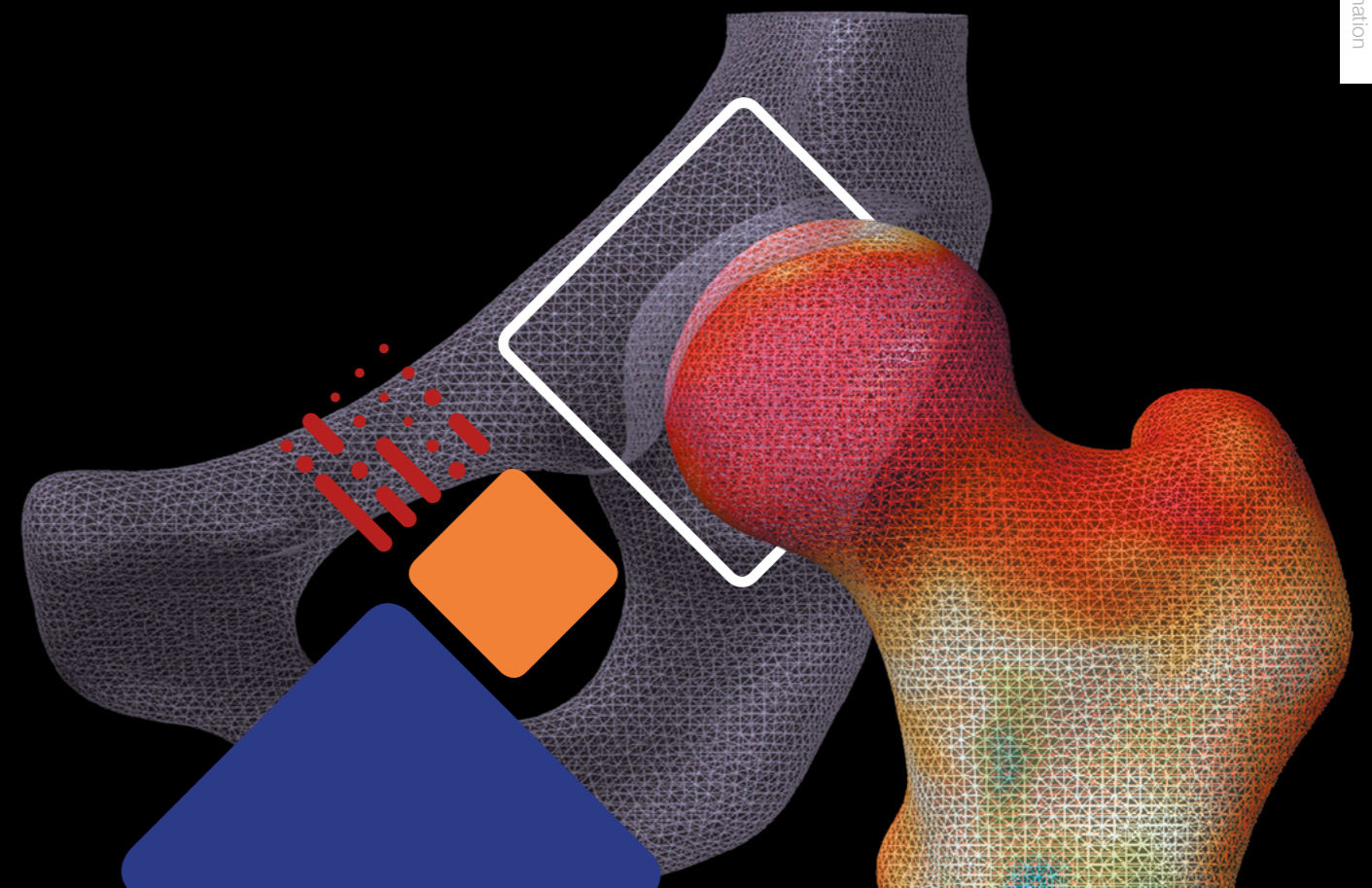
By applying proteomics and RNA sequencing, the scientists identified five genes within the candidate regions whose levels of activity were significantly decreased and, therefore, are likely to contribute to disease progression.

In addition to finding new drug targets the scientists discovered that, within the limits of their study, type 2 diabetes and high levels of lipids in the blood do not have causal effects on osteoarthritis, but reaffirm that obesity does.



### Reference

- Zengini E *et al.* Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. *Nature Genetics*. 2018; doi: 10.1038/s41588-018-0079-y.





# Infection Genomics

## In this section

- 1 Humans, not reservoirs, drive cholera epidemics
- 2 Human cell model shows our genes affect response to *Chlamydia*
- 3 Drug-resistant *E. coli* tracked in a care home
- 4 Slave trade helped melioidosis jump continents
- 5 Genomic surveillance in NHS reveals full picture of MRSA spread
- 6 Genomics cracks Icelandic horse mystery
- 7 Disturbed bacterial populations seek stability
- 8 Penicillin use drove methicillin resistance

1

## Humans, not reservoirs, drive cholera epidemics

**An analysis of historical *Vibrio cholerae* isolates has revealed how virulent cholera strains have spread through South America and Africa, with important implications for disease control strategies.**

Since the 1800s, the world has been swept by seven waves of cholera epidemics, the latest of which began in the 1960s and has killed many thousands of people worldwide. Important new insights into the recent global spread of cholera has come from a major genomic study, published in two papers in *Science*, from an international team led by scientists from the Sanger Institute and the Institut Pasteur in France.<sup>1,2</sup>

The team analysed 1,200 cholera samples from 14 South American and 45 African countries covering a 49-year period, and reconstructed evolutionary trees to show how they were related. The genomic data indicate that the strain responsible for the seventh wave, 7P El Tor (7PET), has been repeatedly introduced from South Asia into Africa since the 1970s and into South America in the 1990s.

Notably, in South America, the pandemic strain seems to behave very differently from those already present. 'Local' cholera strains persist in environmental reservoirs and can trigger outbreaks but do not seem to erupt into full-blown epidemics.

By contrast, 7PET appears to seed rapid explosive outbreaks. Identifying which strain is associated with an outbreak may therefore help public health officials mount the most appropriate response to contain its spread.

The spread of 7PET in Africa has been more complicated, with 11 introductions since the 1970s, triggering epidemics lasting up to 28 years. The last five introductions were all of multidrug-resistant strains originating in South Asia. However, a key similarity with South America is that African epidemics of 7PET appear to be triggered by human transmission of the disease, and are not due to long-term environmental reservoirs or climate events. This discovery will help focus control strategies on areas of greatest impact.



### References

1. Domman D *et al.* Integrated view of *Vibrio cholerae* in the Americas. *Science*. 2017; **358**: 789–793.
2. Weill FX *et al.* Genomic history of the seventh pandemic of cholera in Africa. *Science*. 2017; **358**: 785–789.



We are getting a real sense of how cholera is moving across the globe, and with that information we can inform improved control strategies and basic science to better understand how a simple bacterium continues to pose such a threat."

**Professor Nick Thomson**  
Group Leader at the Sanger Institute

2



## Human cell model shows our genes affect response to *Chlamydia*

By combining stem cell reprogramming and gene editing, Sanger researchers have created an innovative new model to study host genetic factors that influence macrophage invasion by chlamydia. This model has the potential to be applied to other pathogens and tissue systems to explore a wide range of host-pathogen interactions and the role human genetics plays in influencing infection outcome.

*Chlamydia trachomatis* is one of the UK's most common sexually transmitted infections, and also an important cause of blindness worldwide. Chlamydia is difficult to study (and treat) because it tends to invade and replicate within macrophages. Attempts have been made to study the bacteria within mouse macrophages and macrophage cell lines but neither is a true model of natural infections.

To overcome this issue, Sanger researchers and colleagues in Canada have developed a new and more realistic model that

Chlamydia: by employing HipSci cells to create genetically reprogrammed macrophages, Sanger scientists identified two key immune genes that help to prevent chlamydia sheltering in macrophage cells

provides powerful opportunities to explore genetic factors affecting host responses to infection.<sup>1</sup> First, the team used cellular reprogramming technologies to create macrophages from human induced pluripotent cells created by the Sanger co-led HipSci project, and showed that the cells behaved like *bona fide* macrophages in their interactions with chlamydia.

In a second step, the researchers used CRISPR-Cas9 gene editing to eliminate specific genes potentially involved in host responses to chlamydia infection. These studies highlighted the key role of two genes involved in immune system function, *IRF5* and *IL-10RA*, in preventing *Chlamydia* from sheltering within macrophages. As well as suggesting new leads for therapy development, the model provides a tool for dissecting host responses to chlamydia – and, moreover, a similar strategy could be adopted for other intracellular pathogens.



### Reference

1. Yeung AT *et al.* Exploiting induced pluripotent stem cell-derived macrophages to unravel host factors influencing *Chlamydia trachomatis* pathogenesis. *Nat Commun*. 2017; **8**: 15013.

3

## Drug-resistant *E. coli* tracked in a care home

A Sanger-led study has identified extensive transmission of antibiotic-resistant *Escherichia coli* in a care home. Care homes are known to be reservoirs of drug-resistant pathogens, but are not included in current surveillance programmes. The study used genomic methods to track drug-resistant and drug-sensitive *E. coli* strains in 45 care home residents over a six-month period. Slightly more than one third of participants carried antibiotic-resistant *E. coli*, which was spread both within the care facility and to a local hospital.<sup>1</sup> With increasing numbers of people requiring long-term care, the data point to a potentially important new application of genomic surveillance to monitor community reservoirs of drug resistance.



### Reference

1. Brodrick HJ *et al.* Longitudinal genomic surveillance of multidrug-resistant *Escherichia coli* carriage in a long-term care facility in the United Kingdom. *Genome Med*. 2017; **9**: 70.

Care homes are an important reservoir of drug-resistant *E. coli*



Cholera still affects  
**47**  
countries worldwide  
and kills almost  
**100,000**  
people each year



What we do  
Our work  
Our approach  
Other information

Melioidosis kills approximately  
**89,000**  
people a year

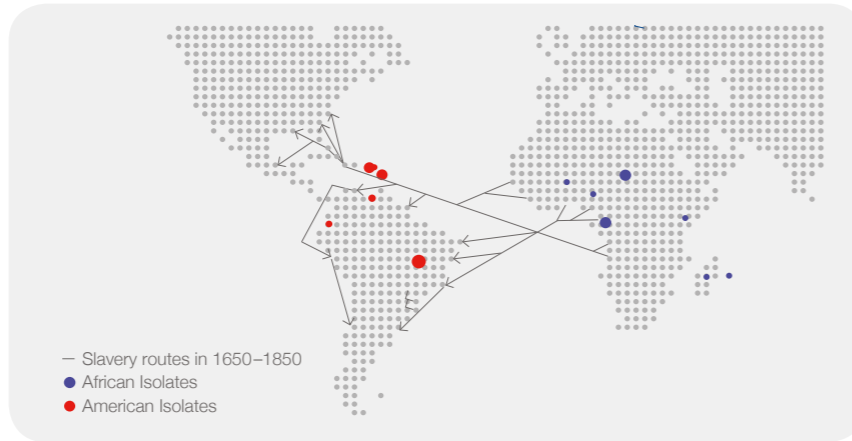
#### 4 Slave trade helped melioidosis jump continents

Genome sequencing has revealed how and when the bacterium causing melioidosis spread from its base in Australia.

Despite a high fatality rate, melioidosis has tended to pass under the global radar. Caused by infection with the soil-dwelling bacterium *Burkholderia pseudomallei*, melioidosis affects around 165,000 people a year and kills an estimated 89,000. As well as the scale of its impact, its potential use in bioterrorism has recently raised its profile.

To provide insight into its global dissemination, an international team led by Sanger researchers sequenced 469 *B. pseudomallei* isolates from 30 countries collected over 79 years, reconstructing its evolutionary history.<sup>1</sup> The sequence data suggest that *B. pseudomallei* originated in Australia and was transferred just once into South-East Asia.

From this single introduction, it then spread repeatedly through countries in South-East Asia and also into Africa and South America.



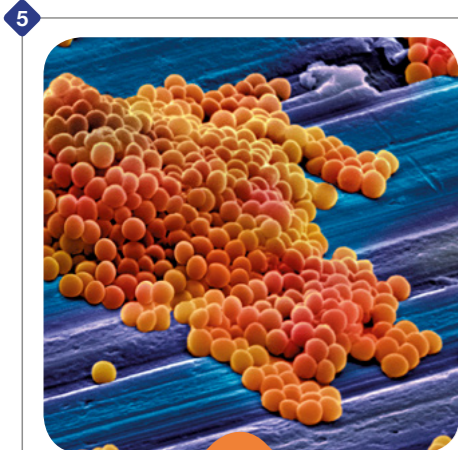
The dates for the likely introduction into South America – between 1682 and 1849 – strongly implicate the Slave Trade in its transatlantic spread.

The genetic data also provide a possible answer to a clinical curiosity. Melioidosis symptoms differ geographically, particularly between South-East Asia and Australia – potentially because the genetic make up

of isolates varies markedly between these two locations.



**Reference**  
1. Chewapreecha C *et al.* Global and regional dissemination and evolution of *Burkholderia pseudomallei*. *Nat Microbiol.* 2017; **2**: 16263.



#### 5 Genomic surveillance in NHS reveals full picture of MRSA spread

Routinely combining genomics with epidemiology within an NHS region would reveal networks of MRSA transmission in hospitals and community settings that would otherwise go undetected, Sanger scientists have found.

An influential 2013 study demonstrated the power of genome sequencing to identify MRSA outbreaks in hospitals and likely chains of infection.<sup>1</sup> But MRSA also circulates outside hospitals, so a full picture of its spread will require analysis of samples from the wider community.

In a study published in *Science Translational Medicine*, Sanger scientists and their colleagues analysed all MRSA samples handled by a large diagnostic laboratory in the East of England over the course of

a year, encompassing samples from three hospitals and 75 GP practices. The study generated both genomic and epidemiological (time and place of sample collection) information for more than 2,000 isolates.<sup>2</sup>

The analysis revealed 173 transmission clusters, most of which had not been previously identified, varying in size from two to 44 individuals and involving nearly 600 people in total. Some clusters were centred on hospitals, others were based in the community, and some involved transmission between the two. While MRSA strains are generally thought to be adapted to the hospital environment, hospital-associated lineages were perfectly able to spread in the community, with transmission seen in households, care homes and at GP practices.

The data point to significant shortcomings in current infection control procedures. They also highlight the importance of extending MRSA surveillance outside the hospital setting, as transmission networks clearly extend into the community.

Analysis involved

**173**

transmission clusters, varying in size from

**2-44**

individuals

#### 6 Genomics cracks Icelandic horse mystery

Genomic sleuthing by Sanger scientists has revealed the origins of a mysterious epidemic affecting the Icelandic horse population.



Icelandic horses have been isolated from the rest of the equine world for more than  
**130 years**

More generally, the study suggests that integrated genomic and epidemiologic surveillance could play a pivotal role in outbreak investigation, to support more effective infection control and rapid halting of transmission. A follow-up study is exploring the potential of real-time sharing of genomic data with clinicians, to enable sequencing to be integrated into routine clinical management.



**References**  
1. Harris SR *et al.* Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect Dis.* 2013; **13**: 130–6.  
2. Coll F *et al.* Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. *Sci Transl Med.* 2017; **9**: pii: eaak9745.

For more than 130 years, the Icelandic horse population has been isolated from the rest of the equine world. This isolation has kept Iceland free of the major diseases that affect horses, but leaves the horse population highly vulnerable to infections they have not encountered before.

Despite strict biosecurity measures, in 2010 the island's 77,000 horses were badly affected by a mysterious respiratory disease. When conventional methods failed to identify a cause, the Animal Health Trust and Sanger Institute were brought in to investigate. Genomic analyses of 257 samples from affected animals suggested that a virulent strain of a bacterium found in healthy horses, *Streptococcus zooepidemicus*, was the likely culprit.<sup>1</sup>

More conventional epidemiological analysis strongly suggested that an equine rehabilitation centre was the source of the Icelandic epidemic. Horses exercised in a communal water treadmill at the centre, which probably provided an ideal breeding ground for the bacterium to thrive and spread from horse to horse.

Since no horses were imported into Iceland, the epidemic strain of *S. zooepidemicus* probably arrived via contaminated tack or an asymptomatic human carrier from overseas; the strain has been found in a horse in Sweden and in an infected person in Finland.



**Reference**  
1. Björnsdóttir S *et al.* Genomic dissection of an Icelandic epidemic of respiratory disease in horses and associated zoonotic cases. *MBio.* 2017; **8**: pii: e00826–17.



## 7 Disturbed bacterial populations seek stability

As well as antibiotics and vaccination, interaction between different strains of bacteria plays an important role in the survival and make-up of bacterial populations.

The alarming rise in antibiotic resistance is often used to illustrate the power of natural selection – resistant microbes prosper at the expense of the susceptible. However, two studies from Sanger scientists in 2017 paint a more complicated picture, with survival also depending on competition between bacterial strains for space in environmental niches.

The first study published analysis of more than 1,500 *Escherichia coli* isolates from national and regional hospital collections of bloodstream infections, from 2001–2011. During this time, a globally disseminated multidrug-resistant strain, ST131, appeared, as did a new drug-sensitive strain, ST69. However, rather than taking

over completely, ST131 numbers stabilised within a few years; a similar pattern was seen with ST69. Hence, after a new strain emerges, population equilibrium is initially perturbed but then stabilises.<sup>1</sup>

The likely explanation is that drug resistance is not always a selective advantage. *E. coli* is a usually harmless bacterium living in the gut, competing for living space. Drug-resistant strains are strongly favoured when bacteria establish bloodstream infections and are treated with antibiotics. But selection appears to act against strains that become too common (an effect known as ‘negative frequency-dependent selection’), halting their expansion and leading to the maintenance of population diversity.

Interestingly, a similar phenomenon is apparent after pneumococcal vaccination. Multiple strains of *Streptococcus pneumoniae* exist, only some of which are targeted by pneumococcal vaccination. An important question, therefore, is what happens to the population structure when particular strains are eliminated by vaccination.

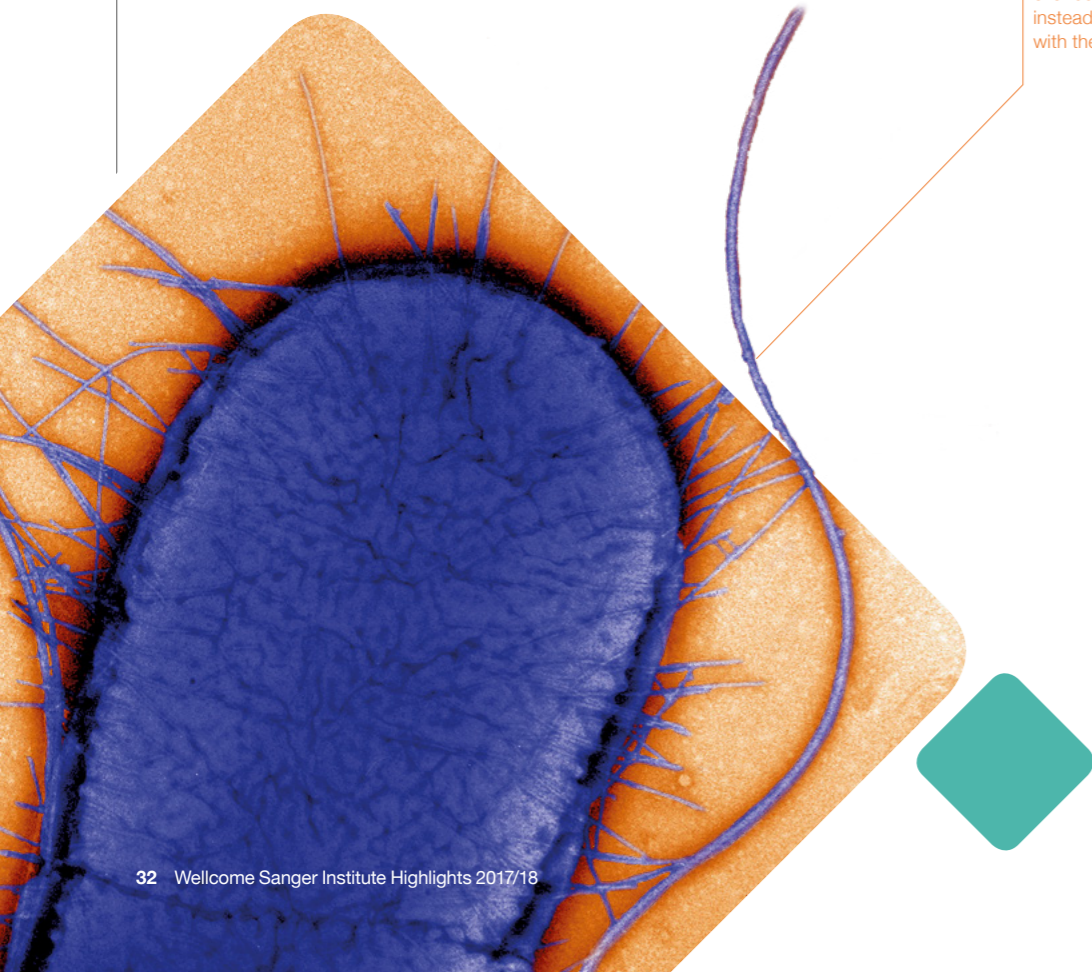
In a study combining genetic analysis and modelling, Sanger researchers and their international colleagues took a gene-centred view. They discovered that the frequency of individual genes was remarkably similar before and after vaccination, even though the strain composition changed markedly.<sup>2</sup> Hence there again appears to be a population equilibrium, driven by competition between strains within an environmental niche, with negative frequency-dependent selection acting as a key stabilising factor both before and after vaccination.



### References

1. Kallonen T *et al.* Systematic longitudinal survey of invasive *Escherichia coli* in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. *Genome Res.* 2017 Jul 18. doi: 10.1101/gr.216606.116.
2. Corander J *et al.* Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. *Nat Ecol Evol.* 2017; **1**: 1950–1960.

In bloodstream infections, no single *E. coli* strain ever completely outcompetes all other strains, instead any new strain finds a stable equilibrium with the others



## Penicillin use drove methicillin resistance

*Staphylococcus aureus* acquired the key gene for methicillin resistance, *mecA*, years before methicillin began to be used therapeutically, a genomic analysis of more than 200 historical *S. aureus* samples has revealed.<sup>1</sup> The analysis suggests that MRSA emerged in the 1940s, more than a decade before the first medicinal use of methicillin. It is likely that the use of penicillin selected for a genetic element including the *mecA* gene, and these strains had a selective advantage once methicillin was introduced in 1959 – leading to the detection of methicillin-resistant strains within a year.



### Reference

1. Harkins CP *et al.* Methicillin-resistant *Staphylococcus aureus* emerged long before the introduction of methicillin into clinical practice. *Genome Biol.* 2017; **18**: 130.

More than

# 200

historical *S. aureus* samples were analysed



What we do

Our work

Our approach

Other information



# Malaria

## In this section

- 1 Mosquito diversity threatens control efforts
- 2 Malaria family ties revealed
- 3 Landmark study of half the genes in the malaria parasite genome
- 4 Missing link could boost RH5-based malaria vaccines
- 5 Genes and cell interactions drive T-cell fate
- 6 Attack all sides of malaria at once
- 7 Why some people are naturally resistant to malaria
- 8 Grand scale reveals hidden life of malaria parasite

## 1 Mosquito diversity threatens control efforts

### The largest ever genetic study of malaria-transmitting mosquitoes has found that they are among the most genetically diverse creatures on Earth – a finding with important consequences for mosquito control.

Control of *Anopheles* mosquitoes, principally through use of insecticides, has made a large contribution to recent falls in malaria infections. However, a surge in insecticide resistance threatens to derail mosquito control efforts.

To gain a better understanding of malaria-transmitting mosquitoes, the international *Anopheles gambiae* 1000 Genomes Project sequenced the genomes of 765 *Anopheles* genomes from 15 locations in eight countries. As reported in *Nature*, the *Anopheles* genome turns out to be remarkably variable, containing more than 50 million sites of single nucleotide variation in the 141 million base pairs of the genome that the team could analyse.<sup>1</sup>

This variation has significant implications for control. It provides abundant raw material from which resistance to insecticides might derive. In addition, it will be an obstacle to 'gene drive' control measures, which seek

to spread sterility genes through mosquito populations. Gene drive relies on gene-editing methods that target specific stretches of DNA sequence, many of which show high levels of genetic variability. More positively, the study identified a set of genes free of genetic variability that could be targeted in gene drive projects.

The data revealed that genetic variants responsible for insecticide resistance have been under intensive selective pressures, and spread more readily than previously thought. Indeed, genes seem to flow with remarkable rapidity across the African continent, not only through independent emergence but also by mosquito migration.



**Reference**  
1. The *Anopheles gambiae* 1000 Genomes Consortium. Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature*. 2017; **552**: 96–100.



## 2 Malaria family ties revealed

Genome sequences of the remaining elusive species of human-infecting malaria parasites have shed new light on *Plasmodium* family relationships and evolution.

In *Nature*, an international consortium led by Sanger researchers has reported the genome sequences of three relatively rare malaria parasites causing human disease – *Plasmodium malariae* and two species of *P. ovale*.<sup>1</sup>

Although less common and deadly than species such as *P. falciparum*, the three malarias still account for some 10 million cases a year and can establish infections that last for years. Because of their rarity and resistance to culturing, very little is known about their biology. The new

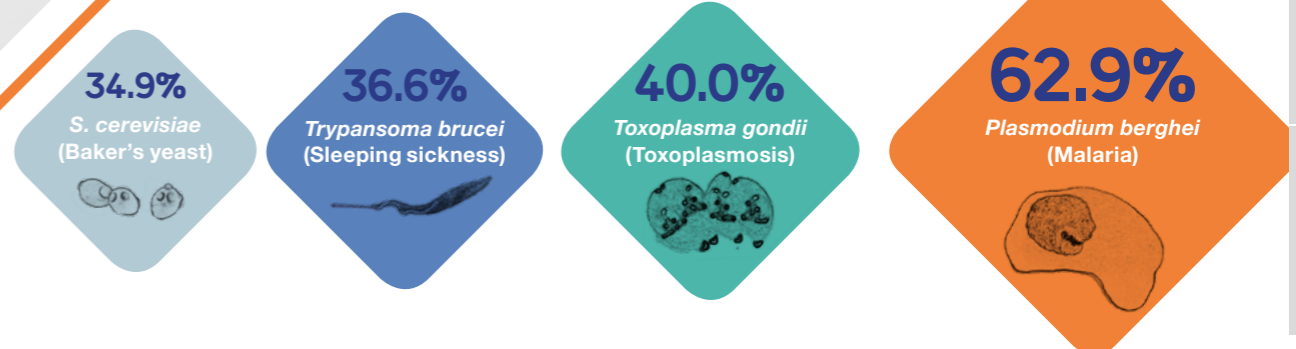
genome sequences will facilitate the development of new diagnostic tests to distinguish these unusual forms of malaria and potentially also new drugs and vaccines.

Furthermore, comparisons between human-infecting and other malaria parasites has revealed how they are related to one another, and shone light on genetic changes associated with adaptation to human hosts. The data provide insight into rapidly evolving genes involved in host invasion, including two new families of genes coding for proteins resembling the RH5 vaccine target.



**Reference**  
1. Rutledge GG *et al.* *Plasmodium malariae* and *P. ovale* genomes provide insights into malaria parasite evolution. *Nature*. 2017; **542**: 101–104.

### Percentage of genes required for normal growth



## 3 Landmark study of half the genes in the malaria parasite genome

New research shows why vaccines often fail, but also reveals targets for effective drug treatment.

A mouse model of malaria infection developed by Sanger researchers has enabled innovative genetic techniques to be used to investigate parasite biology. In *Cell* Sanger researchers and their colleagues in the *PlasmoGEM* consortium report the most extensive study ever undertaken of malaria parasite gene function, assessing the contributions to growth made by more than half the 4,600–4,700 genes in the parasite genome.<sup>1</sup>

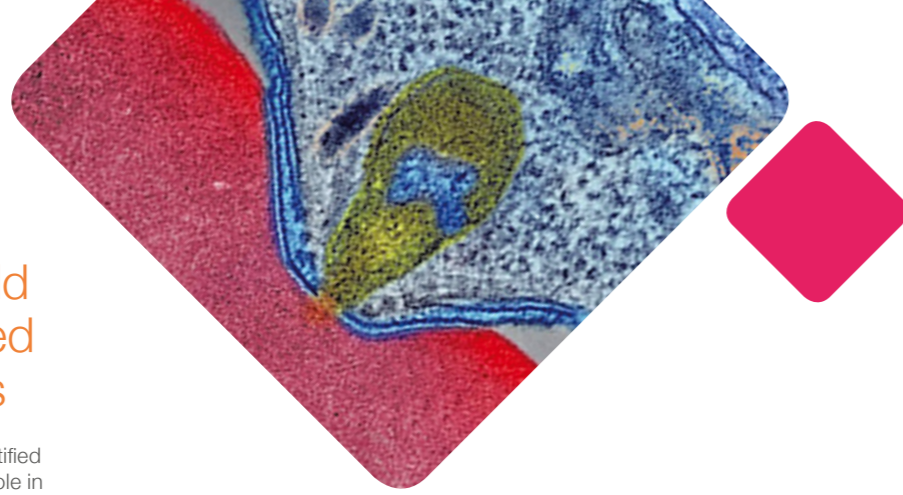
Using the mouse model, the team developed an innovative high-throughput gene knockout strategy for *Plasmodium berghei*. As well as eliminating individual genes, the technique added a unique barcode to the genome of each altered *P. berghei* cell. The barcodes enabled researchers to count the numbers of progeny cells generated in each knockout line, so the effects of each knockout on parasite growth could be determined.

The good news is that a surprisingly large number of 2,578 genes assayed – around two-thirds – were essential or important to growth. This probably reflects the fact that the parasite genome has been slimmed down to the bare minimum number of genes for survival. Encouragingly, this suggests that a relatively large number of genes could be good targets for drug development.

Less positively, genes that encode proteins visible to the host immune system were much more disposable – potentially explaining why it has proven so challenging to develop malaria vaccines targeting single antigens.



**Reference**  
1. Bushell E *et al.* Functional profiling of a *Plasmodium* genome reveals an abundance of essential genes. *Cell*. 2017; **170**: 260–272.e8.



## 4 Missing link could boost RH5-based malaria vaccines

Sanger researchers have identified the protein that plays a key role in the malaria parasite's attachment to red blood cells.

In 2011, Sanger researchers demonstrated that the RH5 protein of *Plasmodium falciparum* is essential for the parasite's invasion of red blood cells.<sup>1</sup> Disrupting the interaction of this parasite's protein and with the basigin protein on the surface of human red blood cells completely inhibits parasite proliferation *in vitro*. Subsequent studies have revealed that targeting this pRH5-basigin interaction offers great promise for developing both vaccines and drugs.<sup>2,3</sup>

However, there has been one vital piece missing from the picture. The parasite RH5 protein lacks any structure that would anchor it to the parasite surface, suggesting that it is a secreted protein. But, if this is the case, how can it mediate the parasite's attachment to the red blood cell?

Sanger researchers have solved this mystery by screening a library of parasite proteins using a sensitive assay, developed

at the Institute, to detect those that bind to RH5.<sup>4</sup> Their work revealed that a parasite surface protein known as P113 binds to one end of RH5, anchoring it to the parasite. When RH5 then binds to basigin on the surface of human red blood cells invasion begins. The P113-binding end of RH5 is cleaved off as the parasite enters the red blood cell, with internalisation driven by other protein interactions between parasite and host cell.

The team also chemically synthesised a peptide corresponding to the P113-binding portion of RH5, and showed that it retained the ability to bind P113 and elicited protective antibodies. Such a peptide has exciting potential as a component of a multisubunit vaccine.

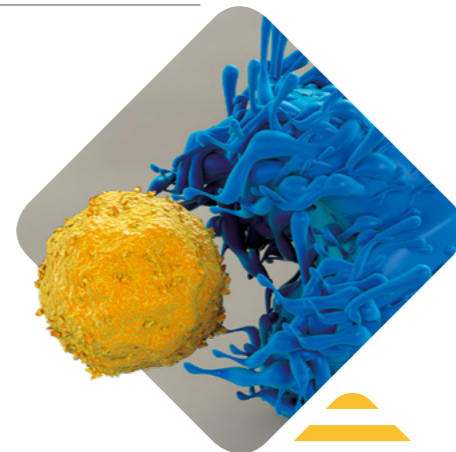
The strategy of targeting RH5 has also received a boost from a phase 1 vaccine trial carried out in Oxford, involving Sanger

researchers. This study showed that RH5 delivered via viral vectors stimulated anti-RH5 antibody responses far in excess of those seen in adults exposed to natural malaria infections.<sup>5</sup>



### References

1. Crosnier C *et al.* Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*. *Nature*. 2011; **480**: 534–7.
2. Zenonos Z *et al.* Basigin is a druggable target for host-orientated antimalarial interventions. *JEM*. 2015; **212**: 1145–51.
3. Wanaguru M *et al.* RH5-Basigin interaction plays a major role in the host tropism of *Plasmodium falciparum*. *Proc Natl Acad Sci*. 2013; **110**: 20735–40.
4. Galaway F *et al.* P113 is a merozoite surface protein that binds to the N terminus of *Plasmodium falciparum* RH5. *Nat Commun*. 2017; **8**: 14333.
5. Payne RO *et al.* Human vaccination against RH5 induces neutralizing antimalarial antibodies that inhibit RH5 invasion complex interactions. *JCI Insight*. 2017; **2**: pii: e96381.



## Insights

into CD4 T-cell differentiation could point to ways to boost immune response to malaria

Single-cell RNA sequencing combined with computational modelling of gene expression data revealed patterns of gene activity associated with a split into either Th1 or Tfh cells, which have distinct roles in immune responses. A third cell type, Tr1 cells, emerged as a spur from the Th1 lineage. Development of Th1 cells was dependent on expression of galectin-1, a protein previously implicated in regulating immune responses and a potential drug target.

The results provide new insights that could potentially be exploited to nudge CD4 T-cell differentiation in ways that protect against malarial infection. More generally, the computational tools developed could be used to dissect the differentiation of other important cell types.



### Reference

1. Lönnberg T *et al.* Single-cell RNA-seq and computational analysis using temporal mixture modelling resolves Th1/Tfh fate bifurcation in malaria. *Sci Immunol*. 2017; **2**: pii: eaal2192.

## 5 Genes and cell interactions drive T-cell fate

Single-cell genomics and computational modelling have been used to track the fate of CD4 T cells after malaria infection in mice.

Single-cell RNA sequencing represents a major technological advance, generating gene expression data on individual cells rather than populations. One key application is in tracking the differentiation pathways followed by cells, including immune cells after infection – work that could identify genes critical to effective immune responses to pathogens.

In an international collaboration, Sanger researchers used single-cell techniques to explore how mouse precursor CD4 T cells – important orchestrators of immune responses – develop into specialised subclasses following infection with the malaria parasite.<sup>1</sup>

## 6 Attack all sides of malaria at once

Creating a malaria vaccine that targets a number of different parasite proteins – each of which operate at a different stage of red blood cell invasion – could be the secret to effective prevention.

Malaria vaccine development faces two major challenges: the complexity of the parasite life cycle and the parasite's genetic diversity. This great diversity has meant that there is no highly effective vaccine currently available to combat the disease. Research from the Sanger Institute and its partners in 2017 suggest that a vaccine targeting a number of different proteins, operating at different stages of red blood cell invasion, would be much more powerful.

In 2014 Sanger scientists and researchers in Kenya reported that children whose immune systems responded to several proteins found on the surface of bloodstream stage parasites were at reduced risk of developing malaria over the following six months.<sup>1</sup> Furthermore, children with good responses across multiple antigens were almost completely protected against malaria.

This highly encouraging study relied on technology developed by Sanger scientists which enables the external domains of *Plasmodium* surface proteins to be produced in large quantities. Using this technology, Sanger researchers generated 29 antigens involved in invasion of red blood cells and, in 2017, showed that antibodies

against several of these proteins inhibited cell invasion across multiple parasite strains.<sup>2</sup>

Importantly, a number acted synergistically and high-resolution videomicroscopy suggested that this compounding of effect was due to targeting different aspects of the invasion process. Notably, with colleagues in Mali, the team also found that people carrying combinations of antibodies against these proteins were at reduced risk of malaria.

The study provides further support for multiple antigen targeting, and has identified at least five new potential targets for vaccine development. Moreover, the technology offers opportunities for additional large-scale studies to identify further targets.



### References

1. Osier FH *et al.* New antigens for a multicomponent blood-stage malaria vaccine. *Sci Transl Med*. 2014; **6**: 247ra102.
2. Bustamante LY *et al.* Synergistic malaria vaccine combinations identified by systematic antigen screening. *Proc Natl Acad Sci USA*. 2017; **114**: 12045–12050.



By bringing together multiple areas of expertise, from genomics to large field studies of patients in Mali, and down to advanced video microscopy observing individual parasites, we have discovered several new vaccine targets.”

**Dr Julian Rayner**  
Group Leader at the Sanger Institute

Sanger researchers generated

# 29

antigens involved in invasion of red blood cells





7

## Why some people are naturally resistant to malaria

Painstaking genetic reconstructions by Sanger researchers have revealed how rearrangements in glycoprotein genes can reduce some people's risk of developing severe malaria by up to 40 per cent.



In 2015, the Sanger co-led Malaria Genomic Epidemiology Network (MalariaGEN), reported in *Nature* that variation at the site of glycoprotein genes was found to provide significant protection against severe malaria.<sup>1</sup> Glycoproteins are a family of proteins that the malaria parasite interacts with as it invades host red blood cells, but it was unclear exactly which genetic features were protective.

To find the answers, the MalariaGEN team collected genome sequence information from an additional 765 people from 10 African ethnic groups. By combining this genomic data with information from the 1000 Genomes Project, and imputing the results into a data set of 4,579 people with

severe malaria and 5,310 controls, the team was able to piece together the stretch of DNA containing glycoprotein genes.<sup>2</sup>

The critical finding, published in *Science*, was that resistance to severe malaria was associated with a complex rearrangement, dubbed DUP4. DUP4 includes five glycoprotein genes, including two copies of a fusion between the *GYP A* and *GYP B* genes. The fusion genes resembled an unusual variant of the MNS blood group system, known as Dantu, and the team went on to show that DUP4 is indeed the basis of the Dantu blood group variant.

Notably, DUP4 was found only in East African populations, suggesting it has

arisen recently and been selected for because of the protection it offers against severe malaria. Further work could reveal the mechanism underlying this protective effect, and hence possible new ways to intervene to prevent severe disease.



### References

1. Malaria Genomic Epidemiology Network *et al.* A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature*. 2015; **526**: 253–7.
2. Leffler EM *et al.* Resistance to malaria through structural variation of red blood cell invasion receptors. *Science*. 2017; **356**: pii: eaam6393.



## Grand scale reveals hidden life of malaria parasite

Two projects in the Malaria programme demonstrate how Sanger scientists are using large-scale sequencing and analysis to discern the most intimate details of the malaria parasite's evolution and life cycle. One is applying whole-genome sequencing to monitor and understand the evolution of drug resistance across countries, while the other is unlocking the individual actions of single parasites at different stages in their life cycles.

In *The Lancet Infectious Diseases*, Sanger researchers detailed how their analysis of 1,500 *Plasmodium falciparum* genomes gathered over 11 years in 11 locations across South-East Asia has revealed how a multidrug-resistant form of the parasite evolved in Cambodia.<sup>1</sup> In 2008 a powerful antimalarial drug combination – dihydroartemisinin and piperaquine (DHA-PPQ) – was made the official first-line treatment in Cambodia and, within a year, the first resistant parasites had emerged. For the next five years, this resistant strain spread under the radar across Cambodia, outcompeting all other parasite strains.

The work shows that population-wide genomic surveillance of malaria can detect the emergence of drug resistance

much earlier than standard procedures, and supports the need for active genomic surveillance in malaria-affected countries.

At the other end of the scale is research applying single-cell technologies to understand every step of the parasite's life cycle. As part of the new Malaria Cell Atlas project now being conducted at the Sanger Institute, researchers analysed more than 500 individual *Plasmodium* parasites during their blood-borne stage.<sup>2</sup>

On average, the team was able to detect the activity of almost 2,000 genes in each parasite, just under half the genetic complement of the parasite and the largest number surveyed in individual malaria cells. This intimate view of the parasite's gene activity has revealed that a long-standing view of the genome's action was wrong. Instead of genes turning on and off in a slow cycle, the scientists discovered that whole sets of genes switched on and off in unison, suggesting that regulatory elements play a vital role.

The information they have gathered is being made freely available on an interactive, open-access Malaria Cell Atlas website.



Visit Malaria Cell Atlas site here [www.sanger.ac.uk/science/tools/mca/mca/](http://www.sanger.ac.uk/science/tools/mca/mca/)



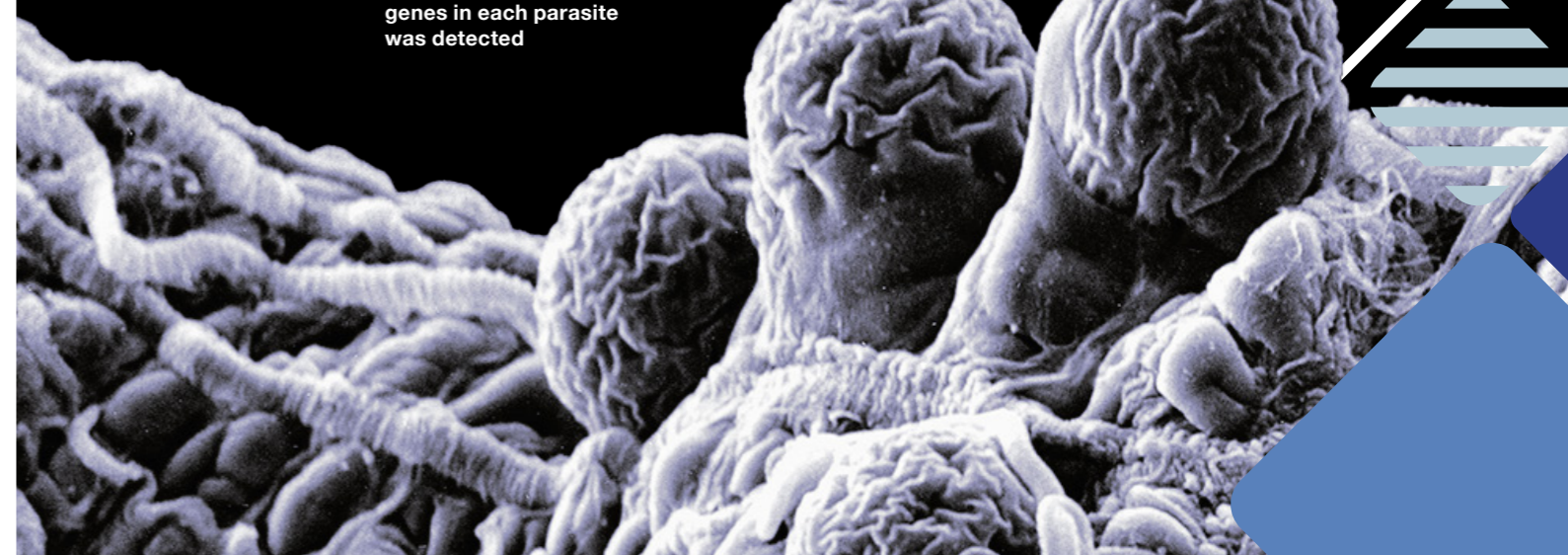
### References

1. Amato R *et al.* Origins of the current outbreak of multidrug-resistant malaria in southeast Asia: a retrospective genetic study. *The Lancet Infectious Diseases*. 2018; **18**: 337–345.
2. Reid A *et al.* Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. *eLife*. 2018; **7**: e33105.

The activity of almost

# 2,000

genes in each parasite was detected



What we do  
Our work  
Our approach  
Other information

# Our approach



We foster strong collaborations with scientists, clinicians, institutions, governments and society for mutual benefit



## 42 Scale

Genomic inquiry requires vast volumes of data, experimental models and computational power. Our Institute's unique, scalable and robust infrastructure delivers – both for us and researchers worldwide.



## 44 Innovation

To take our research findings to the next level and deliver transformative technologies we work in collaboration with biotechnology and pharmaceutical industries and funders.



## 46 Culture

As genomic research begins to impact clinical practice and society, our researchers are crossing traditional divides to work with entrepreneurs, health services and society.



## 48 Influence

By leading global initiatives and facilitating cross-cutting partnerships we seek to lay the foundations for a strong and vital future of genomic research, data sharing and clinical application.



## 50 Connections

We use the power of the internet and collaboration tools to build genomic research capacity worldwide and facilitate the next wave of discovery.

What we do

Our work

Our approach

Other information



# Scale

## In this section

- 1 Understanding human life – one cell at a time
- 2 5,000,000,000,000,000 bases of DNA and counting
- 3 Sanger partners with UK Biobank to sequence 50,000 people's whole genomes

# 1 Understanding human life – one cell at a time

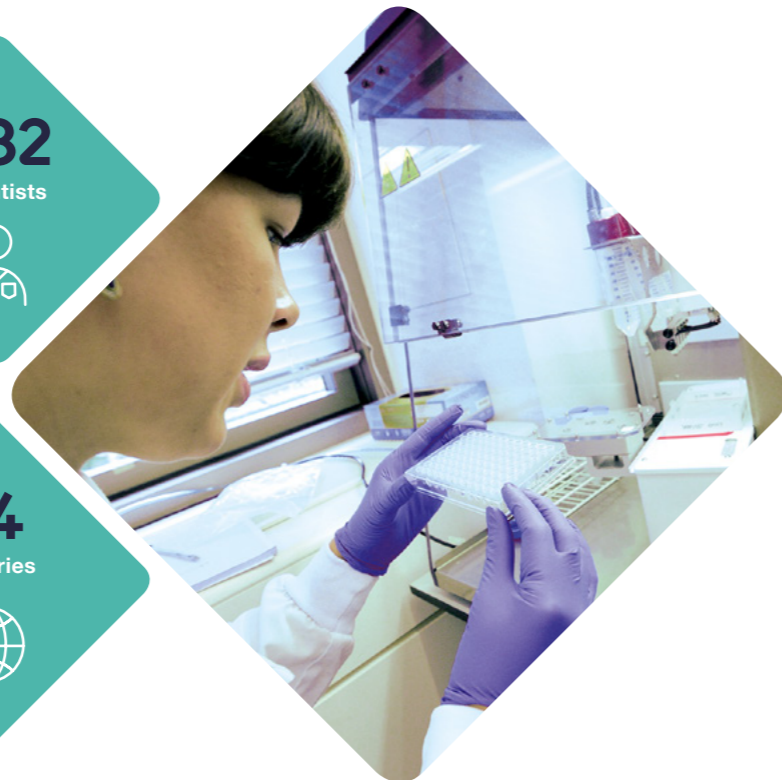
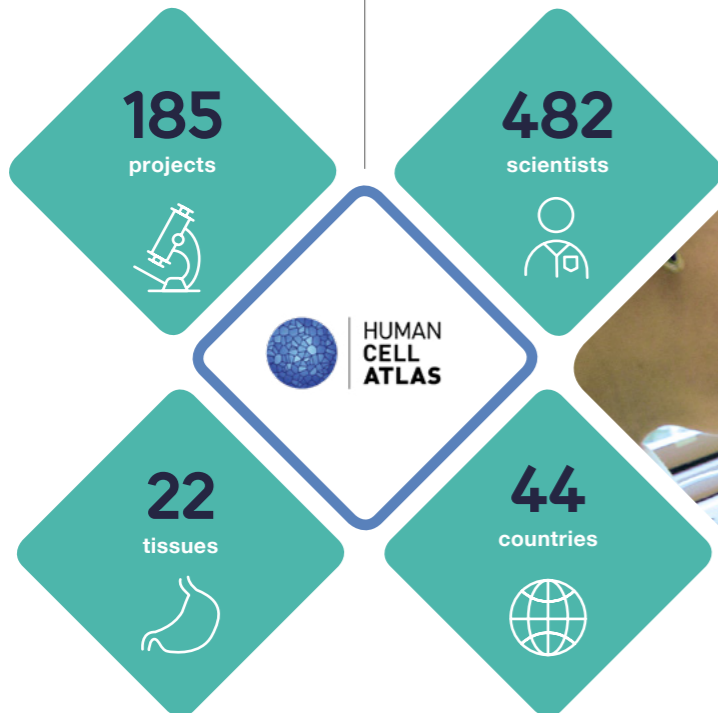
**The Sanger Institute, in conjunction with the Broad Institute, is co-leading the Human Cell Atlas initiative, which seeks to create a comprehensive reference map of every cell in the human body. Just like the Human Genome Project more than 25 years ago, the scope of the endeavour is vast, requiring both worldwide collaboration and the development of new cutting-edge techniques and methods of analysis.**

The ultimate aim of the Human Cell Atlas is to identify, and then investigate the genetic activity of every cell type – that make up an estimated 37 trillion cells in the human body. The work is divided into 185 projects, focusing on 22 different tissues in the body. So far, more than 480 scientists, in 44 countries around the world, are involved and more than 1.5 million cells have been studied.

One programme in which the Sanger is participating is the Human Developmental Cell Atlas, which is exploring the cells that are important for human development. By March 2018, researchers from the Sanger and Newcastle University had sequenced a quarter of a million separate cells from developing tissues including the kidney,

liver, placenta and skin. The ongoing data analyses from these samples will help to reveal which genes are activated, when, and in which cells as the body forms.

The knowledge this project will generate will help to shine a light on the importance of different biological processes at work within cells during development. This could help to unlock understanding of a range of health issues including miscarriage and children's developmental disorders. In addition, the discoveries made could reveal important insights into the biological pathways involved in ageing and cell repair, which could lead to advances in regenerative medicine.



# 2 5,000,000,000,000,000 bases of DNA and counting

The Sanger Institute has one of the world's largest DNA sequencing centres, enabling us to conduct genomic experiments at a scale that few can match. In fact, in 2018, our sequencing teams celebrated a remarkable achievement: they had read more than 5 Petabases of DNA.

To put the figure into context, it is the equivalent of reading 1,666,667 human genomes once. Or, if you strung together all the DNA molecules that would make up 5 Petabases, it would be 3,000km long. That's the same as travelling from Earth to the International Space Station seven and a half times.

Yet these figures do not tell the whole story. Honed by 25 years of continuous development to extract the maximum output from the latest sequencing machines, Sanger's sequencing teams are always searching for more ways to increase

the speed of delivery. Using high-throughput DNA sequencing, the first petabase of DNA took 1,952 days to read, in contrast the fifth petabase took just 169 days – a more than 10-fold increase in speed.

In fact, at this current rate, by the time you have finished reading this sentence, the Sanger's sequencing teams will have read another 273 million bases of DNA code.

## Time taken to sequence 1 Petabase by Wellcome Sanger Institute



# 3 Sanger partners with UK Biobank to sequence 50,000 people's whole genomes

The Sanger Institute will be applying its expertise in high-throughput DNA sequencing to read the full genomes of 50,000 people to gold standard for the UK Biobank. The Sanger was chosen because of its sequencing teams' proven experience in delivering high-quality human genome data at a scale that few in the world can match.

To deliver this work is a major undertaking, and demonstrates the ongoing revolution in DNA sequencing speeds. Reading 50,000 human genomes 30 times over to deliver gold standard reads will require the sequencing of 4.5 Petabases of DNA code. To put this into context, the project will require the Sanger's sequencing teams to read almost the same amount of bases as the entire sequencing output of the Institute over its 25-year history.

UK Biobank's aim is to improve the diagnosis and treatment of a wide range of serious and life-threatening illnesses, including cancer, heart disease, stroke, diabetes, arthritis and osteoporosis. For this reason, all the data generated by the Sanger will be made freely available by the UK Biobank to researchers in the UK and around the world to power future research.

The genomic information generated by Sanger will join UK Biobank's health and wellbeing databases that contain detailed information on medical histories, lifestyles, imaging and biochemical analyses. It is hoped that this resource will drive worldwide research by providing scientists with the high-quality data they need to conduct their investigations, sparing them the time and money that they would otherwise need to spend conducting experiments to gather the same information.

Also, this work may only be the beginning. UK Biobank contains DNA samples from 500,000 individuals. If successful, this project may lead to the sequencing of every UK Biobank volunteer's samples, creating the world's most detailed whole-genome database.





# Innovation

## In this section

- 1 Takeda joins Open Targets
- 2 Sanger spin-outs go from strength to strength
- 3 Kymab is one of Deloitte's UK Technology Fast 50

## 1 Takeda joins Open Targets

**In December 2017 Takeda Pharmaceutical became the fifth member of the pioneering precompetitive partnership – Open Targets – which brings together the Sanger Institute, EMBL-EBI and pharmaceutical companies to speed drug development and delivery through the application of big data.**

The Open Targets initiative was launched in 2014 to address a key challenge in drug development: that almost 90 per cent of all compounds entering clinical trials fail to become licensed medicines due to insufficient efficacy or patient safety. The root of this failure is often a lack of understanding of the biological target the drug is acting on.

Open Targets addresses this issue by marrying the expertise of two scientific worlds to create a critical mass of knowledge and data that could not exist in any one organisation. The initiative draws together the public domain expertise of the Sanger Institute and EMBL-EBI in generating and interpreting data from genomics, proteomics, chemistry and disease biology with the private research of Biogen, GSK and Takeda in areas spanning disease epidemiology, preclinical animal modelling, and biological processes.

The result of this public-private enterprise is that Open Target's partners are able to systematically identify small molecules that are most tailored to a disease's biological targets and prioritise those most likely to succeed for development. The goal is to reduce both the time and cost of drug development.

A cornerstone of this unique endeavour, hosted on the Wellcome Genome Campus, is the Open Targets Platform ([www.targetvalidation.org](http://www.targetvalidation.org)). This site openly shares all the sequence data and information gathered by the initiative to the wider scientific community to drive research and avoid duplication. As of February 2018, the platform offers information on 20,974 targets, 2,306,670 associations and 9,728 diseases from 17 data sources.

Experts in life-science data integration and analysis. Access to vast public domain resources

Expertise in the role of genetics in disease. Extensive experimental capabilities



## 2 Sanger spin-outs go from strength to strength

Two companies founded on long-term research developed at the Sanger Institute continue to garner success.

Congenica, the genome analysis company founded by Dr Matt Hurler and Dr Richard Durbin, successfully completed an £8 million Series B funding round and partnered with US-based Edico Genomics to offer an all-in-one genome analysis solution. It is already supplying services to Genomics England Ltd to help deliver the UK 100,000 Genomes Project and has secured contracts and partnerships around the world.

A key component of Congenica's offering is Sapiaientia™, a clinical decision management support platform that highlights potential disease-causing variants in a patient's genome and links these to observed symptoms and supporting research literature. This platform is being used to help deliver China's 100K Wellness

Pioneer Project, Portugal's In2 Genome genetic diagnosis project and is part of a memorandum of understanding, signed during the UK Prime Minister's tour of China, to bring clinical diagnosis support to China's national digital health platform.

Microbiotica, Sanger's most recent company, has won a number of awards in its first year of operation. Founded in December 2016 by the Institute's Dr Trevor Lawley and Professor Gordon Dougan, the spin-out seeks to develop and commercialise bacteriotherapies and biomarkers based on the human gut microbiome. The company has unique access to the techniques and bacterial libraries generated in the Institute to culture, characterise and phenotype the majority of a patient's gut bacteria.

Since its inception, Microbiotica has won OBN's UK Best Start-Up Biotech 2017 award, Biotech and Money's 'Life Science Spin-out of the Year' 2017 award and the One Nucleus Summer 2017 BioNewsRound Award.



Clinical scientists using Sapiaientia™ by Congenica

## 3 Kymab is one of Deloitte's UK Technology Fast 50

Kymab, a company spun out from the Sanger Institute to develop antibody and vaccine treatments, has been named by Deloitte as one of the UK's 50 fastest growing technology companies in 2017, based on its past four years of revenue growth. The company, founded on genome engineering developed by Sanger's Director Emeritus Professor Allan Bradley, has also been named one of Labiotech's top nine companies to watch in 2018.

The company's core technology is the Kymouse™ platform that recreates the entire diversity of the human immune system's B lymphocyte component in a humanised mouse. Using this platform, Kymab's researchers are exploring the role immunological processes play in health and disease to generate new, fully human, monoclonal antibody drugs and vaccines for a range of diseases ranging from cancer to infectious disease.

Founded in 2010, the company has raised a total of \$220 million in three rounds of investment funding. To deliver new products, the company has developed strategic partnerships in the fields of autoimmunity, cancer, blood-related diseases and infectious disease with Heptares, Novo Nordisk, the Bill and Melinda Gates Foundation, and MD Anderson Cancer Center.

One of its human monoclonal antibody treatments – KY1005 – is currently in clinical trial having shown strong potential in dampening exaggerated immune responses in bone marrow transplant patients, a common and potentially deadly complication. The drug works by addressing an underlying immune system imbalance found in many autoimmune conditions that prolongs T-cell response, causing tissue and organ damage.

Another of its antibodies – KY1044 – is on course to enter clinical trial in 2019. This drug works by improving the ability of a person's immune system to recognise and kill cancer cells. In particular, it has two important effects: it stimulates tumour-fighting immune cells and kills T regulatory cells, which are often found in tumours and reduce the body's immune response.

**\$220 million**  
has been raised by Kymab in three rounds of investment funding

Kymab has been named by Deloitte as one of the UK's **Top 50** fastest growing technology companies

What we do

Our work

Our approach

Other information



# Culture

## In this section

- 1 Sanger to train the next generation of scientific leaders
- 2 Free online course spreads genomic knowledge around the world
- 3 Sanger hosts UK's first data scientist degree apprenticeship scheme

## 1 Sanger to train the next generation of scientific leaders

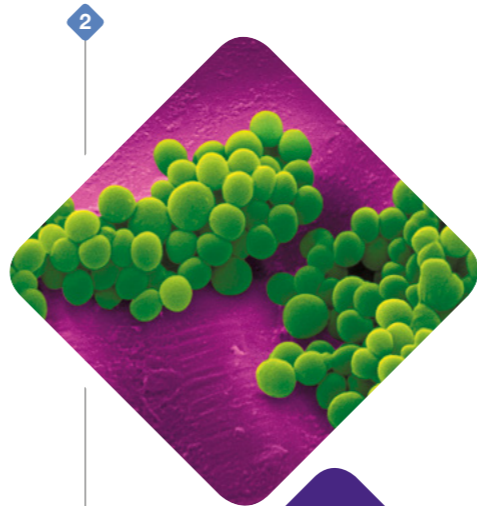
**As life sciences research becomes ever more complex, there is a need to develop leaders who can inspire and build strong communities of researchers dedicated to a common goal – both within single laboratories and across continent-spanning research networks. Yet the skills needed to successfully steer a project to conclusion are not taught to researchers at university, nor are commercially available leadership courses well placed to meet the unique demands of life science research.**

To ensure that the next generation of life science leaders are empowered with the management, negotiating and organisational skills needed for success, the Sanger Institute has started to share its bespoke internal leadership training programmes with the wider life sciences community.

The Institute's Scientific Education and Excellence Development (SEED) Scheme offers five training programmes. Each one is designed to support life science research leaders at very different stages of their career, from newly promoted supervisors and managers, through to more established managers and group leaders.

The courses have been specifically created to meet the needs of the Sanger Institute as it leads global networks of scientists in delivering cutting-edge science at scale. Sanger's long experience of coordinating and motivating diverse groups of scientists in the Institute and across global research networks, while also running one of Europe's largest high-throughput DNA sequencing and analysis pipelines, means that its training is ideally suited to the needs of the genomic and biological research community.

All the programmes focus on helping the participants effectively transfer their learning back to their day-to-day roles. Development is supported through interactive training sessions, feedback and personality profiling. In addition, by structuring the courses to encourage face-to-face discussion, the participants are given plenty of opportunities to build meaningful networks, share best practice ideas and develop behaviours that will drive their own success and that of their teams.



2

The course is the first of a series of

# 10

## Free online course spreads genomic knowledge around the world

For the first time, the Sanger Institute has helped to deliver a freely available web-based genomics course that will boost understanding among healthcare professionals and the public. The course covers the key role that DNA sequencing and genetic analysis can play in tracking and limiting disease outbreaks and the spread of antibiotic resistance.

Over the past eight years the Sanger has pioneered the use of genetic sequencing to track the spread of bacterial diseases through hospital wards, countries, continents and around the globe. Groundbreaking research into the spread of MRSA in a hospital ward demonstrated how applying genomics in a healthcare setting can provide early detection of disease transmission and allow clinical intervention to contain its spread. Additional research studies have shown that genetically analysing a bacteria's drug resistance profile can help to guide clinicians to the

most suitable choice of drug therapy far more quickly than current laboratory-based techniques.

Yet understanding of the use and power of genomic techniques among healthcare professionals is limited.

To address this knowledge gap, Sanger scientists from three research groups in the Infection Genomics Programme have contributed to a pioneering online course. The resulting training package, 'Bacterial Genomes: Disease Outbreaks and Antimicrobial Resistance', has been created by Wellcome Genome Campus Advanced Courses and Scientific Conferences in partnership with FutureLearn, a social learning platform founded by the Open University, and the University of Cambridge.

This course is the first of a series of 10 that will be created in conjunction with the Advanced Courses and Scientific Conferences team and FutureLearn. It has been specifically designed for doctors and nurses both in the UK and around the world but, because the course does not assume any knowledge of genomics or bacterial disease, anyone who wishes to explore the power of genomics to transform healthcare could take part.

3

## Sanger hosts UK's first data scientist degree apprenticeship scheme

As genomics and biodata play an increasing role in life sciences research and healthcare delivery, the need for skilled bioinformaticians becomes increasingly acute. Deloitte estimates that the UK genomics industry will grow by 20 per cent this year, while the report 'Big Data Analytics: Demand for Labour and Skills', has calculated that 56,000 UK-based new big data jobs will be created every year until 2020.

To help address this urgent need for computer scientists with big data skills, the Sanger Institute has partnered with Anglia Ruskin University's Degrees at Work team to run the first-ever UK data scientist degree apprenticeship for the bioinformatics profession. Covering data science, software engineering, biology and mathematics, the BSc (Hons) Bioinformatics undergraduate

course enables students to develop the unique skills necessary to visualise, analyse and interpret biological data through in-work training.

A truly Wellcome Genome Campus-wide endeavour, the programme is supported by a number of companies in the Biodata Innovation Centre, including Sanger spin-out company Congenica, Eagle Genomics, Genomics England, Global Gene Corp, SciBite and Specific Technologies.

The scheme will start for the 2019/20 academic year and is funded by a Higher Education Funding Council for England Degree Apprenticeship Development Fund.

# 56,000

UK-based new big data jobs will be created every year until 2020



What we do

Our work

Our approach

Other information





# Influence



Whipworm: UK school students are helping to analyse the genome of a parasite that affects millions of children in tropical countries

## In this section

- 1 School students join Sanger scientists to decode Whipworm's DNA
- 2 25 Genomes for 25 Years
- 3 Sanger Innovations help guide UK BioIndustry Association policies

# 1 School students join Sanger scientists to decode Whipworm's DNA

**A unique collaborative effort is bringing the world of neglected tropical diseases and bioinformatics out of the laboratory and into classrooms across the UK and Ireland.**

The results of the year-long initiative might even see school students included as contributors on research publications, and could set some of them on the way to filling the UK's bioinformatics skills gap.

Genome Decoders: Whipworm combines the skills of Sanger scientists and researchers in EMBL-EBI with the enthusiasm and curiosity of 10,000 A-level students across the UK to fully annotate the genome of the human whipworm (*Trichuris trichiura*). Coordinated by the Campus' Connecting Science Public Engagement team, in partnership with The Institute for Research in Schools (IRIS), the project aims to annotate all of the estimated 15,000 genes in the worm's genome.

Infection by the worm is called trichuriasis and it affects millions of children in tropical countries where sanitation is poor. It places a substantial health and economic burden on children by preventing them from attending school and damaging their economic futures. It is hoped that their UK peers will help to alleviate this burden by identifying targets for vaccine development and novel therapies.

Not only does this collaboration help to tackle a debilitating neglected tropical disease, it also addresses the pressing need for bioinformaticians. By engaging the students in a real-world scientific endeavour, the participants experience first-hand the excitement and challenge of a career in computational biology.

The high-quality DNA sequence they are interrogating was recently produced by the Sanger's Pathogen Genomics research group, and the analysis software that they are using has been supplied by the Wormbase team at EMBL-EBI and is regularly used by professional bioinformaticians.

But perhaps the most innovative result of the project may take longer to deliver: those pupils whose work substantively contributes to any future research papers will be included in the contributors list on the publications.



*Asterias rubens*: The Common Starfish has remarkable regenerative powers and is one of the genomes being sequenced by the Sanger Institute

# 2 25 Genomes for 25 Years

To celebrate its 25th anniversary, the Sanger Institute is marking the event in a way that uniquely embodies its values and mission – by generating genomes and biodata that are made freely available to the scientific community to drive research that would otherwise not be possible. The resulting project, 25 Genomes for 25 Years, is sequencing the genetic code of 25 UK species whose genomes are not currently available to the scientific community. To read more about the project, see page 51.

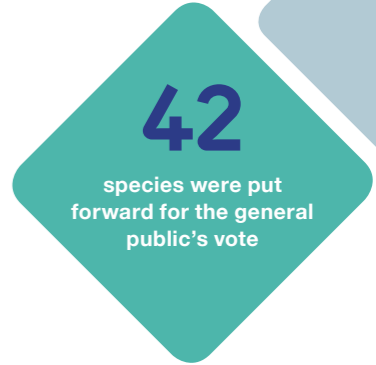
The project also offered the opportunity to embrace a key goal of the Wellcome Genome Campus, on which the Sanger is located: to engage with the public about the power of genomic data and bioinformatics to drive research and, ultimately, improve outcomes.

Organised by the Wellcome Genome Campus Public Engagement Team and partners, an unusual version of 'I'm a Scientist, Get Me Out of Here' gave the public the chance to choose five of the species that will be sequenced.

Throughout November 2017 schoolchildren and members of the public went online to quiz scientists about the 42 species that were put forward for inclusion, and what benefits DNA sequencing would provide.

Experts from across the country kindly gave up their time to answer the nation's questions. Some even replied as if they were a member of the species they were championing: for example the Leathery Sea Squirt, Common Crane and Orkney Vole amongst others.

After five weeks, more than 500 questions, and more than 4,000 votes, the final five species were decided: Common Starfish, Fen Raft Spider, Lesser Spotted Catshark, Asian Hornet and Eurasian Otter.



# 3 Sanger Innovations help guide UK BioIndustry Association policies



The Sanger is playing a central role in helping to guide the UK's support and development of world-leading genomics and biodata businesses. Adrian Ibrahim, Head of the Innovations team at the Sanger Institute, is the inaugural chair of the UK BioIndustry Association's Genomics Advisory Committee. In this role, he will help to guide the group as it provides the Association with expert guidance on the needs of, challenges to, and opportunities for, genomics companies.

Launched at the BIA's Committee Summit in early 2018, the group will provide a united voice to raise awareness of the genomics industry and the support it needs to flourish. In particular it will focus on engaging with both the UK Government and international organisations to:

- ensure that the Life Sciences Industrial Strategy and the Sector Deal are delivered in ways that optimally support the launch and growth of UK-based genomics companies

- collaborate with the Healthcare Advanced Research Programme (HARP) to determine the most productive moonshot programmes and ensure that the funding, infrastructure and industry connections are in place to enable success
- ensure that large datasets are used responsibly to deliver medical benefits, while also ensuring data privacy
- support the NHS in embedding genomic medicine into its healthcare systems to improve health outcomes for patients.

Other members of the committee include the Sanger spin-out company Congenica, and Eagle Genomics and Global Gene Corp, which are fellow tenant companies in the Wellcome Genome Campus Biodata Innovation Centre.





# Connections

## In this section

- 1 Sanger sets up global antibiotic resistance monitoring network
- 2 How 25 genomes are connecting Sanger with another world of biological research
- 3 Sanger is a key partner in Health Data Research UK

## 1 Sanger sets up global antibiotic resistance monitoring network

### A new global genomic surveillance unit is being set up by the Sanger Institute to help tackle the rise and spread of antibiotic-resistant bacteria.

Named as a 'catastrophic threat' by the UK Government's Chief Medical Officer in 2011, the burden of antibiotic-resistant bacteria could cripple healthcare systems if the issue is not successfully managed.

As demonstrated by Sanger researchers, pathogenic bacteria that spontaneously acquire new forms of antibiotic resistance or resistance to many of the most commonly used drugs, can appear anywhere in the world at any time. The work of Sanger Infection Genomics scientists has shown how these bacteria are then carried swiftly around the world by international travel, leading to national epidemics and global pandemics. Their research has also proved that these, often silent, developments could be detected early through genomic surveillance – regularly sequencing the genomes of circulating bacteria in populations across the world. This would allow governments to deploy effective and timely international public health strategies.

The first steps towards such a comprehensive, globally connected early warning system are being taken by the

Centre for Genomic Pathogen Surveillance, within the Sanger Institute. The Centre is an initiative dedicated to developing and disseminating best practice in genomic epidemiology, laboratory techniques and software engineering to improve the global surveillance of pathogenic bacteria.

Following the award of £6.8 million Official Development Assistance funding by the UK National Institute of Health Research (NIHR), the Centre is now working to establish a global genomic monitoring network spread across low- and middle-income countries.

Four sentinel sites are being established to conduct whole-genome DNA sequencing and genomic surveillance of antibiotic-resistant bacteria in Colombia, India, Nigeria and the Philippines. These sites were selected because they are already involved in their nations' antimicrobial programmes and are the laboratory testing stations for a number of local hospitals and healthcare networks. They also occupy strategic positions around the world, acting as gateways to understanding the landscape of antibiotic resistance in their respective continents.

The NIHR Global Health Research Unit on Genomic Surveillance of Antimicrobial Resistance, based at Sanger, will supply resources and training in genomic methods to the sentinel sites. This support will enhance their detection and research capacities, and help to embed cutting-edge genomic practice into their national healthcare systems. Because the data generated are digital, and will be uploaded to the UK centre, the sentinel sites will be able to share their information through online databases to provide clear, interpretable information for healthcare professionals and government officials.

It is hoped that, through this online network, these sentinel centres will benefit their nations' populations by providing cutting-edge insights for local health provision and prevention. In addition, by globally sharing their data, a prototype global early warning system will be established.

The Global Health Research Unit will connect with sentinel stations in Colombia, India, Nigeria and the Philippines to genomically monitor drug resistant bacteria worldwide



## Our approach Connections

## 2 How 25 genomes are connecting Sanger with another world of biological research

For its 25th anniversary, the Sanger Institute is generating a unique present to the UK and wider world: 25 new reference genomes for species that reflect the diversity of life in Britain. In so doing, the Sanger is forming new partnerships beyond its usual sphere of activity and is employing the power of genomics in new research fields to help understanding of ecosystems and preserve endangered species.

In conjunction with the Natural History Museum, the Sanger contacted more than 400 partner groups of wildlife experts to generate a list of 20 species that would be sequenced and a short list of 42 from which the public would select five (see page 49). Many of these species were then championed by these partners, exposing them and the public to the benefits that genomic sequencing could provide to conservation efforts.

In December 2017, the final list of 25 species was decided. Since this time Sanger scientists have been working in partnership with gardeners, university professors, conservation trusts, commercial organisations, environment charities, the Natural History Museum, and the UK Government's Natural England, to gather samples for sequencing. This experience is illuminating for everyone involved, and is revealing just how challenging conservation research can be, especially without a reference genome.

For example, gathering the correct blackberry sample for sequencing was no easy matter. There are approximately 360 species of blackberry and even specialists often need to observe the plant for a full year of its life cycle to determine its species.

It was by forming connections with experts in soft fruit research that the Sanger was able to obtain a verified sample. It was supplied by the National Clonal Germplasm Repository in Oregon, run by the US Department of Agriculture. And the US research connections for this plant continue courtesy of a serendipitous meeting at the Plant and Animal Genome conference in San Diego. There a Sanger scientist discovered a US-based researcher who had just started to sequence the same



There are approximately 360 species of blackberry

species of blackberry. Now both are working together to avoid unnecessary duplication and unlock this fruit's genomic secrets.

## 3 Sanger is a key partner in Health Data Research UK

The Sanger, with co-partners EMBL-EBI and the University of Cambridge, has been selected to join the UK's newest institute for health and biomedical informatics research: Health Data Research UK (HDR UK). The three organisations will form the Cambridge Centre of Excellence, one of six centres established across the UK that will apply big data to challenging healthcare issues.

Coordinated by the Medical Research Council, HDR UK operates in a similar way to Sanger's innovative Open Targets initiative: it seeks to draw together data scientists from a diverse range of disciplines and organisations to tackle challenges that no one research institute could attempt on its own. It will fund research that develops computational approaches that combine clinical data from routine healthcare

records, molecular research data, genomics, environmental and social data to deliver enhanced healthcare approaches and therapeutics.

The Cambridge Hub will blend the genomic talents of the Sanger Institute with the computational biology of EMBL-EBI and the clinical knowledge of the University of Cambridge to deliver fresh insights into health and disease. The ability of these organisations to work in such a complementary fashion demonstrates the value of strategic partnership and long-term investment in genomics and informatics to help to deliver the next-generation of precision medicine.

HDR UK encompasses 22 UK research institutes and organisations and is a joint investment coordinated by the Medical Research Council, working in partnership with the British Heart Foundation, the National Institute for Health Research, the Economic and Social Research Council, the Engineering and Physical Sciences Research Council, Health and Social Care Research and Development Division

(Welsh Government), Health and Social Care Research and Development Division (Public Health Agency, Northern Ireland), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, and Wellcome.

HDR UK encompasses

22

UK research institutes and organisations

What we do  
Our work  
Our approach  
Other information



## Image Credits

All images belong to Wellcome Sanger Institute, Genome Research Limited except where stated below:

Outside	Page 18	– 3D illustration of T cells – Shutterstock
Front Cover – Methicillin-resistant <i>Staphylococcus aureus</i> (MRSA) – Shutterstock	Page 19	– Nerve fibres in adult healthy brain – Zeynep M Saygin, McGovern Institute, MIT, Wellcome Collection
Inside	Page 21	– Yeast cells – Offset
Front Cover – Breast cancer cells – Annie Cavanagh, Wellcome Collection	Page 23	– Microscope – iStock, gut – Shutterstock, pregnant woman – iStock
Page 2 – Woman – H. Kuper, Wellcome Collection, Petri Dish – iStock	Page 24	– Father and son – Getty Images, heart – Shutterstock
Page 3 – Mosquito – CDC/Paul Howell, breast cancer cell spheroid, SEM – Khuloud T. Al-Jamal, David McCarthy & Izzat Suffian, Wellcome Collection	Page 25	– Father and son – Getty Images, Cretian woman – iStock
Page 5 – Baby music – thedanw, Pixabay, Mosquito head – CDC/Paul Howell, 8-cell embryo – K. Hardy, Wellcome Collection, chlamydia cells – Shutterstock, Cretian woman – iStock, Bowel section slide – Wellcome Collection	Page 26	– Papua New Guinean people – Freda Oppenheimer
Page 6 – Canaanite hieroglyphs – Thameen Darby, Wikimedia Commons, Houses of Parliament – iStock, father and son – Getty Images, MRSA cells – Annie Cavanagh, Wellcome Collection	Page 27	– Hip joint – Tom Turmezai, Ken Poole and Graham Treece, University of Cambridge, Wellcome Collection
Pages 8-9 – Health worker and child – Offset and iStock	Page 28	– Mother and child – H. Kuper, Wellcome Collection
Page 10 – Osteosarcoma of the thigh – Wellcome Collection	Page 29	– Chlamydia cells – Shutterstock, hands and walking stick – iStock
Page 11 – Medical discussion – iStock, breast cancer cells – Izzat Suffian, David McCarthy & Khuloud T. Al-Jamal, Wellcome Collection	Page 30	– MRSA cells – Annie Cavanagh, Wellcome Collection
Page 13 – 8-cell embryo – K. Hardy, Wellcome Collection	Page 31	– Icelandic horses – Pixabay
Page 14 – Cancer cell migrating through blood – Annie Cavanagh, Wellcome Collection	Page 32	– <i>E. coli</i> bacterium – iStock
Page 15 – Breast cancer cell spheroid, SEM – Khuloud T. Al-Jamal, David McCarthy & Izzat Suffian, Wellcome Collection, acute myeloid leukaemia cells – Paulo Henrique Orlandi Mourao, Wikimedia Commons	Page 33	– Petri dish – iStock
Page 16 – Medical computing – iStock, Antique tumour in paraffin wax – Great Ormond Street Hospital	Page 34	– Mosquito head – CDC/Paul Howell
Page 17 – Gut organoid – Anne Rios and Florijn Dekkers, Princess Maxima Center	Page 36	– <i>Plasmodium</i> invasion of red blood cell – NIH/NIAID, Dendritic cell and T-cell – iStock
	Page 37	– Injection – iStock
	Page 38	– Child – iStock
	Page 39	– Malaria parasites – Hilary Hurd, Wellcome Collection
	Pages 40-41	– Microscope – Getty Images
	Page 45	– Scientists using Sapientia™ – Congenica
	Page 47	– MRSA cells – Annie Cavanagh, Wellcome Collection
	Page 48	– Whipworm – Dave Goulding, Wellcome Sanger Institute, Genome Research Limited
	Page 49	– Common Starfish – Hans Hillewaert, Wikimedia Commons
	Page 51	– Blackberry – Pixabay

## Wellcome Sanger Institute Highlights 2017/18

The Wellcome Sanger Institute is operated by Genome Research Limited, a charity registered in England with number 1021457 and a company registered in England with number 2742969, whose registered office is 215 Euston Road, London NW1 2BE.

First published by the Wellcome Sanger Institute, 2018.

This is an open-access publication and, with the exception of images and illustrations, the content may, unless otherwise stated, be reproduced free of charge in any format or medium, subject to the following conditions: content must be reproduced accurately; content must not be used in a misleading context; the Wellcome Sanger Institute must be attributed as the original author and the title of the document specified in the attribution.



Printed by Park Communications on FSC® certified paper.

Park is an EMAS certified company and its Environmental Management System is certified to ISO 14001.

This document is printed on Soporset Offset, a paper containing 100% virgin fibre sourced from well-managed, responsible, FSC® certified forests.

Designed and produced by **CONRAN DESIGN GROUP**



**Wellcome  
Sanger Institute**

**Tel:**  
+44 (0)1223 834244

**[sanger.ac.uk](https://www.sanger.ac.uk)**

