Science with impact

Highlights 2019/20

wellcome sanger institute

What we do

Our work

World-class research and innovation that impacts people's lives

Read more about what we do at
sanger.ac.uk

Our approach

Other information

What we do helps us to understand and improve all life.

Gorilla hand
Read how reconstructing a 50,000-year-old parasite gene reveals how malaria jumped from gorillas to humans on page 32

Director's Introduction

As the UK Chief Medical Officer outlined in the 2019 report *Health, our global asset – partnering for progress* we live in a global village where the health issues of one country swiftly impact another. In this world, genomics – powered by open-access data sharing – is central to delivering the rapid, insightful information needed to shape successful therapies and effective healthcare delivery.

As evidenced by the outbreak of the Covid-19 coronavirus, seamless and open sharing of genomic data is guiding international public health responses and powering the worldwide race to develop vaccines. Here at the Sanger Institute, we analyse malaria parasite genomes from both neighbouring countries and continents to track the emergence and spread of drug resistance, guiding public health strategies. By successfully pooling datasets on cancer samples, databases and analysis we can work with fellow scientists across continents to reveal the causes and potential treatments for tumours, driving the development of precision, personalised medicine.

Research of this scale is a truly collaborative endeavour, both within the Institute, nationally and internationally. It is only through the tireless efforts of researchers in the field or in the wards to collect samples and phenotypic data, combined with the sequencing and computational skills of teams around the world, that we can make these scientific advances. To celebrate the invaluable work of our technicians to deliver science at scale, we are proud signatories to the Technician Commitment to develop the talents and careers of our staff. We are also delighted that our International Fellows scheme is enabling new continental networks of genomic research to grow and deliver science.

To deliver truly seamless sharing of genomic and clinical data, we actively support the work of the Global Alliance for Genomics and Health (GA4GH) to create the protocols and frameworks needed to open up the world's genomic databases to the global scientific community. Through innovations such as GA4GH Passports and Data Use Ontology, the process to gain access to much-needed data will now take a matter of days instead of weeks.

As the impact of genomics is more broadly felt across society as a whole, the Institute plays a pivotal role in helping policymakers and clinicians understand and apply its power. The Institute has a long history of providing expert guidance to the UK's policy makers and this year has been no exception, with our researchers advising the House of Commons Science and Technology Committee on Commercial Genomics and contributing to the Chief Medical Officer's annual report. And our researchers have helped to shape online genomics training courses that, in conjunction with Wellcome Genome Campus Advanced Courses and Scientific Conferences, are enabling healthcare professionals to embrace genomic medicine.

Genomic research is still in the foothills of extracting and using the knowledge buried in the six billion letters of code in the human genome. The ever increasing numbers of human genomes sequenced for research or clinical diagnosis will reveal patterns and motifs that will shape health and disease research for decades to come. When we also consider the rest of the genomes on Earth the potential is vast and the Sanger Institute will be in the vanguard of this revolution in science and society.

Professor Sir Mike Stratton, Director
Wellcome Sanger Institute

At a Glance

Sequencing Centre

In 2019, we read the genomes of **313** species

In 2019, we read the equivalent of one gold-standard (30x) human genome every **3.5 mins**

In 2019, the Sequencing Centre provided the equivalent of **2,880** gold-standard human genomes a week

In 2019, the Sequencing Centre outputted an average of **31,000bn** DNA bases a day

The human genome is approximately **3bn** base pairs long

Data Centre

Total number of compute cores **45,000** approx

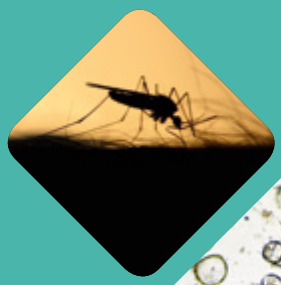
Network backbone speed **400GB/sec**

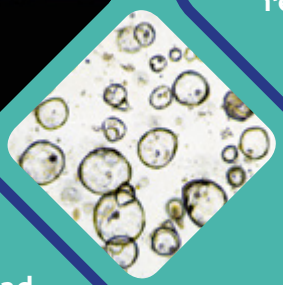
Centre-based high performance compute cores **25,000** approx

Usable storage in the Data Centre **65-70Pb** approx

Centre-based cloud-based flexible compute cores **20,000** approx

2019 timeline





Genome read from just half a mosquito-worth of DNA

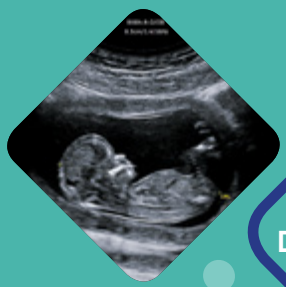
Cancer mutation resource reveals bursts of activity

Global Strep A vaccine one step closer
Page 30

Jan

Ian Dunham becomes Open Targets Director

Combining ultrasound with genomics improves prenatal diagnoses



Feb

100 organoid cancer models developed

100 new gut bacteria discovered

Molecular blueprint of early embryo development

Sanger Institute joins EDIS (Equality, Diversity and Inclusion in Science and Health)

Mar

Sanger Institute signs up to Technician Commitment

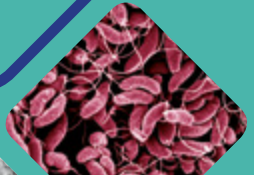
Sarah Teichmann wins 2020 Biochemical Society award

Crusaders made love and war
Page 23

Apr

Matt Hurles elected Fellow of Royal Society

Genetic code of WW1 soldier's cholera mapped



May

Peter Campbell elected to Academy of Medical Sciences

CRISPR catches critical cancer changes

Building blocks of lungs mapped for first time
Page 19

Jun

Nicole Soranzo honoured by EMBO

Gene mutations offer MRSA drug options

Baby biome shows vaginal delivery boosts maternal bacteria transfer

Page 26

Drug targets for liver-stage malaria found

Page 33

Three CT scans could promote cancer development

Page 11

Jul

Institute research featured in CMO's report

Genomic monitoring shows spread of multidrug-resistant malaria in Asia

Houses of Parliament Select Committee evidence on Commercial Genomics

Aug

Genomes reveal Africa's malaria control is at risk

Malaria Cell Atlas reveals parasite's secrets at every stage of life cycle

Diarrhoea-causing bacteria adapted to spread in hospitals

Page 27

Sep

Sanger scientists support LifeLab event across East Anglia

Sanger co-sequencing 450,000 genomes for UK Biobank

Brown Trout genome may reveal species' superpowers

Oct

Mathew Garnett receives National Cancer Research Institute Excellence Award

Jumping genes can cause rare developmental disorders

Pathway to colon cancer mapped

Measles wipes the immune system

50,000-year-old gorilla to human malaria gene resurrected

Page 32

Nov

Wellcome funding drives project to map all life in Britain and Ireland

Sam Behjati is recognised with early career research award

Dec

140 genes linked to immune system control

Successful merge of cancer datasets creates comprehensive resource

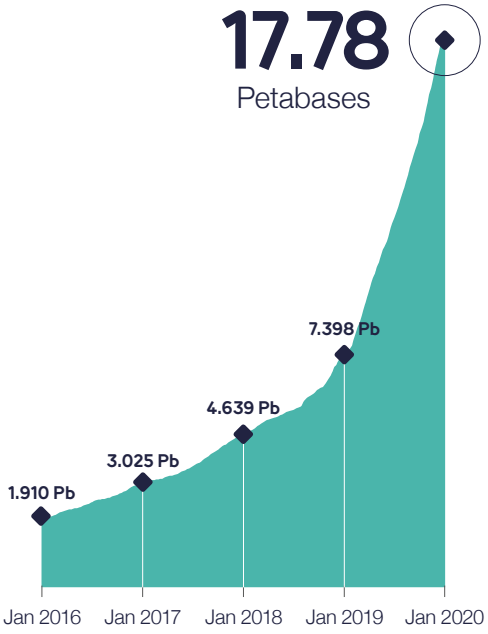
Missing gene reveals two IBD pathways

Year in Numbers

568 research articles in 2019

- *Cell* — 8
- *Lancet* — 1
- *Nature* — 20
- *Nature Genetics* — 21
- *Science* — 10

17.78 Petabases



Who published our work?

Where Sanger staff are from*

How much DNA was sequenced?

*In January 2020

Our work is at the forefront of genomic research.

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

Tree of Life
Discover how we enable researchers to study life at the most intimate detail on page 36

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.

Page 10

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.

Page 16

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.

Page 22

Parasites and Microbes

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.

Page 26

Tree of Life

Investigates the diversity of complex organisms found in the UK through sequencing and cellular technologies. It also compares and contrasts species' genome sequences to unlock evolutionary insights.

Page 36

Cancer, Ageing and Somatic Mutation

In this section

1

Uncovering the pathway to colon cancer

The hidden world of genetic mutations in healthy colon tissue has been uncovered. Sanger researchers shed light on the very earliest stages of cancer and help to understand how a healthy cell becomes a cancerous one.

Researchers at the Sanger Institute have developed technology to sequence the genomes of small numbers of cells, allowing them to study genetic mutations in unprecedented detail.

The team studied cells from 2,035 colon 'crypts', the tiny cavities that form colon tissue. They developed a method to sequence DNA from just a few hundred cells at a time, using over 1,000 times less DNA than standard protocols. The method had to work on cells, rather than the purified DNA the team would normally use, and it had to be effective without any DNA amplification, which could introduce errors that would obscure mutations. The Sequencing Research and Development team at the Sanger Institute developed a new workflow and then fully automated the process enabling tens of thousands of genomes to be swiftly sequenced. The method has since been adapted for other projects where only tiny amounts of DNA are available, for example in studies of parasites.

In the colon study, published in *Nature*, researchers found mutations that represent the very earliest stages of cancer development. The cells were morphologically normal, but had distinct patterns of genetic changes, including in genes known to drive cancer growth.

The team also uncovered mutational signatures in the cells – tell-tale patterns, or 'fingerprints' of specific DNA changes caused by a specific carcinogenic chemical or process. Some signatures were ubiquitous and continuous, others were present in only a few individuals, in particular crypts or during certain periods of life. Six signatures had never been seen before, and further research is needed to determine their cause – potentially uncovering hidden causes of colon cancer.

Reference

Henry L. *et al.* The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* 2019; **574**: 532–537.

2

Why it's never too late to stop smoking

Sanger researchers and their collaborators have uncovered protective cells in the lung that are shielded from the harmful effects of tobacco. These cells re-awaken in ex-smokers to actively repair the airways after someone stops smoking, protecting against lung cancer.

When someone stops smoking, regardless of their age or how long they smoked for, their risk of lung cancer immediately reduces. It continues to fall with time. To understand why, and to quantify the effects of smoking on healthy lung cells, researchers at the Sanger Institute and UCL compared the genetic changes in non-cancerous lung cells from smokers, ex-smokers and people who have never smoked.

The team sequenced the genomes of 632 cell clones from the lungs of 16 people. As expected, they found thousands more genetic mutations in cells from smokers compared with non-smokers. More than a quarter of these damaged cells had at least one cancer-driver mutation, explaining the increased risk of lung cancer in smokers. Unexpectedly, the researchers saw huge genetic variation between individual cells. Smoking added 1,000 to 10,000+ genetic mutations per cell, and cells from the same tiny biopsy of bronchial epithelium could vary 10 fold.

The researchers also found a population of cells in ex-smokers with less genetic damage from tobacco that looked similar to cells in people who had never smoked. Ex-smokers had four times more of these cells than people who still smoked – accounting for up to 40 per cent of lung cells. These cells have escaped the damaging effects of tobacco, and have a much lower risk of developing into a cancer. After stopping smoking, the undamaged cells can reawaken and repopulate the lungs, protecting against cancer.

Reference
Yoshida K. *et al.* Tobacco smoking and somatic mutations in human bronchial epithelium. *Nature* 2020; **578**: 266–272.

Lungs can repair themselves

Never smoked

100% of cells are 'near normal'

Smokers

90-96% of lung cells have up to **10,000** extra mutations compared with non-smokers

Ex-smokers

20-40% of cells are 'near normal'

4x more 'near normal' cells than smokers

3

Medical imaging radiation promotes cancer capable cells

Sanger Institute and University of Cambridge researchers studied the effects of low doses of radiation in the oesophagus of mice. Scientists found that doses equivalent to three CT scans promote the growth of cancer-capable cells in healthy tissue and recommend this risk should be considered in assessing radiation safety.

Sanger researchers previously found that our healthy tissues are battlefields, where mutant cells compete with healthy cells for space. These mutant cells accumulate as we age – though the vast majority do not go on to become cancer.

Reporting in *Cell Stem Cell,* the team investigated the effects of low doses of radiation on the cellular dynamics of healthy tissue. Low doses of radiation, like those experienced during x-rays and CT scans, are generally considered safe as they cause little DNA damage.

To investigate potentially hidden effects of radiation exposure, the researchers gave mice a 50 milligray dose of radiation, equivalent to three or four CT scans. This caused an increase in the number of cancer-capable cells with mutations in the *p53* gene. *p53* is mutated in 5–10 per cent of normal oesophageal tissue, but in almost all oesophageal squamous cell carcinomas – suggesting that these cancers arise from the *p53* mutant cell population in normal tissue.

The researchers then gave the mice an over-the-counter antioxidant – N-Acetyl Cysteine (NAC) – before the radiation exposure. The team discovered that the antioxidant gave normal cells the boost needed to outcompete and replace the *p53* mutant cells.

Until now, the effects of low levels of radiation on the DNA of healthy cells has remained hidden, making assessing the risks difficult. Future research will not only uncover the longer-term effects of radiation on cancer risk, it will also help scientists understand what sets a cell on the journey from healthy to becoming cancerous, shedding light on the very earliest stages of disease.

Reference
David, F. *et al.* Outcompeting *p53*-mutant cells in the normal esophagus by redox manipulation. *Cell Stem Cell* 2019; **25**: 3; 299–300.

p53 is mutated in 5-10% of normal oesophageal tissue

4

Dismantling cancer to find its weak spots

Researchers from the Sanger Institute, Open Targets and their collaborators used CRISPR-Cas9 technology to disrupt every single gene in more than 300 cancer models, representing 30 types of cancer. The team took cancers apart, piece by piece to discover thousands of key survival genes, and prioritised them to produce a hit-list for drug development.

Researchers across the globe are designing new targeted therapies that can kill cancer cells and leave healthy tissue unharmed. But developing such treatments is costly, and the vast majority of new drugs fail during development. Finding the correct targets to start with will massively accelerate the process, bringing new effective treatments to patients faster.

To find potential drug targets, the team, led by researchers from the Sanger Institute, GSK, the European Bioinformatics Institute (EMBL-EBI) and Open Targets, used CRISPR gene editing technology to disrupt each of 18,009 genes in 324 cancer models. The models – cells originally from a tumour but now grown in a laboratory – broadly represent patients' tumours, including the diversity of molecular features across different types of cancer. The cell models include common cancers as well as cancers of particular unmet clinical need.

After the disruption, the cells' growth was assessed and each gene was given a fitness score in each cell model. If cells didn't survive, the gene that had been disrupted was considered essential for that cancer's survival. This process resulted in a list of over 1,000 genes essential for the different cancers.

To prioritise the list, the team systematically integrated additional data, resulting in a ranked list of 628 potential drug targets. Their work is published in *Nature*, and the hugely valuable data resource is openly available to researchers around the world.

One of the top-scoring target genes present in multiple different cancer types was Werner syndrome RecQ helicase (*WRN*). The team found that cancer cells with a faulty DNA repair pathway, known as microsatellite unstable cancers, require WRN for survival. Microsatellite instability occurs in many different cancer types, including colon and stomach cancers. The new identification of WRN as a promising drug target offers an exciting opportunity to develop the first cancer treatments to target WRN – commonly seen in colon, ovarian, and endometrial cancers.

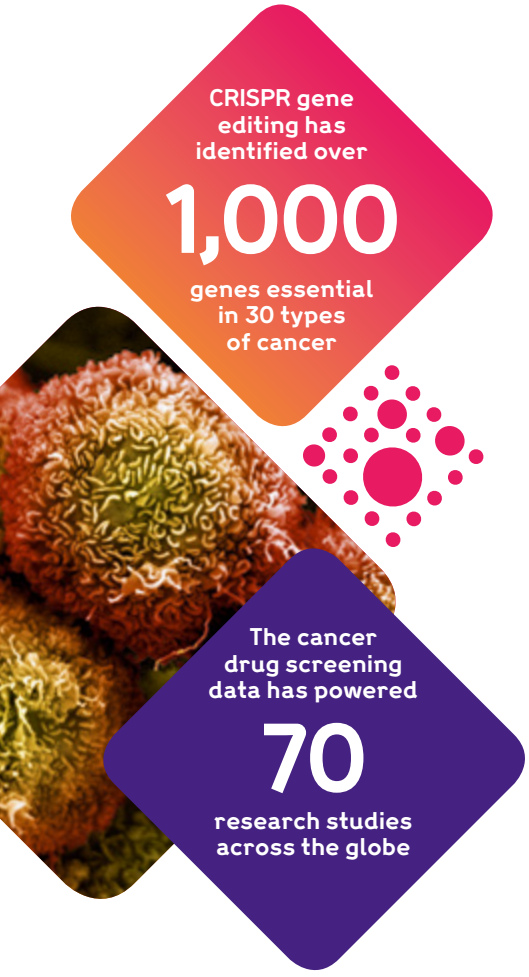
The collaboration between researchers at the Sanger Institute, EMBL-EBI and GSK, the Open Targets partners, bolsters the translation of these research results into new treatments.

The work is part of the Cancer DepMap project, and the analysis provides a resource of cancer dependencies, generates a framework to prioritise cancer drug targets and suggests specific new targets.

Reference
Behan, F. *et al.* Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature* 2019; **568**: 511–516.

Open Targets

Open Targets is a pioneering public-private collaboration that aims to transform drug discovery by systematically improving the identification and prioritisation of drug targets and improving the success rate for developing new medicines. **opentargets.org**



5

Cancer drug data will power next wave of discovery

Sanger researchers have generated and published the world's largest dataset showing how a cancer's underlying genomic landscape influences its response to treatment. The data, freely available to the global research community, will drive new developments in cancer treatment.

Cancer is genetically diverse – hundreds of different genetic changes can drive or inhibit a cancer cell's growth. Predicting which treatment will work for which patient remains a challenge in cancer medicine. Over the past 10 years, Sanger researchers and their partners have shown there is an intimate relationship between the genetic changes present in cancer cells and the way a drug works.

To precisely define those links, the team has screened 498 drugs against 1,000 cancer cell lines, representing 30 different types of cancer. The drugs include licensed treatments and experimental compounds.

To undertake the hundreds of thousands of experiments needed, the researchers miniaturised the assays and developed sophisticated robotics, using acoustic dispensing to 'jump' droplets of drugs onto the cells.

The result is a unique, online, open-access resource for the research and medical communities. The data can be used to optimise the clinical application of cancer drugs as well as the design of clinical trials. The previous releases of data have enabled discoveries that led to drug trials of PARP inhibitors in childhood bone cancer and directly contributed to drug development. So far, the resource has powered 70 research studies across the globe. Working with Massachusetts General Hospital Cancer Center and as part of the Cancer DepMap project, the researchers continue to expand their work to understand how a cancer's underlying genomic landscape influences its response to treatments with the ultimate aim of targeting treatments to individual tumours.

Link
Genomics of Drug Sensitivity in Cancer resource – **cancerrxgene.org**

6

CRISPR catches out critical cancer changes

Researchers from the Sanger Institute, Open Targets and their collaborators used CRISPR-Cas9 in a large-scale analysis of gene fusions to reveal which ones are critical for the growth of cancer cells. As a result, the researchers identified a new drug target for multiple cancers.

Gene fusions, caused by the abnormal joining of two otherwise different genes, play an important role in cancer. As a cancer's genome becomes increasingly dysregulated, gene fusions are common. Identifying gene fusions aids both cancer diagnosis and prognosis, and they are the targets of some of the latest cancer treatments.

In the first large-scale analysis of cancer gene fusions, researchers at the Sanger Institute, EMBL-EBI, Open Targets and GSK analysed more than 8,000 gene fusions in over 1,000 human cancer cell lines, from 43 different cancer types.

The team first analysed RNA-sequencing (RNA-seq) data to define the gene fusions present in the cell lines. They then used genome-wide CRISPR screens to uncover which gene fusions are critical for each cell line's survival and tested the cell lines against more than 350 cancer drugs.

The results, published in *Nature Communications,* show that 90 per cent of gene fusions do not play an essential role in cancer – an important finding for clinicians when inferring causes of a cancer from the genome sequence of a patient's tumour.

The team also discovered a new fusion, *YAP1-MAML2.* It is essential for the progression of multiple cancer types, including some brain and ovarian cancers, and so represents a new treatment target. The researchers also found fusions involving the genes *RAF1, BRD4* and *ROS1,* in cancer types previously unknown to harbour them – pancreatic, breast and lung. This opens up the possibility of re-purposing existing treatments designed for other cancer types for patients with these particular gene fusions.

The collaboration between researchers at the Sanger Institute, EMBL-EBI and GSK, the Open Targets partners, bolsters the translation of these research results into new treatments.

Reference
Gabriele, P. *et al.* Functional linkage of gene fusions to cancer cell fitness assessed by pharmacological and CRISPR/Cas9 screening. *Nature Communications* 2019; **10**: 2198.

Stories 4-6

cancer dependency map

Cancer Dependency Map
The Sanger Institute and the Broad Institute are producing cancer dependency maps that will enable cancer researchers around the world to discover new drug targets and precisely match an individual's cancer mutations to specific vulnerabilities.

Only **10%** of gene fusions play a key role in cancer

7

Clarifying mutations in the driving seat

A clearer picture of how DNA changes lead to cancer has emerged, following the most comprehensive evaluation of non-coding driver mutations.

Discovery of cancer drivers, genetic mutations that drive a cancer to grow, has largely focused on genes that code for proteins. To investigate driver mutations in the 98 per cent of the genome that doesn't code for proteins, researchers at the Sanger Institute and the Broad Institute analysed 2,658 cancer genomes from the Pan-Cancer project.

The study is the most comprehensive evaluation of non-coding driver mutations to date, in terms of the number of methods employed, number of samples analysed, and the number of cancer, genome region and mutation types studied. Types of mutations studied include promoters, enhancers, which control gene activity, non-coding RNAs and other elements. The team combined the results of several driver-discovery methods to overcome limitations of any individual method.

The research, published in *Nature,* found several new cancer driver mutations in non-coding genes, including a point mutation in the *TP53* promoter region, which resulted in inactivation of the gene.

The overall conclusion, however, reaffirms the fact that the vast majority of cancer drivers occur in protein-coding regions of the human genome. Some non-coding drivers identified in previous studies were found to be the result of flaws in methodology. This unexpected finding has implications for future research and the development of new cancer treatments.

For cancer patients, these results mean that the vast majority of clinically-relevant mutations in a cancer are likely to be found in protein-coding sequences, which will simplify efforts for the clinical use of genome sequencing in cancer diagnosis and treatment.

Reference
Rheinbay E *et al.* Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* 2020; **578**: 102–111.

8

Cancer causing culprits will be caught by their DNA fingerprints

Sanger researchers together with international collaborators have created the most detailed list of mutational signatures in cancer to date. It will be used to understand how cancer develops and discover new causes of cancer, helping to inform public health strategies to prevent the disease.

The processes that change DNA leave patterns of damage, or mutational signatures, behind. Those factors that can lead to cancer, from UV light and tobacco smoke, to faulty DNA repair, each leave a distinct signature – their unique fingerprints can be seen in the DNA.

To build a comprehensive picture of mutational signatures in cancer, researchers from the Sanger Institute, the Broad Institute and their collaborators analysed almost every publicly available cancer genome sequence. The team studied data from 4,645 whole genome and 19,184 exome sequences, encompassing most types of cancer. They analysed a total of 84,729,690 genetic mutations, 10 times more than any previous study.

The team used multiple methods to classify and characterise the mutational signatures. They identified 49 single base substitution, 11 doublet base substitution, four clustered base substitution, and 17 small insertion and deletion signatures. Many had not been seen before. While the causes of many signatures remain unknown, this analysis provides a systematic perspective on the range of mutational processes contributing to cancer. The work is part of the Pan-Cancer project, published in *Nature.*

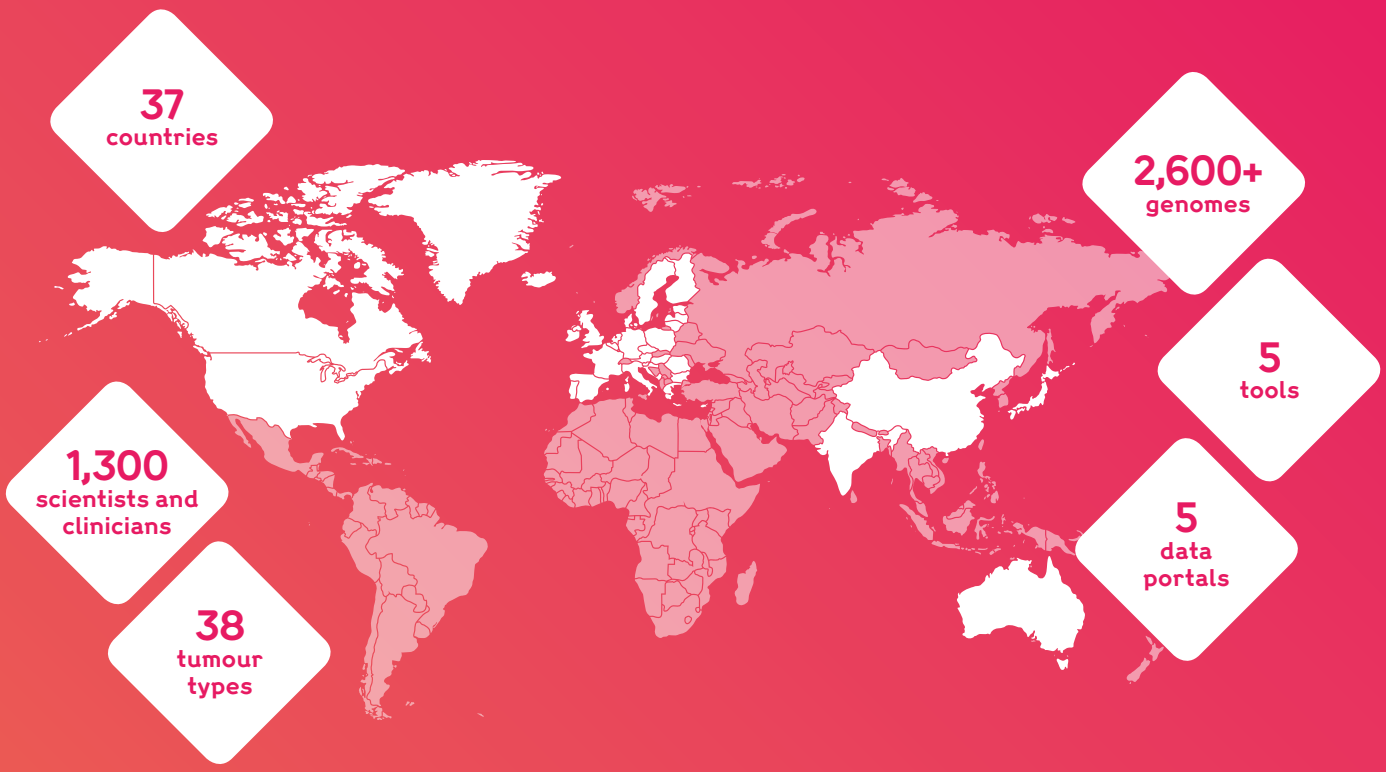
Another Pan-Cancer study by Sanger researchers discovered that larger, more complex rearrangements of DNA could also act as mutational signatures, and point towards causes of cancer.

It is likely that a substantial proportion of the naturally-occurring mutational signatures found in human cancer have now been described. This comprehensive repertoire provides a foundation for research into why cancer rates vary across location and time, the mutational processes operating in normal tissues, clinical and public health applications of signatures, and mechanistic understanding of the mutational processes underlying cancer.

References
Ludmil A *et al.* The repertoire of mutational signatures in human cancer. *Nature* 2020; **578**: 94–101.

Li Y *et al.* Patterns of somatic structural variation in human cancer genomes. *Nature* 2020; **578**: 112–121.

81
mutational signatures were identified by studying
84,729,690
mutations



9

International scientists come together for unprecedented exploration of cancer genomes

The global Pan-Cancer project has completed the most comprehensive study of whole cancer genomes. Researchers uncovered causes of previously unexplained cancers, new mechanisms of cancer development and have signposted new directions for diagnosis and treatment.

The decade-long Pan-Cancer project involved over 1,300 scientists and clinicians from 37 countries. Co-founded and co-led by researchers at the Sanger Institute, the collaboration of scientists analysed more than 2,600 genomes of 38 tumour types, including breast, bone and blood cancers.

Cancer is driven by genetic changes, and the rapid advancement of sequencing technologies means that documenting those changes across whole genomes is now possible. The Pan-Cancer project explored the entire genome, including regions that don't code for proteins, regulatory regions that control gene activity, non-coding Ribonucleic acids (RNAs), and large-scale structures.

To facilitate the comparison between diverse tumour types, all of the genome data were subjected to uniform processing, and passed rigorous quality control tests. The Pan-Cancer project advanced methods for analysing cancer genomes and the software for genome analysis, together with the raw genome sequencing data, are freely available to the scientific community.

Their research is published in 23 papers in *Nature* and affiliated journals. Several themes have emerged; firstly, the cancer genome is finite and knowable. Researchers can characterise every genetic change found in a cancer and all of the processes that have generated those mutations, including the order of key events during a cancer's life history.

Secondly, scientists are close to cataloguing all biological pathways involved in cancer – from changes in single DNA letters to the reorganisation of whole chromosomes. Thirdly, through a new method of dating mutations, researchers can identify those which occurred years before the tumour appeared. This could lead to earlier cancer detection. Finally, tumour types can be identified accurately according to the patterns of genetic changes seen across the genome, potentially aiding diagnosis.

This landmark project will accelerate the development of precision medicine for cancer – where patients are matched to therapies targeted at their tumours, using genomics.

Links
The project website **docs.icgc.org/pcawg** provides links to data resources for online browsing, analysis and download of data and results

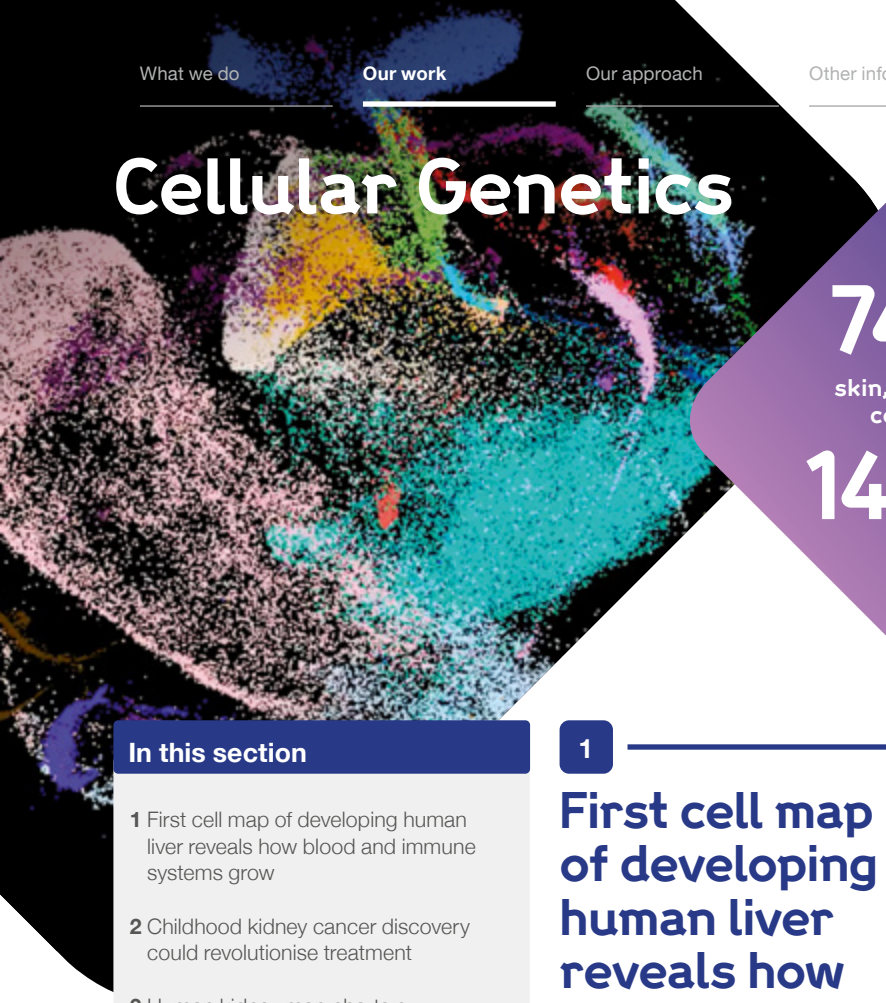
The research papers can be accessed via: **nature.com/collections/pcawg**

Stories 7-9

PCAWG

Pan-Cancer Analysis of Whole Genomes Study
The study aimed to examine the similarities and differences among the genomic and cellular alterations found across diverse tumour types.

Cellular Genetics



74,000
skin, kidney and yolk sac cells were studied

140,000
liver cells were studied

In this section

1

First cell map of developing human liver reveals how blood and immune systems grow

Sanger researchers and their colleagues have mapped the cellular landscape of the developing liver. The data will enhance understanding of how the blood and immune system develop, and support efforts to tackle diseases such as leukaemia.

It was not previously known precisely how the blood and immune systems develop in humans – a process known as haematopoiesis. In adults, bone marrow creates the blood and immune cells. But in early embryonic life, the yolk sac and liver play a major role.

To understand the process, Sanger scientists, together with teams from the Wellcome – MRC Cambridge Stem Cell Institute, Newcastle University, University of Cambridge and others, used single cell technology to analyse 140,000 developing liver cells and 74,000 skin, kidney and yolk sac cells.

They characterised individual cells isolated from the developing liver using morphology, and single-cell RNA sequencing to determine which genes were active. Haematopoietic cells in sections of developmental liver were tagged using heavy metal markers so researchers could map each cell to its location.

Reported in *Nature*, the team charted changes in the cellular landscape of the liver between the first and second trimesters of pregnancy. They discovered that during development, first generation haematopoietic stem cells stay in the liver. The next generation cells – progenitor cells – travel to other tissues. They mature in places such as the skin, where they develop into red blood cells to help meet the high oxygen demands of a growing foetus. These cells subsequently seed other peripheral tissues and finally, bone marrow.

The team also studied genes known to be involved in immune deficiencies to see which cells were expressing them. The result is a comprehensive, high-resolution resource – the developmental liver cell atlas. It provides crucial insights into how the blood and immune systems develop. It will also enhance research aiming to tackle diseases which can form during development, such as leukaemia.

In a related study, Sanger researchers collaborated with a team at the University of Edinburgh and identified new sub-types of cells that, when they interact, accelerate the scarring process in diseased livers. It is hoped that the findings, also published in *Nature*, will help researchers develop new treatments for liver diseases.

The work is part of the Human Cell Atlas initiative.

References

Popescu, D. *et al.* Decoding human fetal liver haematopoiesis. *Nature* 2019; **574**: 365–371.

Ramachandran, P. *et al.* Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* 2019; **575**: 512–518.

2

Childhood kidney cancer discovery could revolutionise treatment

The genetic origins of childhood kidney cancer have been uncovered by Sanger researchers. The findings indicate precursors to the cancer could be removed, rather than sacrificing the whole kidney, which would revolutionise treatment.

Wilms' tumour is the most common form of childhood kidney cancer, mainly affecting children under five. Most cases are curable by removing the kidney, but in 10 per cent of cases, both kidneys are affected. Without complete removal of the affected kidneys, there is a strong likelihood of the cancer returning.

To uncover the causes of Wilms' tumour, Sanger researchers worked together with colleagues at Addenbrooke's Hospital in Cambridge and Great Ormond Street Hospital in London.

Reporting in *Science*, the team compared the whole genome sequences of 66 tumour samples with 163 normal kidney samples. In two thirds of children with Wilms' tumour, they found DNA mutations associated with the disease in both the morphologically normal kidney tissue and the tumour tissue.

The team showed that the patches of genetically abnormal cells within the normal tissue, known as precursor lesions, had developed from a single 'rogue' cell. Phylogenetic analyses of the cells, a technique first pioneered by Sanger researchers in 2012, indicate that they evolve in early development, before the left and right kidney diverge.

More than 60 per cent of the rogue cells contained a mutation in the *h19* gene. The mutation suppresses *h19*'s activity, causing the cell to multiply more quickly than those around it. These mutated rogue cells are the likely precursors to Wilms' tumour.

Uncovering the genetic root of Wilms' tumour represents a shift in the understanding of childhood cancer – researchers didn't expect to find the root of childhood kidney cancer in normal-looking tissue. The results could help improve treatments, as it may be possible to remove precursor lesions, rather than sacrificing the whole kidney, to effectively treat the disease.

References

Coorens THH *et al.* Embryonal precursors of Wilms tumour. *Science* 2019; **366**: 1247–1251.

Young, M. *et al.* Single cell transcriptomes from human kidneys reveal the cellular identity of renal tumours. *Science* 2018; **361**: 6402: 594–599.

Mutations in *h19* were found in 60%+ of Wilms' tumour associated cells

Stories 1-5

HUMAN CELL ATLAS

The Human Cell Atlas

The Human Cell Atlas is a mission to create comprehensive reference maps of all human cells—the fundamental units of life—as a basis for both understanding human health and diagnosing, monitoring, and treating disease. The initiative is coordinated by researchers at the Sanger Institute, the Broad Institute in the US, and the Riken and Karolinska Institutes in Japan and Sweden.
humancellatlas.org

3

Human kidney map charts our growing immune defence

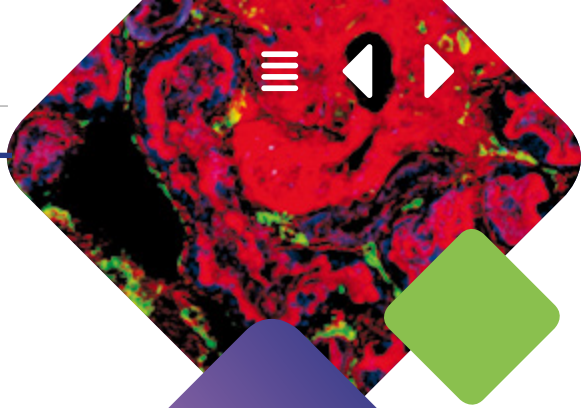
Sanger researchers and their collaborators mapped 70,000 individual cells to create the first cell atlas of the human kidney. The research reveals zones of immune cells in the kidney and the map will help scientists understand chronic kidney disease and why transplants are rejected.

Kidneys filter the blood, removing waste and excess water, and maintain the levels of salts, minerals and water that other organs need to function. Worldwide, 850 million people are affected by chronic kidney disease – a condition where kidneys gradually lose their ability to function. It can progress to kidney failure, for which dialysis or transplant are the only treatment options. The immune system plays a critical role in responding to kidney tissue damage, but very little is known about how it works.

To map the position, development and functions of immune cells that reside in the kidney, researchers studied nearly 70,000 individual kidney cells from developmental, child and adult kidney tissue. The research, conducted by teams from the Sanger Institute, University of Cambridge and Newcastle University, is part of the Human Cell Atlas – a project to map every cell type in the human body.

Cells were assessed with single-cell RNA sequencing, to identify which genes were active. The team mapped the cells over time and space, to understand the development and organisation of the kidney's immune system. They discovered that the very earliest cells that populate the developing kidney are macrophages – large white blood cells that consume pathogens – that remain in the kidney as we grow older.

The team were able to chart which types of immune cells were present in particular zones of the kidney. The researchers discovered that antimicrobial immunity is spatially zonated in the adult kidney, with a strong 'defence zone' at the base, near where urine leaves the kidney, which fights against urinary tract infections. There were few active immune cells in the developing kidney, which aligns with the view that a developing baby is relatively sterile.



850m

people worldwide have kidney disease

The findings, published in *Science*, will help scientists understand the immune system in the kidneys, what happens when tissue is damaged or infection occurs, how this can lead to chronic kidney disease, and why kidney transplants are rejected.

Reference
Stewart, B. *et al*. Spatio-temporal immune zonation of the human kidney. *Science* 2019; **365**: 1461–1466.

4

Chan Zuckerberg Initiative boosts Human Cell Atlas research

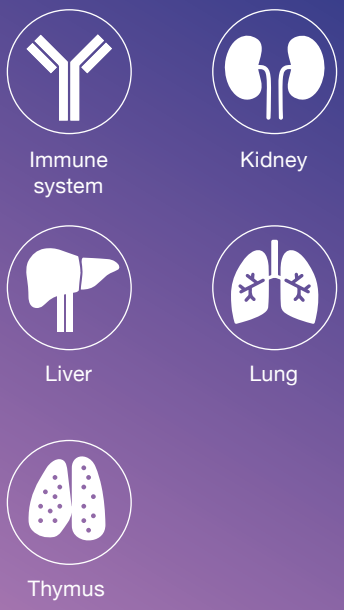
Sanger researchers have received funding from the Chan Zuckerberg Initiative (CZI) for five collaborative projects supporting the Human Cell Atlas (HCA).

The HCA is a global initiative to map every cell type in the human body. In 2019, the Chan Zuckerberg Initiative funded 38 new collaborative projects to support the HCA, with grants totalling $68 million.

The grants support networks of scientists from different disciplines who study a variety of human organs. Bringing together scientists, computational biologists, software engineers, and physicians, these collaborative groups will work towards building an atlas of the whole body – mapping the types, location and functions of cells in unprecedented detail.

The funding supports five projects at the Sanger Institute to investigate specific tissues in the thymus, lung, liver, kidney and immune system. The aim is to understand the tissues in health and disease. All the data from the research will be made available via the HCA Data Coordination Platform, a unified resource that will enable free data sharing across researchers and research institutes.

The Sanger Institute is involved in five HCA projects funded by CZI



5

First lung map uncovers new insights into asthma

Researchers from the Sanger Institute, Open Targets and their collaborators mapped the location, type, and activity of cells in the lungs, uncovering a new cell state responsible for creating mucus. The results could help pinpoint new drug targets for asthma.

Asthma, a common lung condition that makes breathing difficult and triggers coughing, wheezing and shortness of breath, is caused by swelling of the tubes that carry air to the lungs. It affected more than 350 million people worldwide in 2015, and at least 5 million people in the UK are currently receiving treatment.

As part of the Human Cell Atlas (HCA) project to map every cell type in the body, researchers studied more than 36,000 individual lung and nasal cells from people with and without asthma. Using single-cell sequencing technology, they analysed the gene activity in each cell. The study revealed clear differences between asthmatic and normal lungs.

A symptom of asthma is an overproduction of mucus. The team discovered a previously unknown mucus-creating cell state – the mucociliated state – in asthmatic lungs.

The researchers also revealed that asthmatic lungs had many more inflammatory Th2 cells than non-asthmatic lungs. These cells sent the vast majority of cellular signals in the lungs, dominating communication pathways. In normal lungs, a broad range of cells send communication signals to keep the airways functioning well. The team's finding highlights the importance of Th2 cells and inflammatory signals in asthma, and that knowledge could help inform drug development.

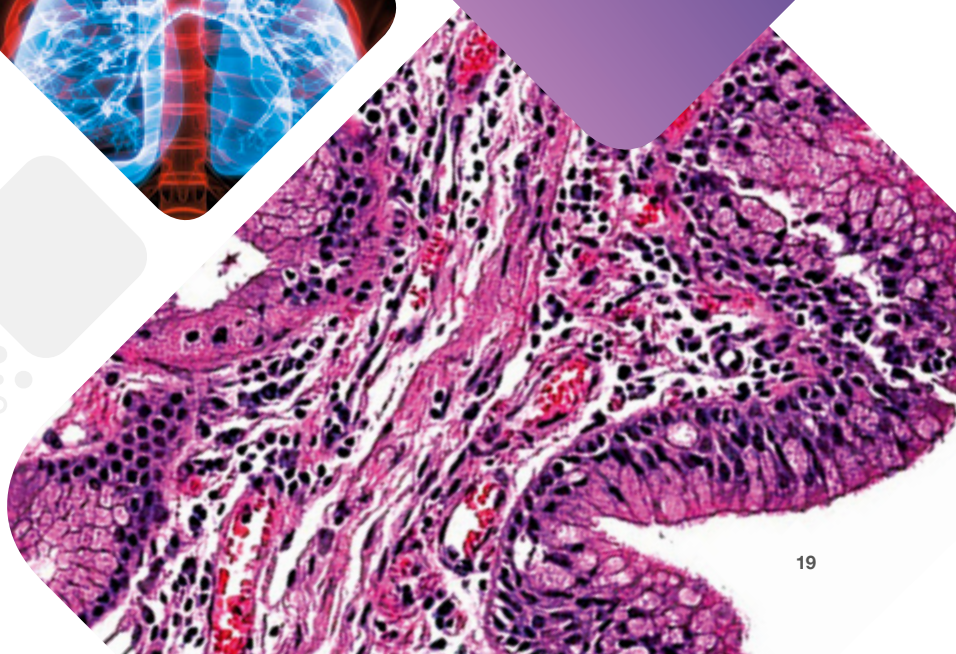
The research from the Sanger Institute, University Medical Center Groningen, Open Targets, GSK and others was published in *Nature Medicine*. The large-scale data set is openly accessible to researchers worldwide, providing an important resource for research into lungs, asthma, and potential new treatments.

Reference

Vieira Braga, F. *et al.* A cellular census of human lungs identifies novel cell states in health and in asthma. *Nature Medicine* 2019; **25**: 1153–1163.

36,000+

individual lung and nasal cells from people with and without asthma were studied



6

Scientists hone in on DNA differences behind immune diseases

Discovery that early 'switch-on' of certain immune cells triggers inflammatory diseases could lead to new drug targets for conditions like asthma.

Immune diseases, including arthritis, asthma, multiple sclerosis and inflammatory bowel disease, are caused by immune cells mistakenly targeting the body's own cells, rather than an invading pathogen, and causing inflammation. These conditions affect millions of people worldwide, yet it is not known what triggers the immune system to act in this way, or which cell types are involved.

Previous research has found thousands of genetic changes, or variants, that are more common in patients with immune diseases than in healthy people.

The variants are frequently in regions of the genome that control gene activity in immune cells, specifically T cells and macrophages, but the variants' roles in disease remain poorly understood.

To understand the significance of genetic variants, researchers from the Sanger Institute, together with GSK and Biogen within the Open Targets collaboration, stimulated T cells and macrophages from healthy donors. They included different combinations of inflammatory molecules, called cytokines, with the activation, to mimic 55 different cellular states.

Using advanced DNA sequencing methods, the team looked at which parts of the genome were active in three types of immune cells from healthy volunteers, and cross-checked these positions against all the genetic variants implicated in different immune diseases.

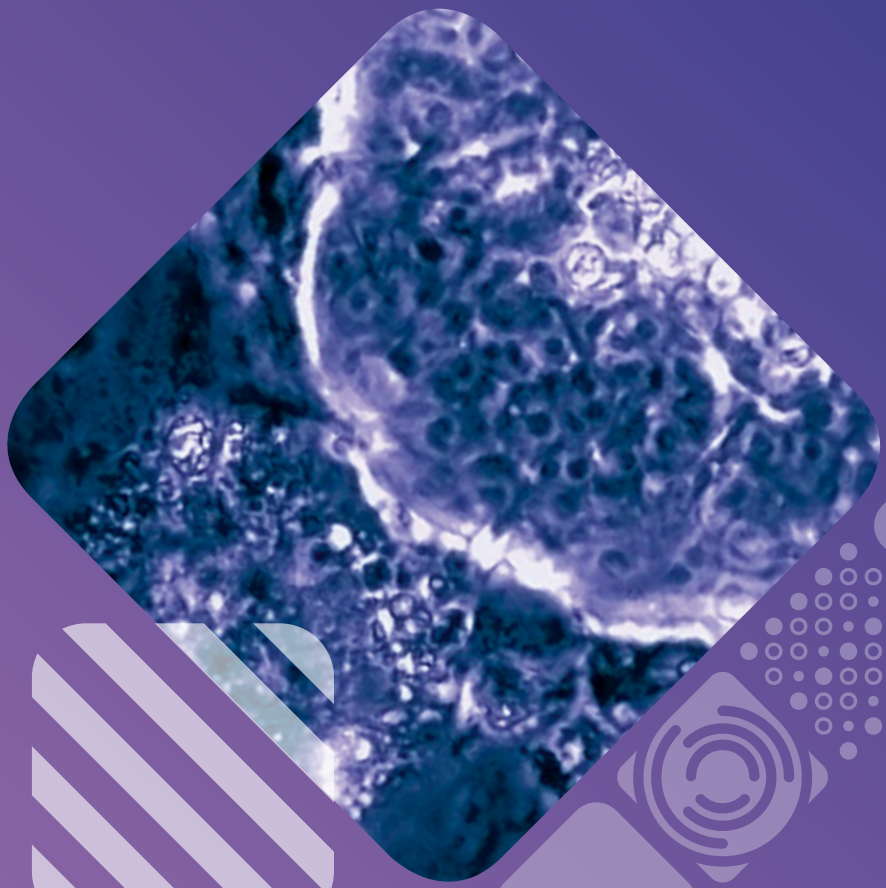
To enable this complex analysis, they developed a new computational method, called CHEERS. Openly available, this resource could also be used to find links between genetic variation and mechanisms for other complex diseases.

One particular cell type and its state – early activation of memory T cells – had the most active DNA across the same regions as the genetic variants implicated in immune diseases. Surprisingly, the research showed that the cytokines generally only had subtle effects on the DNA activity, and played a lesser role in most of the diseases studied.

The results point towards the initial activation of these memory T cells being important in disease development. The findings, published in *Nature Genetics*, will enable researchers to study in great detail how the 'switching on' of T cells and other immune cells is regulated and discover genes and pathways that could be used as new drug targets for immune conditions.

Reference
Soskic B. *et al*. Chromatin activity at GWAS loci identifies T cell states driving complex immune disease. *Nature Genetics* 2019; **51**: 1486–1493.

Study mimicked **55** different cellular states



7

Major stem cell discovery boosts research into development and regenerative medicine

Scientists from the University of Hong Kong, Sanger Institute and their collaborators have created human and pig Expanded Potential Stem Cells (EPSCs) – stem cells that look like the very first cells in the embryo and can develop into any type of cell. The achievement presents a unique cellular platform for research into biotechnology and regenerative medicine.

Stem cells have the ability to develop into other cell types, and existing stem cell lines – characterised populations of stem cells maintained in laboratories – are already extremely useful for research into development and disease. However, the types of stem cell lines currently available have limitations; they are not usually capable of producing extra-embryonic cells including the placenta or yolk sac.

Teams of researchers from LKS Faculty of Medicine of the University of Hong Kong, the Sanger Institute, and the Friedrich-Loeffler-Institut in Germany have previously created EPSCs in mice. They achieved this by inhibiting the critical molecular pathways that predispose the cells' differentiation. The EPSCs have the features of the very first cells in the developing embryo, and can develop into any type of cell, including placental cells.

Using the same technology, the researchers have now derived stem cells from early pig embryos, creating porcine stem cells – something that had not previously been achieved. Domestic pigs have great potential for biomedical research due to their genetic and anatomical similarities to humans, including comparable organ sizes.

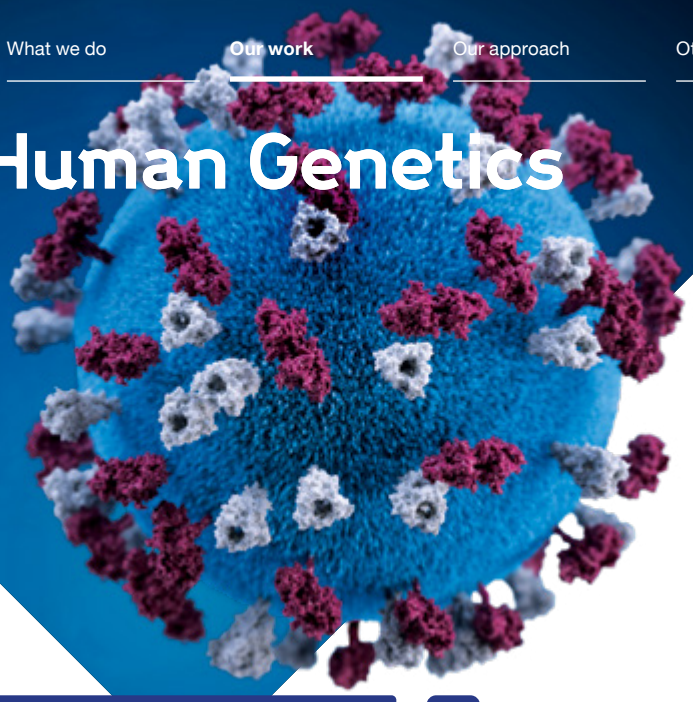
Under similar conditions, the team also used human stem cells to create EPSCs that display the molecular and functional attributes reminiscent of porcine EPSCs. The pathway-inhibition method opens an avenue for generating mammalian pluripotent stem cells.

The research, published in *Nature Cell Biology*, offers incredible potential for studying human development and regenerative medicine.

Reference

Gao, X. *et.al.* Establishment of porcine and human expanded potential stem cells. *Nature Cell Biology* 2019; **21**: 687–699.

Human Genetics



Measles causes
100,000+
deaths per year worldwide

In this section

After infection some children were immune suppressed for up to
5
years

1

Measles infection wipes immune system's memory

Sanger scientists have found how measles causes long-term damage to the immune system and leaves people vulnerable to other infections, emphasising the importance of vaccination.

Measles is highly contagious, causing 100,000 deaths per year worldwide in unvaccinated communities. A measles infection weakens the immune system and leaves people much more susceptible to other infectious diseases. This immunosuppression can last for months or years after the infection has passed.

To understand the impact of measles on the immune system, Sanger scientists, together with collaborators at the University of Amsterdam, studied a group of non-vaccinated people in the Netherlands.

Blood samples were taken from 26 children before and after a natural measles outbreak. The researchers sequenced and analysed specific DNA sequences from immune B cells. These B cell receptor (BCR) genes are responsible for producing antibodies which aid the immune system to recognise different pathogens.

The team discovered that a person's specific memory B cells, which swiftly tackle infections that have been previously encountered, were depleted after a measles infection. This is despite the fact that total blood cell counts – the measurement taken in clinics – had recovered. The depletion of memory B cells had not previously been seen in measles patients and this would leave the children vulnerable to a whole range of infectious diseases they had previously been immune to.

The researchers discovered that the measles virus also affects naïve B cells, which are those yet to be exposed to an infection. After infection with measles, the naïve B cell population was reset to an immature 'baby-like' state, able to produce only a limited range of antibodies. This would leave the children with a reduced ability to fight any new infections. The team's work provides the explanation for 'immune amnesia' caused by measles.

The study, published in *Science Immunology*, has huge implications for vaccination and public health. It shows that not only does measles vaccination protect people from measles, it also protects from other infectious diseases.

Reference

Petrova V N *et al.* Incomplete genetic reconstitution of B cell pools contributes to prolonged immunosuppression after measles. *Science Immunology* 2019; **4**: eaay6125.

2

Out-of-Africa and back again: ancient human migrations

Sanger scientists resolve a long-running debate about out-of-Africa migration after genomic analysis of three Nigerian Y chromosomes.

According to current models, present-day humans outside Africa descend mainly from a single migration out of the African continent some 50,000–70,000 years ago. The estimation is supported by genetic data from genome-wide, mitochondrial DNA, and Y-chromosomal analyses, but many details remain unclear, including the movements of the male-specific Y chromosome lineages at the time.

To investigate, Sanger researchers studied a rare African Y-chromosomal lineage. The Y chromosome, passed directly from father to son without genetic recombination, gives researchers a clear picture of ancestry and the rate of DNA change over time.

Three Nigerian men representing this lineage were first studied in 2003. Since then, genome sequencing technology has advanced, and so the team sequenced the genomes again. They analysed the genome sequences against the globally-recognised Y chromosome family tree.

Their analysis, published in *Genetics*, shows that the three Y chromosomes represent a new lineage – termed 'D0', which the team included in an update to the human ancestral family tree. These results provide key information to add to the debate about the factors that influence the geographical locations of the Y lineages 50,000–100,000 years ago and the mode of expansion out of Africa.

The researchers considered three scenarios for the expansion of Y lineages out of Africa, incorporating migration back to Africa where necessary, to explain present-day Y-lineage distributions. Considering published archaeological evidence for modern humans outside Africa, and mixing with Neanderthals, the most favoured scenario is an exit from Africa between 50,300–59,400 years ago.

This work resolves a long-running debate about Y-chromosomal out-of-Africa/back-to-Africa migrations, increasing our understanding of the movement of modern humans out of Africa. Sanger researchers demonstrate how DNA sequences from just three chromosomes can provide deep understandings and insights into human evolution.

Reference
Haber M *et al.* A Rare Deep-Rooting D0 African Y-Chromosomal Haplogroup and Its Implications for the Expansion of Modern Humans Out of Africa. *Genetics* 2019; **212**: 1421–1428.

3

Crusaders made love and war

Sanger researchers sequenced the genomes of 13th century skeletons discovered in Lebanon. The skeletons were found to be Crusaders who had families with local people in the Near East, and that they fought and died side by side with their sons.

During the medieval period, hundreds of thousands of Europeans migrated to the Near East to fight in the Crusades. These religious wars, which took place between 1095 and 1291, were led by Christian nobility who tried to claim the Near East. But historical records lack details of the ordinary soldiers who travelled to, lived and died there.

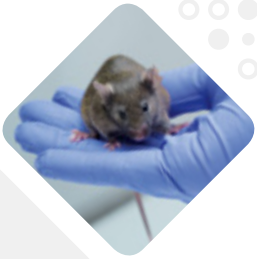
To study these important historical events, Sanger scientists worked closely with archaeologists at the Sidon excavation site in Lebanon who had recently uncovered 25 skeletons, believed to be Crusaders killed in battle in the 13th century. The bones of nine skeletons were transferred to a specialist ancient DNA laboratory in Cambridge.

Small portions of the surviving 800-year-old DNA were extracted from the temporal bone in the skulls. The bodies had been burned and buried in a warm and humid climate, where DNA degrades quickly. Only recent advances in DNA extraction and sequencing technology made the study possible.

Genetic analysis showed all of the individuals were males; some were Western Europeans from diverse origins, some were locals (genetically indistinguishable from present-day Lebanese), and two individuals were a mixture of European and Near Eastern ancestries. This provides direct evidence that the Crusaders mixed and had families with the local population, and that their sons joined them in battle.

These ancestral mixtures appear to have had no genetic consequences on Lebanese people today, as genetic signals of admixture with Europeans were quickly lost over time. The study, published in the *American Journal of Human Genetics*, shows how collaborations between scientists and archaeologists and genomic analysis of ancient DNA can provide exceptional historical insights.

Reference
Haber M *et al.* A Transient Pulse of Genetic Admixture from the Crusaders in the Near East Identified from Ancient Genome Sequences. *American Journal of Human Genetics* 2019; **104**: 977–984.



CNOT1
gene mutation identified in people lacking a pancreas

4

Rare genetic change gives clues to pancreas and brain development

Understanding how the pancreas forms could aid research into treatments for type 1 diabetes.

Very rarely, babies are born without a pancreas – a condition called pancreatic agenesis. Forming part of the digestive system, the pancreas makes various hormones, including insulin that controls the amount of sugar in the blood. To survive, those with pancreatic agenesis need immediate and lifelong treatment with insulin and other hormones.

Sanger researchers, together with a team at the University of Exeter, studied 107 people with pancreatic agenesis. In those people where the genetic cause of their condition was unknown, the researchers sequenced the protein coding regions of the genome to uncover the genetic changes responsible. They discovered an identical mutation in the *CNOT1* gene of three people. The three individuals had similar clinical features, including holoprosencephaly – a condition where the forebrain fails to develop into two hemispheres.

To investigate further, the team bred mice to have the same mutation in the equivalent mouse gene – *Cnot1*. As a result, the mouse embryos had a much smaller pancreas than usual, directly linking *Cnot1* with pancreas development. The researchers also saw changes in the brain development of the mice, including structural defects. This is the first time *CNOT1* has been identified as important in pancreatic and neurological development.

The *CNOT1* gene had previously been implicated in keeping human and mice embryonic stem cells in a pluripotent state, meaning they are capable of developing into any type of cell. In the mice with the *Cnot1* mutation, the researchers found the activity of a key developmental factor had changed, which kept stem cells as stem cells and prevented them from developing into pancreas cells.

The work, published in the *American Journal of Human Genetics*, adds crucial understanding to how the pancreas forms. The results could also help researchers develop replacement insulin-producing cells to treat patients with type 1 diabetes in the future.

Reference

De Franco E, Watson R. *et al.* A specific *CNOT1* mutation results in a novel syndrome of pancreatic agenesis and holoprosencephaly through impaired pancreatic and neurological development. *American Journal of Human Genetics* 2019; **104**: 985–989.

5

‘Jumping genes’ can cause rare developmental disorders in children

After analysing the role of ‘jumping genes’, three more children with rare developmental diseases received a diagnosis for the first time through the Deciphering Developmental Disorders (DDD) study.

The DDD study began in 2010, and recruited nearly 14,000 children with rare, undiagnosed, genetic conditions. Since then, genetic analyses by Sanger scientists, NHS Clinical Geneticists and researchers around the world have provided diagnoses for over 4,500 of the children taking part. Dozens of new conditions have been identified. A genetic diagnosis gives families improved access to support networks and, in some cases, opportunities to participate in clinical research in the future.

Mobile genetic elements, also known as mobile elements, or MEs, are pieces of DNA that can ‘jump’ from place to place in the genome via a process called retrotransposition. MEs can cause disease if they land in a gene and disrupt its function, but it has not been previously known if MEs could cause developmental disorders.

To uncover the potential role of MEs in developmental disorders, Sanger researchers, together with colleagues from NHS Regional Genetics services, studied the protein-coding sequences from 9,738 parent and child trios (28,132 individuals) of DDD participants.

In this latest study, published in *Nature Communications*, the Sanger team developed and validated an analytical process for the rapid assessment of MEs in genome data. They used the software to search for MEs in DDD patients and identified four children with MEs that were the likely cause of their symptoms, providing a new diagnosis for three of them.

While events like MEs are rare, they represent an important class of genetic variation to investigate in clinical assessments. The software is openly available and could be used to help diagnose other patient groups.

Reference
Gardner E *et al.* Contribution of retrotransposition to developmental disorders. *Nature Communications* 2019; **10**: 4630.

14,000
children were studied

4,500
have received a genetic diagnosis

Parasites and Microbes

100+
new species have been identified

In this section

1

How you are born determines your gut bacteria

In the largest ever study of new-borns' gut bacteria, Sanger researchers uncovered the different species that colonise a baby's gut at birth.

The gut microbiome – the genetic material of all the microbes living in the gut – is vital for human health. Imbalances are thought to contribute to conditions such as obesity, asthma and Crohn's disease. The very first micro-organisms to colonise a baby's gut could influence the subsequent development of its microbiome and immune system.

To understand where and how babies acquire their gut bacteria, and the subsequent development of their microbiomes, Sanger researchers studied 596 babies and 178 mothers.

Reporting in *Nature*, genomic analysis of 1,679 faecal samples revealed delivery method had the biggest impact on gut microbiota during the neonatal period. Babies born vaginally acquired their mother's gut microbes, whereas babies born via C-section had more hospital bacteria, including pathogenic and antibiotic resistant strains. However, these differences appear to largely even out after a year.

This type of research has previously been restricted by technical limitations of sequencing just a small portion of DNA (16S rRNA gene profiling), small sample size or limited sampling during the first month of life. Taking a whole genome sequencing approach and longitudinal sampling allowed the researchers to identify the bacteria present. The study draws on the team's recent work to isolate and sequence the genomes of gut bacteria, where they identified over 100 new species.

The consequences of disruptions to early-life microbiota, and the presence of immunogenic pathogens during this critical window of immune development, are yet to be uncovered. The study provides insight into the source and species of a baby's gut microbiome, and future work will determine if this early stage has a role to play in health.

References

Shao *et al.* Stunted gut microbiota and increased pathogen colonisation associated with caesarean birth. *Nature* 2019; **574**: 117–121.

Forster *et al.* Human Gastrointestinal Bacteria Genome and Culture Collection. *Nature Biotechnology* 2019; **37**: 186–192.

2

Diarrhoea causing bacteria adapted to spread in hospitals

Sanger researchers have discovered that the gut-infecting bacterium, *Clostridium difficile* is evolving into two separate species. One of these species has adapted to thrive on sugary diets and evade disinfectants in order to spread in hospitals.

Antibiotic treatments wipe out normal gut bacteria, leaving people vulnerable to *C. difficile* infection. The infection is difficult to treat, and can cause bowel inflammation and severe diarrhoea. *C. difficile* bacteria are the leading cause of antibiotic-associated diarrhoea worldwide.

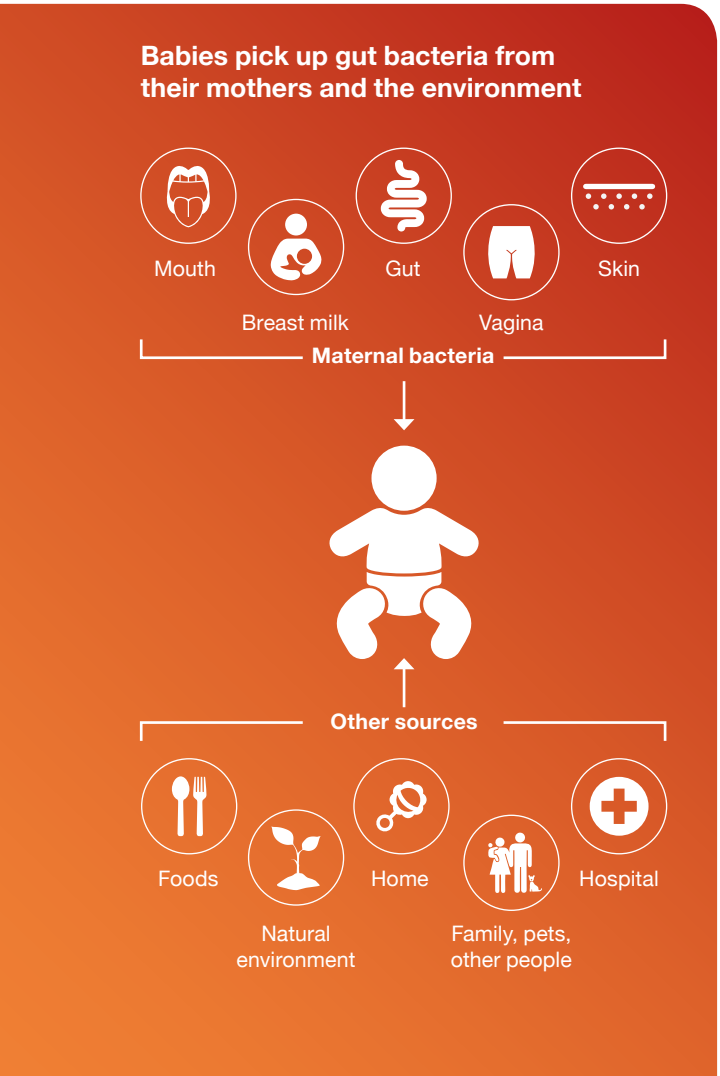
Researchers at the Sanger Institute, the London School of Hygiene & Tropical Medicine and their collaborators undertook the largest ever genomic study of *C. difficile*. They collected 906 strains of bacteria from humans, animals and the environment, across 33 countries around the world. Genomic sequencing and analysis showed that *C. difficile* is currently evolving into two separate species.

Reporting in *Nature Genetics*, the team identified genetic changes in the emerging species that allow it to thrive on Western sugar-rich diets. It has also evolved differences in genes involved in forming spores, which help the bacterium evade common hospital disinfectants, allowing it to remain on surfaces. It has become specialised to spread in healthcare environments.

Researchers estimated this species first appeared about 76,000 years ago; it was primed to take advantage of modern healthcare practices and diets before hospitals even existed. The bacterium has continued to evolve and is now thriving, accounting for over two thirds of healthcare *C. difficile* infections.

The findings give new insights into bacterial evolution and demonstrate how human behaviour shapes species. The team also demonstrates the importance of genomic surveillance of bacteria. Their work could help shed light on how other dangerous pathogens evolve and could help inform infection control in hospitals.

Reference
Browne H *et al.* Adaptation of host transmission cycle during *Clostridium difficile* speciation. *Nature Genetics* 2019; **51**: 1315–1320.



3

Combination of common drugs could be used to treat superbugs, like MRSA

Sanger researchers have revealed many MRSA strains are susceptible to the widely-available antibiotic, penicillin, when combined with another drug – clavulanic acid.

Antibiotic resistance poses a severe threat to the future of medicine. Methicillin-resistant *Staphylococcus aureus* bacterium – MRSA – is already a serious problem. The superbug has acquired widespread resistance to β-lactam antibiotics which include penicillin and its derivative, methicillin, the most clinically important and widely-used group of antibiotics.

S. aureus has become resistant to β-lactam antibiotics by acquiring a gene encoding a β-lactamase, or, in the case or MRSA, a gene encoding penicillin-binding protein 2a (PBP2a).

Previously, scientists in Cambridge identified an isolate of MRSA that showed susceptibility to penicillin when combined with clavulanic acid. Clavulanic acid is a beta-lactamase inhibitor, already used to treat kidney infections during pregnancy.

Reporting in *Nature Microbiology* the team, together with scientists from the Sanger Institute and others from Denmark, Germany, Portugal, and the USA, sequenced the genome of this particular MRSA strain to identify which genes make it susceptible to the combination of drugs. They found several mutations centred around the gene that codes for the penicillin-binding protein 2a (PBP2a).

PBP2a is crucial to MRSA strains, enabling them to grow in the presence of penicillin and related antibiotics. Two of the mutations reduced PBP2a expression, whereas two other mutations increased the ability of PBP2a to bind to penicillin in the presence of clavulanic acid. As a result, a combination of penicillin and clavulanic acid could overcome the resistance to penicillin in some MRSA strains. The researchers used the drugs to successfully treat MRSA infections in moth larvae and then mice.

The team then sequenced the genomes from a diverse collection of MRSA strains. They found a significant number of MRSA strains – including the dominant strain in the USA – contained both mutations, making them vulnerable to the combined treatment.

This makes it possible that one of the most widespread strains of MRSA could be treated by a combination of drugs already licensed for use. The next step is to prepare clinical trials in humans.

Reference

Harrison, EM *et al*. Genomic identification of cryptic susceptibility to penicillins and β-lactamase inhibitors in methicillin-resistant *Staphylococcus aureus*. *Nature Microbiology* 2019; **4**: 1680–1691.

4

Monitoring dangerous bacteria around the world to stop the spread and save lives

Sanger scientists use genomic surveillance to study the global spread of dangerous bacterial pathogens and improve control and treatment practices.

During a Europe-wide survey, researchers analysed the genomes of almost 2,000 samples of *Klebsiella pneumoniae*, an opportunistic pathogen that can cause respiratory and bloodstream infections. They found that antibiotic-resistant strains are spreading through hospitals in Europe. Some strains are resistant to the carbapenem antibiotics that represent the last line of defence in treating infections.

In 2007, an estimated 341 people died from infections with carbapenem-resistant *K. pneumoniae*; by 2015 the number of deaths was 2,094. The increase is down to the fact that once carbapenems are no longer effective, there are few other treatment options left. Infants, the elderly and immuno-compromised individuals are particularly at risk.

Researchers at the Centre for Genomic Pathogen Surveillance, based at the Sanger Institute, and their partners, published their findings in *Nature Microbiology*. The results will inform public health efforts to control the spread of these infections in hospitals across Europe.

In a separate global genomic survey, scientists from the Sanger Institute, Emory University and the US Centers for Disease Control and Prevention sequenced DNA from nearly 20,000 samples of *Streptococcus pneumoniae*, the most common bacterial cause of pneumonia. The study is the first time *S. pneumoniae* has been assessed on a global scale.

Pneumonia is responsible for hundreds of thousands of deaths each year. Many countries have introduced the pneumococcal conjugate vaccine, and while highly effective against some strains, pneumococcal pneumonia rates remain high.

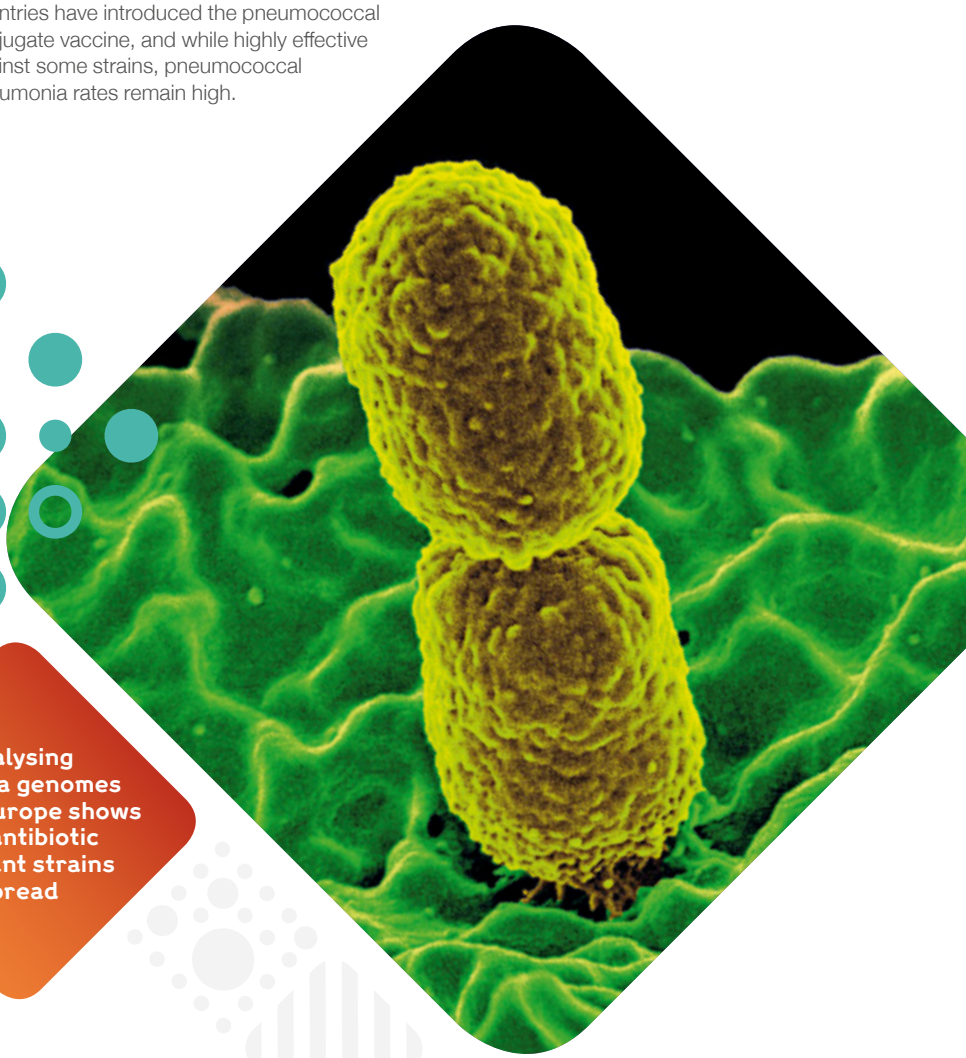
Samples were analysed from both before and after the introduction of the vaccine, allowing the team to see how the bacterium has evolved in response. The work will help predict which strains are likely to cause a threat in the future, and it will be vital for researchers reformulating the next generation of vaccines and developing vaccine strategies worldwide.

Both studies highlight the importance of ongoing global genomic surveillance to inform practices and save lives.

References

David S *et al*. Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread. *Nature Microbiology* 2019; **4**: 1919–1929.

Lo S *et al*. Pneumococcal Lineages Associated with Serotype Replacement and Antibiotic Resistance in Childhood Invasive Pneumococcal Disease in the Post-PCV13 Era: An international whole genome sequencing study. *The Lancet Infectious Diseases* 2019; **19**: 759–796.



Analysing bacteria genomes across Europe shows how antibiotic resistant strains spread

5

Genomics points to universal Streptococcus A vaccine

A decade-long genomic study of highly diverse, dangerous bacteria has revealed potential vaccine targets.

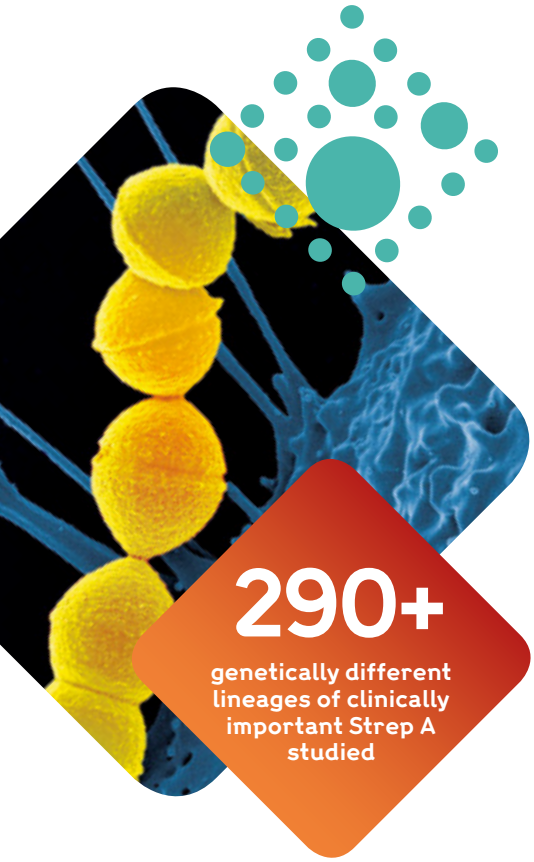
Group A *Streptococcus* bacteria, commonly known as Strep A, is an unjust killer. In low-income regions of the world it is thought to be responsible for more than half a million deaths per year, where it causes infections leading to rheumatic heart disease. Yet in high-income countries it is more likely to cause milder issues, such as a sore throat or impetigo.

To aid the search for a global vaccine, researchers from the Sanger Institute, the University of Cambridge and collaborators in Australia, sequenced and analysed the genomes of over 2,000 Group A *Streptococcus* samples from 22 countries.

Reporting in *Nature Genetics*, the team revealed the existence of more than 290 genetically different lineages of clinically important Strep A from countries spanning Africa, the Pacific, New Zealand and Australia.

Until now, little has been known about the genetic makeup of Strep A in areas with the greatest burden – most information has come from areas such as the UK and US. This new data showed that current leading Strep A vaccine candidates would be of little use in lower-income areas of the globe where Strep A is endemic.

However, the team found a number of molecular targets that were present in all of the bacterial strains, offering new targets to develop a global vaccine. In addition, further analysis of the genomic data may help researchers understand why the bacteria cause such differing symptoms and diseases in different regions of the globe.



Reference
Davies MR *et al.* Atlas of group A streptococcal vaccine candidates compiled using large scale comparative genomics. *Nature Genetics* 2019; **51**: 1035–1043.

6

Deadly drug-resistant Salmonella discovered by African surveillance

International collaboration and long-term monitoring has revealed the rise in dangerous strains of *Salmonella* bacteria in the Democratic Republic of Congo (DRC) that are now becoming resistant to the last-line-of-defence drug.

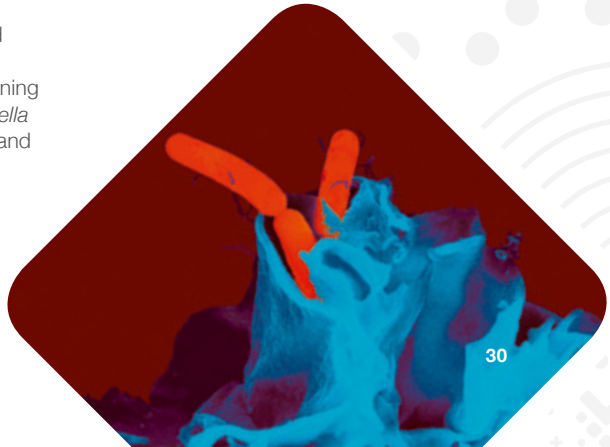
Every year invasive non-typhoidal *Salmonella* infections affect 3.4 million people, resulting in more than 680,000 deaths. The majority of these blood-borne infections are caused by *Salmonella* Typhimurium. It is a major health burden in sub-Saharan Africa, where it disproportionately affects children under five and adults with HIV.

To understand the disease, its evolution and spread, researchers from the Institut National de Recherche Biomédicale in the DRC and the Institute of Tropical Medicine in Antwerp established a hospital-based surveillance system that has been operating for the past 10 years. Working with scientists at the Sanger Institute and others, the teams analysed both the bacterium's physical characteristics and its genome sequence.

Reporting their findings in *Nature Communications*, the teams uncovered the emergence of a new, extensively drug-resistant form of *S.* Typhimurium. Laboratory analysis identified that this new sub-group, named ST313 lineage II.1, is resistant to all but one of the commonly available drugs in the DRC; one sample even showed reduced susceptibility to this final antibiotic. The findings also suggest that the bacteria are evolving to become more invasive.

To see whether genomic insights could help with effective healthcare planning, Sanger researchers used machine learning to look for DNA patterns in the *Salmonella* genome associated with invasiveness and drug resistance. The algorithm they developed successfully identified the changes that enabled the bacterium to be more dangerous and they hope that such approaches could predict the emergence and spread of drug-resistant strains in the future – guiding more effective interventions.

Reference
Van Puyvelde *S et al.* An African *Salmonella* Typhimurium ST313 sublineage with extensive drug-resistance and signatures of host adaptation. *Nature Communications* 2019; **10**: 4280.



7

World War One soldier's cholera revived and sequenced, 100 years on

Sanger researchers have read the genetic code of a century-old *Vibrio cholerae* bacterium, stored by the National Collection of Type Cultures, providing insights on bacterial evolution.

The bacterium was isolated during World War One (WWI) from a British soldier convalescing in Egypt. It is the oldest publicly available strain of these cholera-causing bacteria. As part of a collaboration to understand bacterial evolution, researchers revived the bacterium and sequenced its genome.

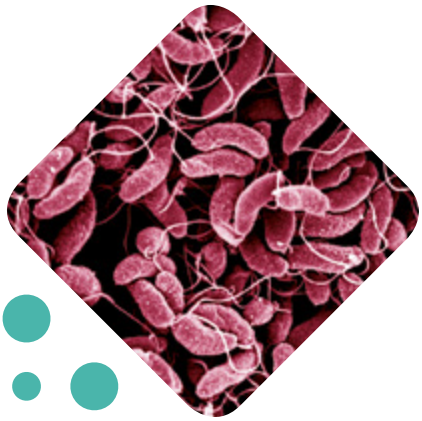
The results, published in *Proceedings of the Royal Society B*, show that this strain is a unique, non-toxigenic strain of *V. cholerae*. It was not capable of causing epidemic cholera, and was unrelated to the classical *V. cholerae* that caused the sixth pandemic and was active at the time. The genomic findings align with historical records of the soldier's illness.

The team also found that this strain of *V. cholerae* possessed a gene for ampicillin resistance, adding to evidence that bacteria possessed genes for antibiotic resistance long before antibiotics were developed. One explanation is that bacteria needed such genes to protect themselves against naturally-occurring chemicals.

Comparing historical bacterial species with their modern-day counterparts can give deep insights into the evolution of bacteria, and the roles they played in human history.

The bacterium from WWI was collected during a significant period of cholera history. The disease has killed millions since the 19th century and remains a global threat to public health today – up to 5,000 people die every year from the disease.

Reference
Dorman M *et al*. The history, genome and biology of NCTC 30: a non-pandemic *Vibrio cholerae* isolate from World War One. *Proceedings of the Royal Society B* 2019 DOI: 10.1098/rspb.2018.2025.



Ampicillin resistance gene existed before the drug had been developed

8

Resurrecting 50,000-year-old malaria gene shows jump from gorillas to humans

Sanger researchers have uncovered the molecular pathway that enabled the world's deadliest malaria parasite to jump from gorillas to humans. The research is important for understanding how pathogens are able to move between species.

Malaria's most deadly form is caused by *Plasmodium falciparum*. The species has arguably been responsible for more human deaths than any other disease.

P. falciparum is one of seven members of the sub-genus *Laverania*, which originated in African great apes. *P. falciparum* only infects humans, after switching host from gorillas around 50,000 years ago.

To investigate the origins of *P. falciparum*, researchers from the Sanger Institute and the University of Montpellier studied the genomes of all *Laverania* species. Analysis revealed a section of DNA, containing the *rh5* gene that had transferred from a gorilla parasite to the ancestor of *P. falciparum*.

The gene encodes the RH5 protein that binds to the basigin receptor on human red blood cells – a critical interaction for *P. falciparum* to infect humans. The team used ancestral sequence reconstruction to 'resurrect' the 50,000-year-old *rh5* DNA sequence and created synthetic copies. The resulting protein was able to bind to both gorilla and human basigin, providing the molecular explanation as to how the parasite switched host.

The study, published in *PLOS Biology*, also identified six differences between the current and ancestral RH5 DNA sequence in *P. falciparum*. One of the changes stops RH5 being able to bind to gorilla basigin, explaining how *P. falciparum* became restricted to humans.

In a separate study, published in *Nature Communications*, a second Sanger Institute team investigated proteins that interact with RH5, and found that a change in the *EBA165* gene may also have played a key role in the emergence of *P. falciparum* as a human-infective species.

References

Galaway F *et al*. Resurrection of the ancestral RH5 invasion ligand provides a molecular explanation for the origin of *P. falciparum* malaria in humans. *PLoS Biol* 2019; **17**: e3000490.

Proto WA *et al*. Adaptation of *Plasmodium falciparum* to humans involved the loss of an ape-specific erythrocyte invasion ligand. *Nature Communications* 2019; **10**: 4512.

Stories 8-12

Malaria

Malaria, caused by *Plasmodium* parasites, is a global killer. It is responsible for at least 400,000 deaths a year. Nearly 228 million people were infected in 2018, with children under five most at risk. Species of *Plasmodium* are rapidly and repeatedly becoming resistant to antimalarial drugs in many areas, meaning elimination efforts are under threat and new treatments are urgently needed.

Ancestral sequence reconstruction was used to 'resurrect' the

50,000

year-old RH5 DNA sequence

9

Success in controlling malaria at risk from spreading drug resistance

Regional populations of malaria parasites in Africa are sharing genetic material in all directions – including genes that can confer resistance to anti-malarial drugs. African scientists together with Sanger Institute collaborators provide crucial evidence as to how resistance to malaria drugs is developing across the continent.

In the first continent-wide genomic study of malaria in Africa, scientists have uncovered the genetic features and evolution of the deadly *Plasmodium falciparum* parasite, including how drug resistance spreads.

Over the last 15 years, the ongoing drive to eliminate malaria has seen worldwide deaths halve to 429,000 per year. But recently, the decline in prevalence has stalled and malaria remains a global problem. In 2015, 92 per cent of global malaria deaths were in Africa, with 74 per cent of these in children under five.

Researchers, led by those from the Plasmodium Diversity Network Africa, worked with the Sanger Institute to study genetic features of *P. falciparum* parasite populations from 15 African countries. The work is part of the global MalariaGEN data-sharing network, which aims to build surveillance capacity in low and middle income countries.

The findings, published in *Science*, show that *P. falciparum* parasites are present in genetically distinct regional populations, and these are consistent with human and mosquito population divergence.

Using single-nucleotide polymorphism (SNP) analysis on whole genome sequence data, researchers found that the populations are sharing genetic material in all directions – previously it was thought that the transfer of genes went from East to West Africa. Genes that confer resistance to antimalarial drugs are part of that transfer, with new types of drug resistance emerging in different parts of Africa.

Of most concern were genetic signatures detected on chromosome 12 in *P. falciparum* from Ghana and Malawi which could compromise the effectiveness of artemisinin-based combination therapies (ACTs). ACTs combine multiple antimalarial drugs in one treatment to overcome resistance to one or more individual drugs.

This information is crucial for understanding how resistance to malaria drugs is developing in Africa, and it can help inform control strategies across the continent. Ongoing, large-scale genomic surveillance is going to be vital to track the emergence and spread of drug resistant malaria.

References

Amambua-Ngwa A *et. al.* Major subpopulations of *Plasmodium falciparum* in sub-Saharan Africa. *Science* 2019; **365**: 813–816.

Miotto O *et al.* Genetic architecture of artemisinin-resistant *Plasmodium falciparum*. *Nature Genetics* 2015; **47**: 226–34.

Continent-wide study shows genetic material shared in all directions

10

New malaria drug targets identified in liver stage of life cycle

A huge collaborative project has identified essential malaria survival genes as drug targets, aimed at disrupting the parasite's invasion of the liver. The results open the possibility of new ways to tackle the disease, alongside the existing blood-stage drugs.

Most current anti-malarial drugs target the blood stage of the parasite's complicated life cycle, with very few targeting the liver stages. Liver-stage drugs hold the potential to attack malaria parasites that can lay dormant in the liver for years causing recurrent symptoms – such as *Plasmodium vivax*.

Using *Plasmodium berghei*, which infects mice, scientists from the Sanger Institute, the University of Bern, Switzerland and Umeå University, Sweden studied over 1,000 individual genes in the malaria parasites.

Building on their previous work, the team studied parasites which had individual genes that had been deleted and tagged using a molecular barcode. The team assessed the effect of each deletion on the parasite's behaviour throughout its life cycle, including in the liver stages. They cross-referenced the results with previously published data on metabolites and parasite gene activity.

This computational approach enabled the researchers to identify 461 genes that are essential for parasite transmission from mosquitoes, through the mouse liver, and back into the bloodstream of mice. The team then verified 20 of the findings in the laboratory – demonstrating the validity of their method. They discovered seven metabolic pathways that the parasite needs to infect the liver.

The study, published in *Cell*, is the first systematic study of gene function in the *Plasmodium* malaria parasite family.

Reference

Stanway R *et al.* Genome Scale Identification of Essential Metabolic Processes for Targeting the *Plasmodium* Liver Stage. *Cell* 2019; **179**: 1112–1128.

11

Malaria deconstructed: cell by cell, stage by stage

Sanger researchers have built the first Malaria Cell Atlas, giving unprecedented detail of how the parasite works at each stage of its complex life cycle.

The parasite that causes malaria is complex. It lives in both mammals and mosquitoes and goes through multiple stages in mammalian livers and red blood cells, and mosquito guts and salivary glands. The parasite itself is a single cell, which takes on many forms. Understanding which parasite genes are active at each of the different stages of the life cycle is key to developing much-needed antimalarial drugs, vaccines, and transmission blocking strategies.

Previous malaria studies used pooled samples of parasite cells to study gene activity, but this could mask potentially important differences, both between and within life stages, because it was not possible to ensure all the cells were at exactly the same point in the cycle.

To overcome this, Sanger scientists used advanced single-cell RNA sequencing technologies to study more than 17,000 individual parasites from across the life cycle. Not only did this provide an accurate picture of gene activity for each individual stage, it also allowed the gene activities of different parasites within the same stage to be compared, giving exceptional detail.

The function of 40 per cent of *Plasmodium* genes is unknown but the team were able to infer the roles for many, based on their pattern of activity across the life cycle and compared to the activity of genes whose function is known. These data could help researchers understand how parasites evade the host immune system, decide when to switch their developmental trajectory, react to stress or use different red blood cell invasion pathways.

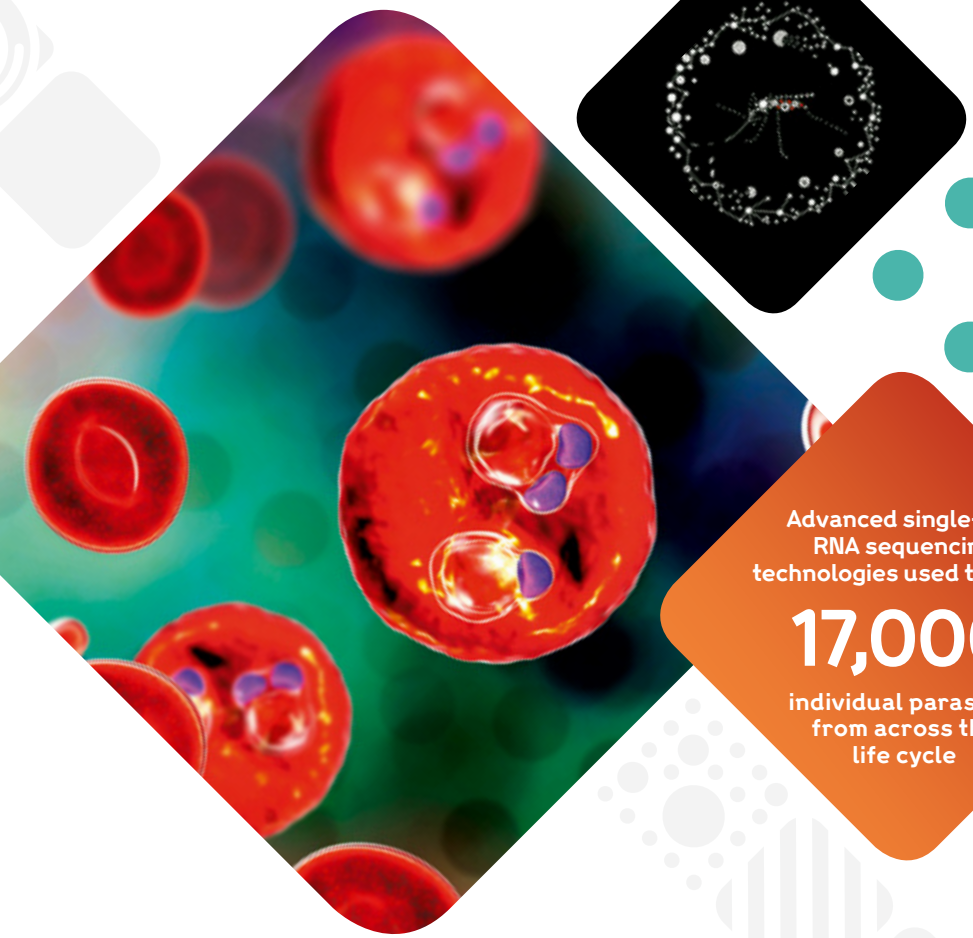
The research is published in *Science*, and all the data is freely available for use. The atlas builds on initial work published in 2018 where the team analysed over 500 individual parasites during the blood stages of the parasite's life cycle.

References

Howick VM *et al.* The Malaria Cell Atlas: Single parasite transcriptomes across the complete *Plasmodium* life cycle. *Science* 2019; **365**: eaaw26190.

Reid A *et al.* Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. *eLife* 2018; **7**: e33105.

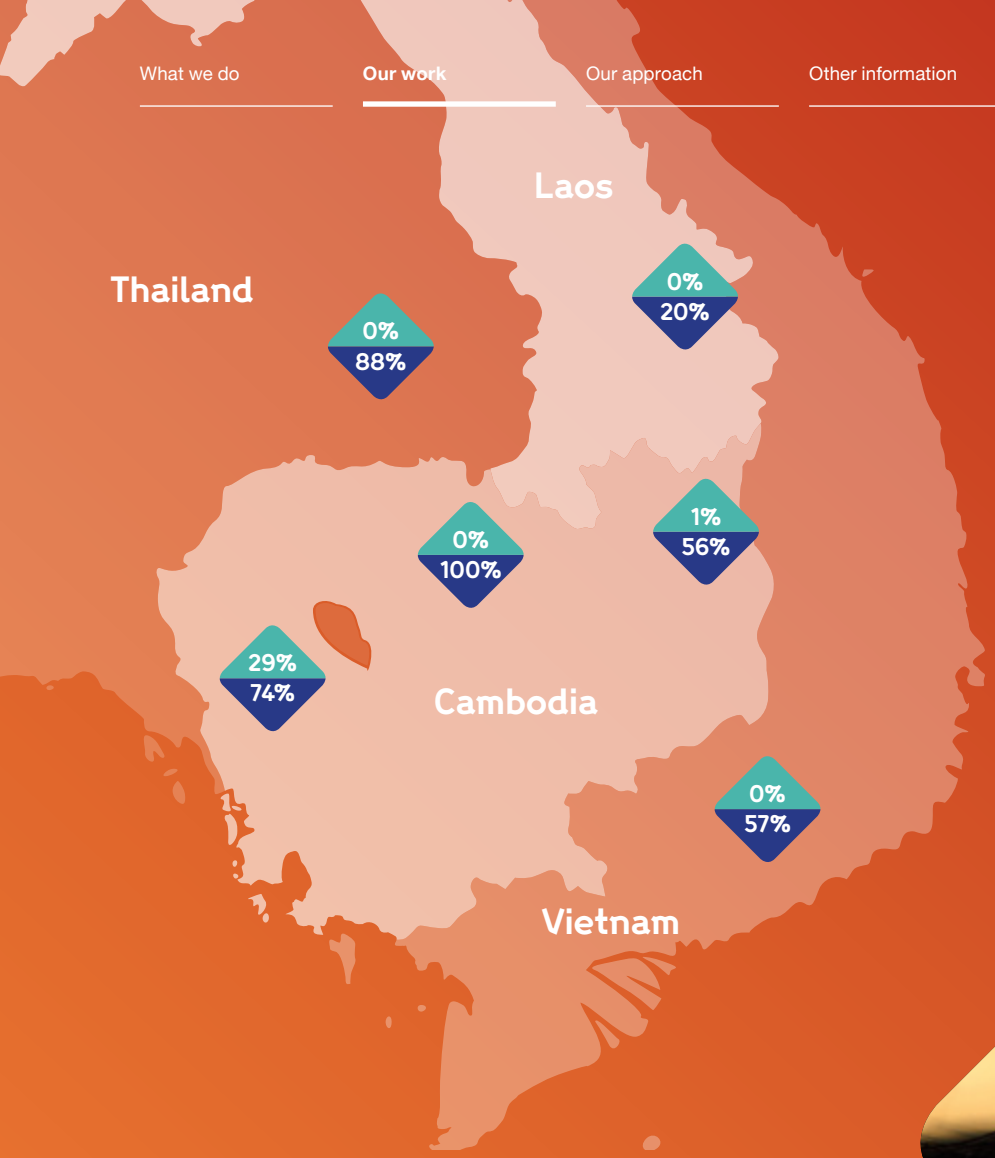
Tool: Malaria Cell Atlas



Advanced single-cell RNA sequencing technologies used to study

17,000+

individual parasites from across the life cycle



Rise of multidrug resistant KEL1/PLA1 malaria strain

Before 2011

After 2016

80%+

parasites in some regions were multidrug-resistant

12

Multidrug resistant malaria spreading in Asia

Sanger scientists used genomic surveillance to reveal that malaria resistance to two first-line antimalarial drugs has spread rapidly from Cambodia to neighbouring countries in Southeast Asia.

Over the last decade, the first-line treatment for malaria in many areas of Asia has been a combination of dihydroartemisinin and piperaquine (DHA-PPQ). In 2018, researchers from the Sanger Institute, University of Oxford and Mahidol University, Bangkok, identified a *Plasmodium* strain resistant to DHA-PPQ, named KEL1/PLA1. It had been had spreading across Cambodia, unnoticed, between 2007 and 2013.

In the most up-to-date and comprehensive genome study of malaria parasites in Southeast Asia, the team has now sequenced and analysed the DNA of 1,673 *Plasmodium falciparum* parasites. Their analysis, reported in *The Lancet Infectious Diseases*, revealed that the situation has worsened. The multidrug-resistant KEL1/PLA1 parasites have spread internationally, replacing the local parasite populations in Vietnam, Laos and Northeastern Thailand.

The researchers also discovered that the resistant strain has picked up additional genetic changes in the chloroquine resistance transporter gene which may be enhancing resistance even further. A related report on clinical outcomes revealed that these mutations were associated with complete treatment failure of DHA-PPQ.

The findings highlight the rapid and worrying spread of this deadly pathogen. Active genomic surveillance is now vital to inform national malaria control programmes and to help reduce the risk of a major global outbreak.

Reference
van der Pluijm RW *et al.* Determinants of dihydroartemisinin-piperaquine treatment failure in *Plasmodium falciparum* malaria in Cambodia, Thailand, and Vietnam: a prospective clinical, pharmacological, and genetic study. *The Lancet Infectious Diseases* 2019; **19**: 952–961.

Tree of Life

60,000
species in Britain and ireland will be sequenced as part of the Darwin Tree of Life

In this section

The Darwin Tree of Life project is part of a global project to sequence

1.5m
species

1

Darwin Tree of Life sets down roots

In 2019 the Sanger Institute welcomed Professor Mark Blaxter as the Tree of Life Programme Lead and the Darwin Tree of Life project secured £9.4 million in funding from Wellcome.

The Darwin Tree of Life project is one of several initiatives across the globe working towards the ultimate goal of sequencing the genomes of all complex life (eukaryotes) on Earth. The project is the first effort to sequence the species of one geographical area – Britain and Ireland. The project joins other global initiatives in a historic venture known as the Earth BioGenome project, which aims to sequence 1.5 million species of plants, animals, fungi and protists.

Exploring the genomes of these organisms will give an unprecedented insight into how life on Earth evolved and uncover new genes, proteins and metabolic pathways as well as new drugs for infectious and inherited diseases. At a time when many species are under threat from climate change and human development, these data will also help characterise, catalogue and support conservation of global biodiversity for future generations.

The Sanger Institute will serve as one of the hubs for sequencing and assembling the genomes of 60,000 species from Britain and Ireland. Professor Mark Blaxter joined the Institute to set up and lead the new Tree of Life Programme.

The Sanger Institute will collaborate with partners at the Universities of Cambridge, Edinburgh and Oxford, the Earlham Institute, EMBL's European Bioinformatics Institute (EMBL-EBI), The Marine Biological Association, Natural History Museum, and the Royal Botanic Gardens at Kew and Edinburgh. Together, teams will identify and collect species, set up new pipelines and workflows to process large numbers of species through DNA preparation, genome sequencing and assembly, gene finding and annotation.

New methods will be developed for high-throughput and high-quality assembly of genomes and their annotation, and data will be shared openly through existing data sharing archives and project specific portals.

The £9.4 million funding from Wellcome will support researchers to launch the first phase of sequencing 60,000 species in Britain and Ireland. This will see the teams collect and barcode around 8,000 key species, and deliver high-quality genomes of 2,000 species.

These data will be of enormous value to the international scientific community, including those working in life sciences, medicine, alternative energy and climate research. The data will also act as a global resource for public engagement experts, naturalists, citizen scientists, university students and schools.

Link
Darwin Tree of Life project news story

2

Sequencing an entire ecosystem

Species living in Wytham Woods in Oxfordshire will be among the first to be sequenced for the Darwin Tree of Life project.

Wytham Woods is one of the most studied areas of woodland in the world. Owned and maintained by the University of Oxford, its 1,000 acres are a designated Site of Special Scientific Interest (SSSI). Home to a rich diversity of species, there is a wealth of long-term biological data available about the ecosystem.

As part of the Darwin Tree of Life project, the species living there will be collected and catalogued, and their genome sequences will add to the wealth of data available on the ecosystem.

It's thought there are approximately 4,000 eukaryotic species in the woods, though this is likely to be an underestimate. There are over 700 species of Lepidoptera (butterflies and moths), and these will be among the first species to have their genomes sequenced.

The goal over the first two years of the project will be to sequence the dominant players in the ecosystem, and a broad taxonomic spread of species.

For most species, good quality genetic or genomic information is unavailable, so the data will be hugely valuable to the research community worldwide. Having data from an ecosystem will allow scientists to discover how different species have adapted to living in the same environment. For example, genome sequence data could give insights into how different insects break down the same defensive plant toxin. It may be that there are different solutions to the same problems; alternatively species may have evolved the same solutions independently.

The data will also provide a basis for analysing genetic responses to environmental changes and human disruptions to habitats, as researchers will be able to study and compare how gene expression changes over time.

Link
Tree of Life Programme at the Sanger Institute

1,000
acres are a designated Site of Special Scientific Interest

Approximately
4,000
eukaryotic species in Wytham Woods

The woods contain
700
species of moths and butterflies

3

Brown trout genome will help explain species' genetic superpowers

The newly sequenced brown trout genome will help settle the debate around whether it is one species or several and aid conservation efforts as we face a warming planet.

Brown trout (*Salmo trutta*) is one of the most genetically diverse vertebrates. Taxonomists once classified brown trout as up to 50 distinct species, though now they are considered one. Different populations have adapted to exploit particular biological niches, with some living their whole lives within a 200 metre stretch of freshwater stream while others migrate from the stream where they were born to the open sea.

Researchers at the Sanger Institute and their collaborators have sequenced the brown trout genome for the first time. The data will allow scientists and conservationists to better understand the genetic roots of this highly specialised species. The genome will also help settle the long-standing debate about whether the physically-varied brown trout is a single species or several.

The brown trout genome will enable scientists to compare DNA from different populations of trout to the reference genome sequence – giving insights into how particular genetic variations allow certain trout to live in habitats that would be fatal to others. Pinpointing genetic variations that allow Scottish loch trout to adapt to living in relatively acidic waters, for example, may be useful in guiding conservation efforts to protect populations affected by increasing acidity in rivers and oceans as a result of global warming.

Link
Brown trout genome news story

Squirrel pox can kill up to

80%

of a red squirrel population in an affected area

4

Genome could hold the key to red squirrel survival

Sanger researchers and their collaborators have sequenced the genomes of red and grey squirrels, boosting research and aiding red squirrel conservation efforts.

Red squirrels are native to Britain and Ireland, but they are under serious threat of extinction in the UK because of habitat loss, competition with grey squirrels and fatal squirrel pox. While grey squirrels are immune to the disease, squirrel pox outbreaks can kill up to 80 per cent of a red squirrel population in an affected area.

Red squirrels are now confined to isolated pockets in northern England and Wales, with a more connected population in Scotland and Ireland. They are at great risk of being pushed out of their territories by greys, outbreaks of squirrel pox and inbreeding.

Sanger scientists extracted DNA from red and grey squirrel samples sent by collaborators. They used PacBio SMRT® and Illumina sequencing technology to generate the first, high-quality red and grey squirrel reference genomes.

It is hoped that the genomes will be used to identify the genetic basis of immunity to squirrel pox in grey squirrels, as well as in the reds that survive outbreaks. The genome can also be used to help ensure that populations are genetically mixed, helping to select animals for reintroduction based on their genetic compatibility with existing populations to give them the best chance of survival through successful breeding.

The red and grey squirrels are two of the species to be sequenced as part of the Sanger Institute's 25 Genomes project and will contribute to the ambitious Darwin Tree of Life project.

Reference
Mead D. *et al*. The genome sequence of the Eurasian red squirrel *Sciurus vulgaris* Linnaeus 1758. *Wellcome Open Research*. DOI: 10.12688/wellcomeopenres.15679.1.

Stories 3-4

25 Genomes for 25 Years – Institute's anniversary project
The brown trout, the red squirrel and the grey squirrel are some of the UK species to have been sequenced as part of the Sanger Institute's 25th anniversary 25 Genomes project. The project laid the groundwork for the ambitious Darwin Tree of Life project, which will sequence 60,000 complex species in the UK.

Our approach helps to nurture the next generation.

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

Enabling all to thrive
Read more about how we empower our staff on pages 46 and 47

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.

Page 42

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.

Page 44

Connections

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

Page 50

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.

Page 46

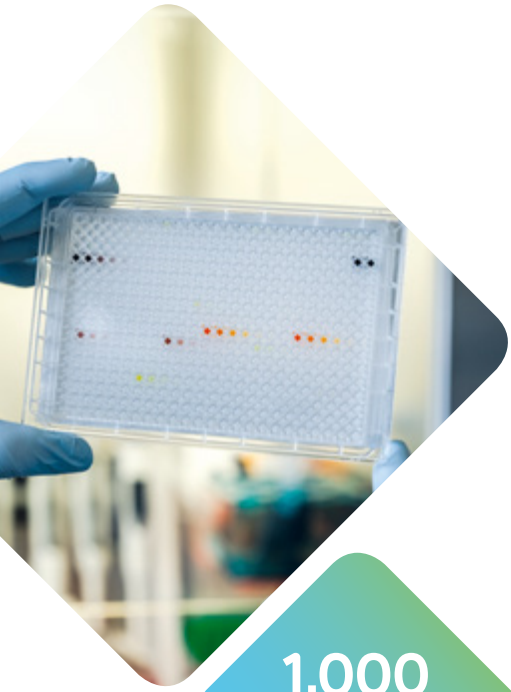
Influence

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.

Page 48

Scale

In this section



1,000
new cancer cell models are being generated

1

Cracking cancer at scale

The Sanger Institute is making the Cancer Dependency Map: a genome-wide chart of every cancer type's vulnerabilities to drive precision medicine.

The mutations that drive a cancer's development also change the way its cells function – making the tumour vulnerable in ways that normal tissue isn't. Charting these cancer-specific dependencies across the entire genome, in every cancer type and tissue, will provide the rulebook for precision medicine in cancer. Both the Sanger Institute and Broad Institute are producing cancer dependency maps that will enable researchers around the world to discover new drug targets, precisely match an individual cancer's mutations to specific vulnerabilities, and change the way we treat cancer.

To create such a map requires bench experimentation at an unparalleled scale and innovative computer-based analytics. To achieve this, the Sanger Institute has founded four streams of research: DepMap Models, DepMap Drugs, DepMap Genes and DepMap Analytics.

DepMap Models – by building on the 1,600 cell lines the Institute has already made available to scientists worldwide, this stream is creating the cell lines, 3D organoid cultures and genetically engineered cancer models required. The work encompasses the Human Cancer Model Initiative, which is generating approximately 1,000 new cancer cell models, and Cell Model Passports, which enables researchers to access key information on the genetic changes and drug sensitivities of a particular cancer type.

DepMap Drugs – the research extends the Genomics of Drug Sensitivity in Cancer project to systematically screen hundreds of anti-cancer drugs and drug combinations against thousands of cancer cell lines to identify and associate genomic characteristics with dependencies and treatment options.

DepMap Genes – this stream is using CRISPR technology and programmable single-guide RNAs to knockout every single gene across the human genome in many cancer types to map gene function and determine individual cancer cell dependencies. The results are made available via ProjectScore, which ranks and prioritises vulnerabilities and has already highlighted a promising new drug target.

DepMap Analytics – this work is creating the new computational approaches and software needed to process, combine, analyse and visualise the data being produced by the Models, Drugs and Genes research.

All the models, data and algorithms generated by the Institute are being made available to the worldwide research community to provide the foundations for precision oncology.

Link
Cancer dependency map at the Sanger Institute website

2

Cancer research has a COSMIC future

After more than 15 years of providing vital cancer gene data, the Catalogue of Somatic Mutations in Cancer (COSMIC) database has become self-sustaining and is driving innovation in precision oncology.

COSMIC started in 2002 as a spreadsheet for the Sanger Institute's cancer genome project to keep track of the genetic mutations being discovered in cancer. Updated weekly, it proved to be so useful that it was developed and launched as a freely available, open-access database for the global cancer research community just two years later.

COSMIC has grown rapidly. At launch it contained data on mutations across just four key cancer-related genes. It now comprises tens of millions of genetic variants across the entire human genome, encompassing every type of mutation mechanism, across every form of human cancer.

Currently more than 4,000 independent researchers use the COSMIC website every day to help drive innovation in precision oncology. Many more benefit from COSMIC's highly annotated data through such global genetic resources as NIH GDC, NCI CGC, OpenTargets, cBioPortal, and Ensembl.

COSMIC helps to power much early-stage pharmaceutical research and cancer diagnostics development, and new resources and insights are regularly developed. One example is COSMIC-3D, developed with Astex Pharmaceuticals, which enables visualisation of cancer-associated proteins to aid drug development. Another is the integration of methylation and expression information in collaboration with Bayer. Further updates and innovations are planned for 2020.

The costs of growing and sustaining COSMIC are substantial and, while initially funded by Wellcome, the COSMIC team have deployed a range of commercial licensing opportunities. The result is that COSMIC is now self-funding by pioneering the balance of commercial data sales with academic open-access to all academic and charitable institutions across the world. This model will ensure that its data will continue to power genomics research for many years to come.

Link
COSMIC website

COSMIC: then and now

April 2004 – Version 1

57,444 tumours

10,647 mutations

September 2019 – Version 90

1,412,466 samples

34,320 whole genomes

26,829 papers

9,733,455 coding mutations

13,099,101 samples

19,396 gene fusions

1,207,190 copy number variants

9,197,630 gene expression variants

7,929,101 differentially methylated CpG

3

Sequencing 500,000 genomes

Having successfully delivered 50,000 UK Biobank human genome sequences on time and on budget, the Sanger Institute is now using its power to deliver the most ambitious human genome sequencing project to date.

The Sanger Institute is helping to produce a game-changing genomics resource: the UK Biobank's Whole Genome Sequencing project of 500,000 volunteers' genetic code. The data will be allied with their clinical and lifestyle data, providing a wealth of information to power the next wave of research into human health and disease worldwide.

The project is the most ambitious sequencing programme undertaken by a public-private partnership. It follows on Sanger Institute's successful initiation of the UK Biobank's proof-of-concept pilot project to sequence 50,000 volunteers' whole genomes.

The £200 million collaboration brings together the UK Government, charities, researchers, and four leading pharmaceutical companies. The sequencing work is being shared by the Sanger Institute and deCODE genetics in Iceland.

The resulting database of anonymised genomic information, linked with comprehensive clinical characterisation, will help to provide a unique insight into why some people develop particular diseases and others do not.

Building on the work of the pilot programme, the plan is to complete sequencing of the remaining 450,000 participants in two tranches. The first tranche of up to 125,000 whole-genome sequences is expected to be accessible in Spring 2021 and the entire cohort should be available in early 2023.

Link
UK Biobank Whole Genome Sequencing project at the Sanger Institute

Innovation

In this section

Multi-year strategic collaboration with Genentech worth up to

$534 million

1

Microbiotica: fast maturing company grows at a pace

Innovative Sanger Institute spin-out company Microbiotica has established itself as a leading player in microbiome-based live therapeutics and biomarkers.

Microbiotica was founded on gut microbiome research carried out in the Institute's Parasites and Microbes Programme. It is now fast maturing into a major force in gut research. In 2018, its extensive culture collection of human gut bacteria and reference genome database of gut bacteria species attracted the interest of biopharma world leader Genentech and secured multi-year strategic collaboration worth up to $534 million.

In 2019, the magnitude and potential of this partnership was recognised when it was announced as a Scrip Awards finalist in the Best Partnership Alliance category.

The collaboration brings together patient samples from Genentech's clinical trials of potential inflammatory bowel disease (IBD) medicines and Microbiotica's precision metagenomics microbiome platform to discover and develop diagnostic biomarkers and treatments for the condition. In addition, Microbiotica is using the samples to expand its bacteria collection and reference genome database.

Using bioinformatics and artificial intelligence, Microbiotica's researchers identify the bacterial species present in individuals' guts to discover microbiome 'signatures' that can be linked to specific diseases and conditions. They then validate these signatures in mice with a humanised microbiome.

The Microbiotica team and functions are currently located across the Wellcome Genome Campus in dry-lab compute and wet-lab experimentation spaces that were provided to the company as it established itself. To aid the company's continued growth, they will be moving to the nearby Chesterford Research Park, releasing space for the next innovative start-up to grow at the Wellcome Genome Campus.

Link
Microbiotica website

2

Congenica grows globally

Congenica, the Sanger Institute's spin-out clinical decision support company, has continued its national and international growth through strategic alliances, data integration and new funding.

Based at the Wellcome Genome Campus' BioData Innovation Centre, Congenica delivers accelerated interpretation of complex genomic data to improve disease diagnosis for the UK NHS and a range of hospitals and research projects across the globe.

During 2019, the company partnered with Chinese firm BGI Genomics to launch BGI-Xome, an all-in-one whole exome sequencing and interpretation service. The offering combines BGI Genomics' laboratories with interpretation by certified clinical scientists using the Congenica analysis and clinical decision support platform to provide a gold-standard service for its users.

Congenica also teamed up with Nonacus Ltd to create Exome CG, a clinically validated exome capture kit that enables whole exome sequencing and targeted copy number analysis in a single assay. The product simplifies generating and interpreting molecular and cytogenomic data by removing the need for additional frontline tests such as microarrays.

The company continues to develop its analysis software and has increased the platform's power by integrating reference data from the DECIPHER project and Mastermind. To enable further growth the company secured an additional £13.25 million in equity investment from new strategic investors, including Digital China Health Technologies Corporation Limited, alongside follow-on funding from existing investors to reach £23.3 million in series B financing.

Closer to home, Congenica is also supporting the Next Generation Children project – a collaboration between the NHS East Anglian Medical Genetics Service, the NIHR Bioresource – Rare Diseases, and the University of Cambridge to investigate the clinical effectiveness of rapid whole genome sequencing for children in intensive care. The approach is providing fast and accurate genomic diagnoses that may help children to receive appropriate treatment and avoid further invasive tests, and is likely to become the template for future NHS services.

Congenica

Launched
all-in-one
whole exome sequencing and interpretation service with BGI Genomics

Link
Congenica website: congenica.com

3

Funding future technologies

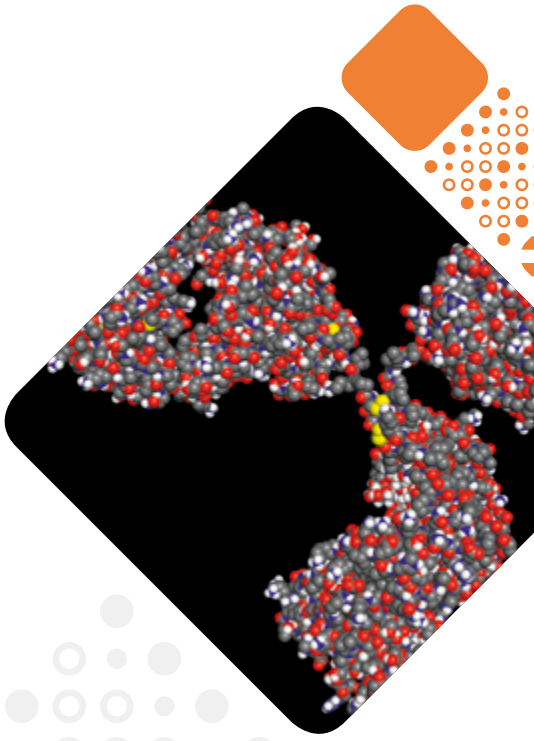
In 2019, the Sanger Institute provided more than £500,000 in development funding to early stage in-house research as part of its commitment to deliver scientific and healthcare benefits.

The goal of the funding, awarded by the Sanger Institute Translation Committee, is to allow the intellectual property of the discovery, technique or process to be sufficiently developed to attract further translation by an external partner. The projects currently being supported cover a wide range of scientific endeavour, including equipment design, immunology and novel gene knock-out library construction.

Examples of the science that are being developed for translation include FLIP and IsoTyper.

FLIP is a technique that dramatically improves the speed, affordability, scale and efficiency of producing conditional human and mice gene knock-out models. The approach places a cassette of custom DNA sequence into a gene which will not affect the gene's functioning. However, when a recombinase enzyme is used to flip the cassette's orientation, the gene is disrupted and knocked out in a single step. The effect is also completely reversible – a second recombinase will flip the cassette back to its original orientation and reinstate gene function.

IsoTyper allows researchers to profile the antibody diversity and functioning of isotype-specific B cells in a single experiment. By using specially-designed barcoded DNA primer sets and DNA sequencing, the system can identify antibody-producing cells that flow cytometry techniques cannot. IsoTyper could be used to monitor leukaemia progression and responses to therapy, to optimise vaccine development, or to provide early detection of transplant rejection or developing autoimmune or anti-drug responses.



Link
Sanger Institute Innovations: sanger.ac.uk/innovations

Culture

In this section

Almost
500
Sanger Institute technicians make our science happen

1

Institute ramps up its support for technicians

The Sanger Institute is indebted to its highly skilled workforce of almost 500 technicians whose diverse range of skills are crucial to the delivery of its science. To nurture their career development, the Institute is a member of the Technician Commitment.

From cutting-edge cell manipulation carried out on the latest laboratory equipment, to novel bioinformatics analysis of genomic data using our high-performance IT infrastructure, the Institute's technicians deliver the Institute's world-leading research. To ensure that the Institute's technicians receive the training, career development opportunities and recognition they deserve, the Institute signed up to the Technician Commitment in 2018.

The commitment is a university and research institution initiative that aims to ensure visibility, recognition, career development and sustainability for technicians working in higher education and research, across all disciplines. In May 2019, the Institute submitted its action plan to address these key issues.

The plan is now being executed by a Technician Commitment manager who champions the Institute's technicians both internally and externally. The Institute has launched new training courses and has joined Higher Education and Technician's Educational Development (HEaTED) to expand the range of resources, specialist training, advice, and networking opportunities to support technicians' development. In addition, the Institute has developed procedures to enable staff to become members of professional bodies covering animal technology, biomedical sciences and computer technology.

Over the course of 2020 a programme of events, internal policies, procedures and processes, and seminars will be established to further promote and empower the Institute's technicians.

Link
Technician Commitment at the Sanger Institute

2

How DORA supports Sanger scientists

Signing up to the Declaration on Research Assessment (DORA) reinforces the Institute's approach to recognising the value of all its staff's scientific contributions.

The Sanger Institute produces research that powers genomic and medical exploration and discovery around the world. Its staff create and disseminate cell lines, genome manipulation and analysis techniques, data repositories and software tools – available free of charge and open access wherever possible. The scientific impact of the Institute's research output is much greater than simply counting its published scientific papers – and the impact factor of the journals they are in.

The Institute is keen to ensure that its researchers are free to produce science that produces the greatest possible benefit to research, healthcare or society, regardless of the form it takes. For this reason, it was delighted to sign up to the San Francisco Declaration on Research Assessment in 2019.

To avoid placing unnecessary pressure on staff to publish in high-impact journals, the Institute employs a progressive approach that draws on expert peer review and a range of information to assess the scientific merit of a researcher's work. To this end, the Sanger Institute broadly defines research excellence to include data resources, leadership of large consortia and mentoring, and to recognise societal impact through translation into healthcare or scientific benefit and public engagement.

These guiding principles also align with Wellcome's new policy on Open Access, which will launch in 2021.

Link
DORA at the Sanger Institute

3

Enabling all to thrive, valuing diversity

The Sanger Institute is a genomic research 'ideas factory', powered by the imagination, skills, experiences and insights of our diverse, interdisciplinary staff. By valuing difference and nurturing talent, we seek to enable everyone to reach their full potential and deliver world-changing science.

In common with many industries, scientific research can suffer from a lack of diversity, but the Sanger Institute is committed to fostering an environment where the talents of all genders, races and neurodiversity can thrive. At a structural level, the Institute is addressing inequalities in pay and career opportunities, with the aim to establish equal pay across all nine pay grades by 2022. By updating our staff's job families to introduce a new job family and link pay to market rates, the mean pay gap has already reduced from 20% mean (2016) to 16.2% mean (2017).

The Institute is committed to supporting women's careers and it has addressed potential obstacles by providing generous parental leave, return-to work-fellowships, paid carers leave and flexible working. Our organisation is a bronze award member of the Athena SWAN Charter, and we have recently applied for silver.

In addition, the Institute has introduced measures that are increasing the proportion – and interview success rate – of women at senior grades. These include unconscious bias training, expanding our management and leadership programmes, and increasing access to promotion opportunities.

To further develop our inclusive culture, the Institute has partnered with AdvanceHE to embed the principles of the Race Equality Charter in our workplace, through staff support and tracking the ethnicity pay gap. In 2020 we will join the Stonewall Diversity Programme, and will use the Stonewall Workplace Equality Index to benchmark our support of our LGBTQ+ staff against best practice.

But there is more we can do to support diversity in our workplace and the Sanger Institute is exploring how we can support neurodiversity within our staff.

Link
Equality in Science at the Sanger Institute

The mean pay gap reduced from **20% to 16.2%**

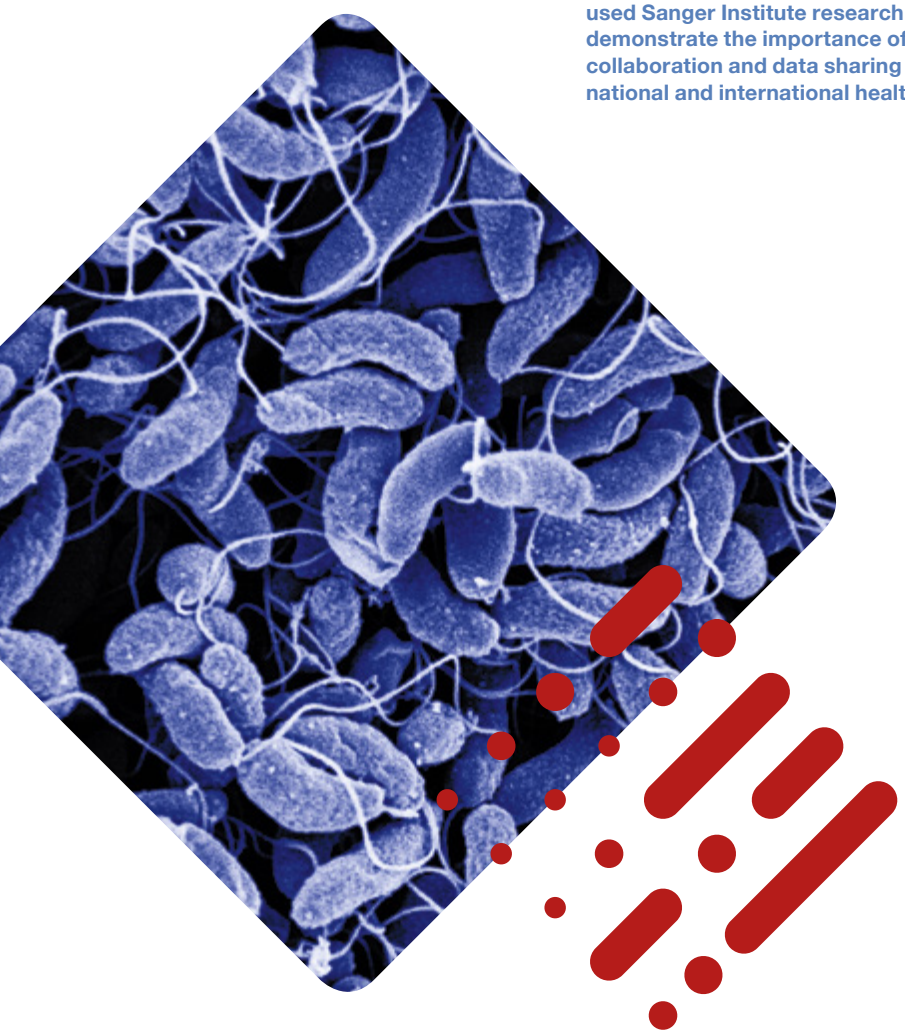
Influence

In this section

1

Sanger's work highlighted by Chief Medical Officer

In her final annual report, the outgoing UK Chief Medical Officer (CMO) used Sanger Institute research to demonstrate the importance of global collaboration and data sharing for national and international health.



The speed and ease of international travel means that continued health and prosperity is now a shared global endeavour. Focusing solely on domestic diseases brings the risk of being unable to respond to global threats when they arise, the CMO warned in her 11th annual report: *Health, our global asset – partnering for progress.*

To highlight the value of international collaboration, the report featured an article from a leading Sanger Institute Parasites and Microbes Programme researcher. It detailed how a rapid-response global team for cholera successfully used genomic surveillance to trace the source of 2017's Yemen outbreak.

The work brought together skills and experience from around the world. In Yemen, Médecins Sans Frontières (MSF) and local healthcare workers collected *Vibrio cholerae* samples from the Yemeni population and from a temporary refugee camp. Researchers at the Institute Pasteur, Paris sequenced the bacterial genomes and Sanger scientists analysed the data. By comparing the sequences to those from the current, ongoing global pandemic, the team showed that the Yemen outbreak was caused by the same 7PET lineage, and was likely to have entered the region from Eastern Africa. The team also found that the Yemeni samples were missing four genes responsible for resistance to commonly used antibiotics, making the bacterium more susceptible to treatment.

UK-based researchers play a vital role in helping to understand emerging global health threats. The Institute is a leading player in building the partnerships needed for global health by developing local genomic research capacity in low- to middle-income countries through training and sharing technical knowledge.

Link
Sanger Institute contribution to the 2019 CMO report

2

Institute advises UK Parliament on commercial genomic testing

As at-home DNA testing kits become increasingly accessible, and commercial providers seek to provide genetic testing for the NHS, the Sanger Institute provided expert advice to the UK Government on the opportunities and risks.

In 2019 the House of Commons Science and Technology Committee set up an inquiry into commercial genomic testing. To help guide UK lawmakers through the issues around health diagnoses outside of clinical support, result accuracy, implications for the NHS, and the safety of sensitive data, the Institute provided written and oral evidence.

The Institute pointed out that commercial genomic services fall into two main areas:

- direct-to-consumer tests – such as for genomic ancestry
- diagnostic services for the NHS – including prenatal screening, complex disease risk assessment, and searching for rare genetic disorders.

In terms of direct-to-consumer tests, the Institute counselled that an independent regulatory body should be set up to oversee testing standards, accuracy and the privacy and security of individuals' genomic data.

Regarding commercial genomic services for the NHS, the Institute highlighted the need to develop a highly skilled workforce able to interpret the complex genomic data being produced with polygenic risk scores to deliver meaningful diagnoses. Without adequate resources for training or additional staff these services may place increased pressure on the healthcare system.

In addition, the Institute encouraged the NHS to adopt standard file formats for its genomic data and to store them in accessible databases. In this way, with appropriate security and governance, the research community could draw on the collective power of these data to further benefit healthcare and develop new treatment options.

Link
Sanger Institute contribution on Commercial Genomics inquiry

3

Nagoya: how to balance fairness with effectiveness

As part of its core principles, the Sanger Institute uses the Nagoya Protocol to handle its genetic samples and discoveries fairly and ethically. But the Institute is concerned that proposed expansions in scope could harm the future of global genomic research.

The Sanger Institute is committed to generating and sharing freely available open-access data to benefit the worldwide research community and wider society. To ensure that the Institute leads the way when appropriately gathering and handling samples from around the world, it has embedded a Nagoya Protocol officer into its processes.

Much of the Institute's research uses samples from other countries: from monitoring the global emergence and spread of infectious diseases to studying the influence of environmental factors on cancer. The Nagoya Protocol officer works across the Institute at all levels to ensure that its researchers develop, and use, best practice structures and processes.

Recently the international Convention on Biological Diversity (CBD) proposed extending the Nagoya Protocol to include digital sequence information. While supporting fair use of samples, the Institute is concerned that such a move could cause more societal and scientific harm than good.

In counselling the CBD, the Institute cited the vital role of globally-accessible, free, open access databases such as the European Nucleotide Archive in facilitating collaborative research. If the ability to share digital sequence information freely and rapidly through such resources was restricted, then it might prove impossible to respond effectively to global public health emergencies, deliver actionable public health strategies, or address conservation challenges.

Link
Sanger Institute response to Nagoya Protocol proposal

Connections

In this section

1

GA4GH: powering genomic research's future

The Sanger Institute is helping to build the foundations of easy access global genomic data sharing to enable genomic enquiry at scale.

The power of genomic research lies in analysing enormous datasets, often created by aggregating data from multiple data sources around the world. Yet this work is hampered by the relevant data sources being worldwide, in countries with different regulatory and data protection processes, and different clinical and genomic terminology. Finding, obtaining access to, and aligning the relevant datasets to analyse them are major barriers to future discovery.

To remove these obstacles, the Sanger Institute and the Broad Institute helped to found the Genomic Alliance for Global Health (GA4GH) in 2013 to enable genomic data sharing for the benefit of human health.

In the same way that the World Wide Web Consortium develops the standards needed to enable the internet to deliver easy-to-find, globally available information that can be read on any computer, the GA4GH is developing the standards needed for an internet of genomic data.

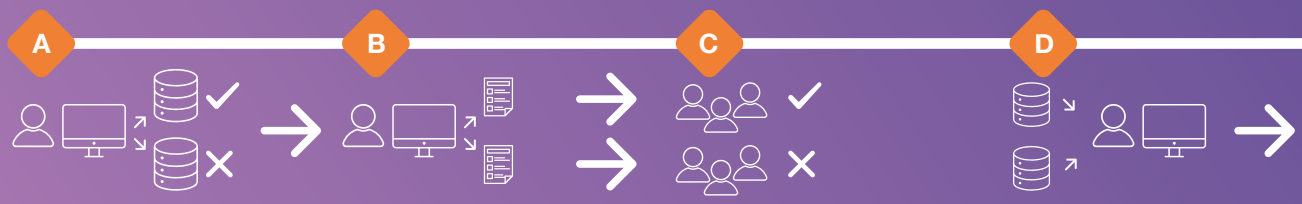
The Sanger Institute provides expert guidance and administrative support to the GA4GH. The Institute's Policy team has helped the Alliance develop its Regulatory and Ethics Framework, based on human rights. In addition, in 2019, Institute researchers helped to deliver the GA4GH Passports and Data Use Ontology standards which work in harness to enable all researchers to automatically access data worldwide by demonstrating that they fulfil the necessary restriction criteria.

Life today – powered by the internet – is radically different to life just two decades ago: through standards such as Passports and Data Use Ontology, the GA4GH will enable a paradigm shift in global genomic research.

Link
GA4GH website

Workflow of accessing and analysing genomic data

Currently



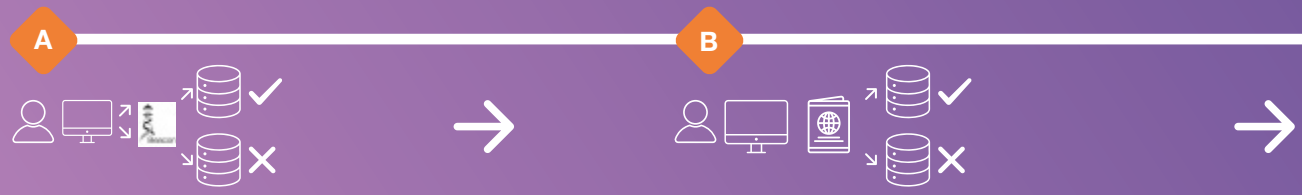
A: Search for individual sites for information

B: Apply for access by filling forms to show that the researcher matches the restrictions of the data in terms of data use agreements, data protection and security

C: Wait for permission to be granted by a committee overseeing the data

D: Download the data

Using GA4GH approaches, including Passports and Data Use Ontology



A: Use Beacon protocol to quickly search all relevant databases

B: Data providers check researcher's Passport (using Data Use Ontology to detail organisation, level of access, data protection and regulation criteria) automatically to grant access

2

Free online learning for all

Campus-based partnership provides free genomic training for everyone around the world.

Over the past 15 years, the Wellcome Genome Campus Advanced Courses and Scientific Courses (ACSC) team and Sanger scientists, along with researchers around the world, have successfully collaborated to train other researchers worldwide in bioinformatics and genetic techniques. But the number of people who can benefit has always been limited by the physical constraints of location and capacity.

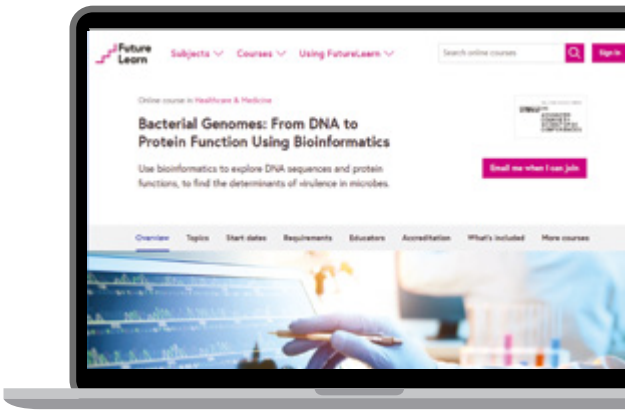
To enable everyone to benefit from the revolution in genomic medicine, ACSC has partnered with FutureLearn to harness the power of the internet to provide training across the globe for free. Designed and delivered in partnership with the Institute's pathogen researchers, the series of four courses explore how genomics is used to understand bacterial genomes to monitor the spread of diseases and drug resistance.

Sponsored by ACSC, each three-week-long course is completely free to use, including assessment and certification. The courses can be taken via computer, mobile phone or tablet and require no genomic knowledge. Aimed at doctors, nurses, microbiologists, secondary school teachers and pupils, the content is delivered via a mix of videos, articles, tests and quizzes to validate learning. Each course is presented by expert educators who also support the students.

So far, more than 13,500 have benefited from the online training and further courses are planned throughout 2020 and 2021.

To find out more, visit:
coursesandconferences.wellcomegenomecampus.org/event-type/online-courses

13,500
researchers and healthcare professionals have completed the online courses



A couple of months

E Convert data to a form that researcher can query

F Compare ontologies and try to match terminology

G Analyse data on computer/ organisation's system

H Get results

Matter of days

C Query data in the cloud using online servers (as data is in standard file formats)

D Get results

Image Credits

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

Outside

Front Cover – Butterfly wing – iStock
Inside
Front Cover – Cancer cell – Shutterstock

What we do

Page 2 – Gorilla hand – Getty images
Page 4 – Pancreatic cancer cells – Anne Weston, Francis Crick Institute
Page 5 – Mosquito – Emphyrio, pixabay.com
Baby ultrasound – Dr Wolfgang Moroder, Wikimedia Commons
Lebanese Crusader skeleton – Dr Claude Doumet-Serhal
Steptococcus A computer images – CDC/PHIL
Child with asthma inhaler – Adobe Stock
Page 6 – CT scanner – Adobe Stock
Big Ben Elizabeth Tower – Ukyo Katayama, pixabay.com
Surgeon washing hands – Piron Guillaume, Unsplash.com
Brown trout in water – Shutterstock
Newborn baby – Iuliia Bondarendo, Pixabay.com
Gorilla – Adobe Stock

Our work

Page 8 – Mosquito – Adobe Stock
Page 11 – CT scanner – Adobe Stock
Page 12 – Cancer cells – Anne Weston, Francis Crick Institute
Page 16 – Liver Cell Development Atlas – Newcastle University
Page 17 – Child's hand with intravenous line – Adobe Stock
Page 18 – Kidney cortex – Menna Clatworthy, University of Cambridge
Page 19 – Computer imagery of lungs – Adobe Stock
Lung cells section – Professor Wim Timens, UMCG
Page 20 – Child using asthma inhaler – Shutterstock
Page 21 – Expanded pluripotent stem cells – Xuefei Gao
Page 22 – Human genetics section – measles virus computer imagery, CDC/PHIL
Child with measles – CDC/Molly Kurnit, M.P.H
Page 23 – Crusader skull – Dr Claude Doumet-Serhal
Nigerian man – Ovinuchi Ejiohuo, unsplash.com
Page 24 – Blood sugar level testing – Adobe Stock
Page 25 – Child with building blocks – pixabay.com
Deciphering Developmental Disorders project logo – Deciphering Developmental Disorders project
Page 26 – Bacteria computer imagery – CDC/ Antibiotic Resistance Coordination and Strategy Unit
Mother and newborn baby – Iuliia Bondarendo, pixabay.com
Page 27 – Surgeons washing hands – Shutterstock
Page 28 – Blister packs of medications – Shutterstock
MRSA scanning electron microscopy – National Institute of Allergy and Infectious Diseases (NIAID)
Page 29 – *Klebsiella pneumoniae* scanning electron microscopy – National Institute of Allergy and Infectious Diseases (NIAID)
Page 30 – Strep A scanning electron microscopy – National Institute of Allergy and Infectious Diseases, National Institutes of Health
Page 31 – World War 1 soldiers – Imperial War Museums (Wikimedia Commons)
Page 32 – Gorilla – Adobe Stock
Page 33 – African family receiving antimalarial medication – William Brieger, Jhpiego
Page 34 – Malaria parasites in red blood cells computer imagery – Shutterstock
Page 35 – Mosquito on arm – Emphyrio, pixabay.com
Page 36 – Bumblebee – iStock
Page 37 – Butterfly on flower – iStock
Wytham Woods – Sandra Bouwhuis
Page 38 – Brown trout in water – Shutterstock
Page 39 – Red squirrel – Adobe Stock

Our approach

Page 40 and 41 – Child and parent hands – iStock
Page 45 – Congenica logo – Congenica Ltd
Immunoglobin molecule computer imagery – Adobe Stock
Page 49 – Big Ben Elizabeth Tower – Ukyo Katayama, pixabay.com
World connections map – Pete Linforth, pixabay.com

Wellcome Sanger Institute Highlights 2019/20

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, 2020.

Designed and produced by

CONRAN DESIGN GROUP

Wellcome Sanger Institute
Tel: +44 (0)1223 834244

sanger.ac.uk