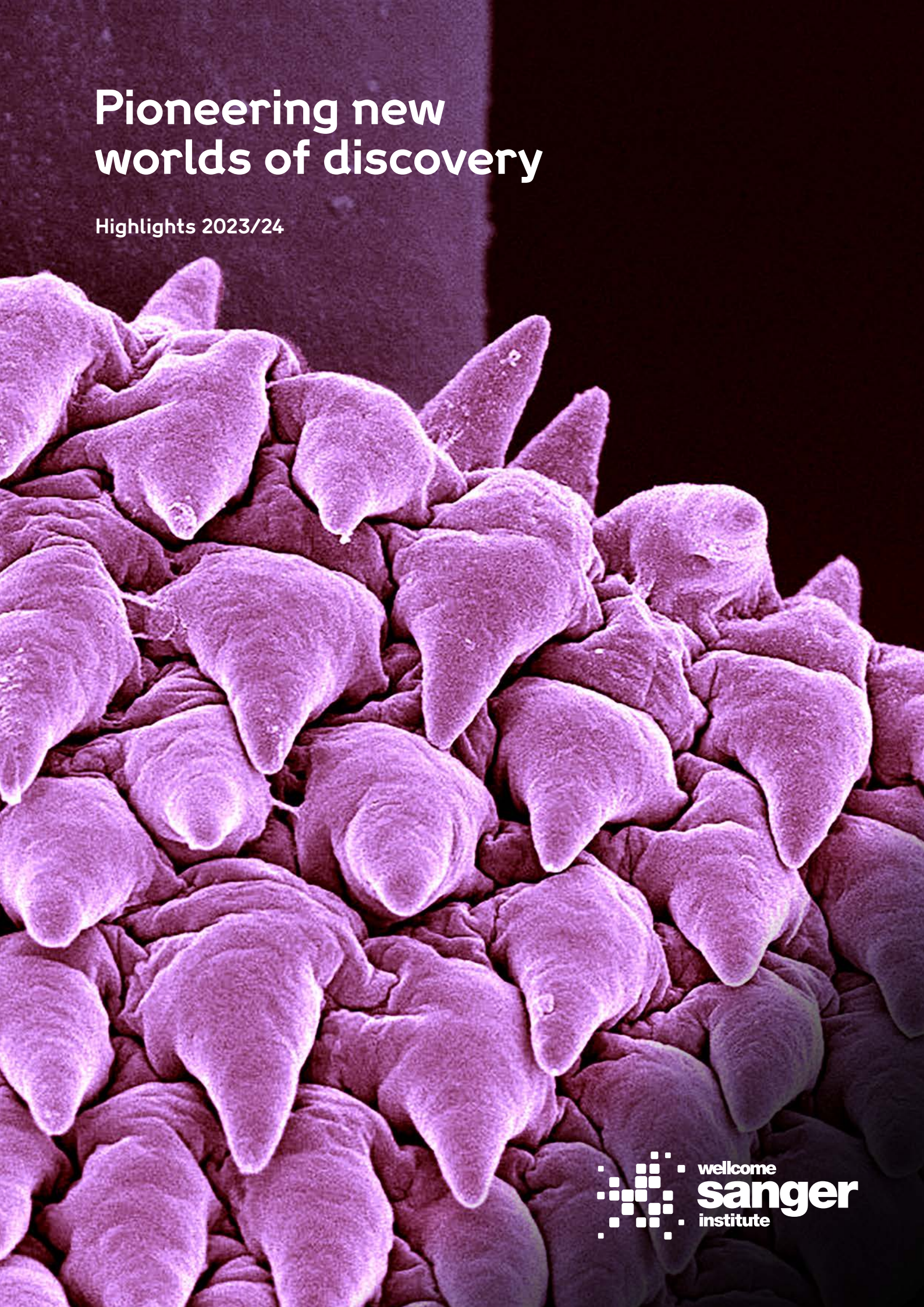
Pioneering new worlds of discovery

Highlights 2023/24



wellcome sanger institute

Contents

We apply genomic technologies at scale to advance understanding

Unlocking new fields of discovery

Throughout the history of science, seismic changes in knowledge and understanding have come about from the coupling of new technologies with bold, creative ambition. From astronomy to zoology, advances such as the telescope, the microscope, computation and transportation have driven paradigm shifts in discovery and exploration. The Sanger Institute was born out of, and is powered by, these two forces.

Just over 30 years ago, the Sanger Institute was created by Wellcome to exploit and drive technological advances in DNA sequencing to deliver the seemingly impossible: a reference human genome. Its completion ushered in a new age of genomic research, founded on the techniques, ingenuity and open-access research principles developed by this Institute and its collaborators.

Since then, the Institute has pioneered the use of new sequencing, cellular, imaging and analysis technologies to provide ever fine-grained insights into the biology of life. Powered by the leap from capillary sequencing to next-generation flow-cell techniques, Sanger scientists have envisioned and delivered projects of world-leading scale. First, we studied 1,000 genomes, then 10,000 and recently delivered over half of UK Biobank's 500,000 genomes. In a similar vein, the same approaches have given genetic diagnoses for 5,500 children with rare developmental disorders, in partnership with the NHS.

At the other end of the scale, our researchers, drawing on the ingenuity of our technical staff, are exploiting DNA sequencing to build the reference genome library of all eukaryotic species in the UK and Ireland. Their work, which employs the latest approaches to automation and computation, is laying the foundations for impactful conservation of biodiversity across the UK and the world. The research is also discerning the underlying mechanisms that drive the evolution of life on earth.

Advances in imaging and cell-characterisation methodologies are enabling Sanger scientists to track the role and destination of every cell from conception to old age. These insights, as part of the Human Cell Atlas initiative, is driving previously unattainable discoveries that open new therapeutic possibilities. Exquisitely intricate maps of lung, placenta and heart development covering space and time are revealing new cell types and interactions that could have profound impacts on medicine.

Combining creativity, technology and scale has powered transformative research by the Institute's cancer scientists. Applying CRISPR gene editing, organoid models, cell lines and prime editing to interrogate cancer genomes has uncovered many new therapeutic targets and diagnostic approaches. Additionally, the application of saturation genome editing offers an unparalleled opportunity to explore the effect of variation at every letter of a cancer gene – opening up a future of truly personalised diagnosis and treatment.

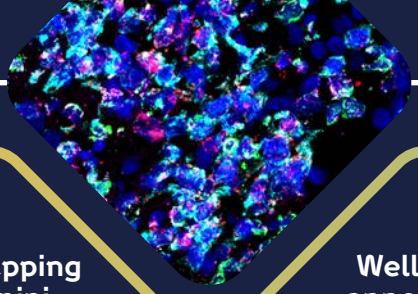
In the field of infectious diseases, real-time genomic surveillance of global pandemics is fast-becoming reality as our scientists build on the techniques, processes and analyses that they created during the COVID-19 pandemic. Amongst others, these approaches have delivered key insights into the value of vaccination in refugee communities to prevent devasting outbreaks of cholera, guiding future public health decisions.

And now the Institute is breaking ground in a new arena: generative and synthetic genomics. By combining the vision of our scientists with recent advances in DNA synthesis, gene editing and machine learning at scale, we seek to make biology programmable. The fruits of this work promise to deliver diagnostics, therapeutics and biomaterials at greatly reduced cost and environmental impact.

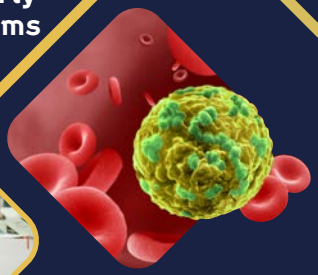
Yet none of the Institute's research would be possible without the creativity, collaborative spirit and dedication of the Institute's staff and our collaborators. For this reason, we seek to foster a culture of openness, teamwork and transparency. From providing practical and financial support for those with parental or caring responsibilities to encouraging transparency and accountability in our research relationships both at the Institute and in our collaborations and partnerships around the world, we look to embed equality and mutual benefit into the way we deliver our science.

Professor Matthew Hurles, Director
Wellcome Sanger Institute

2023 Timeline



Respiratory Virus and Microbiome Initiative launched to develop early warning systems for new viruses



Cell mapping and mini placentas give new insights into human pregnancy

Wellcome announced major investment in Genome Campus expansion

5,500 people diagnosed with rare genetic disease by major research study

Ocean Census initiative launched to discover 100,000 new species within 10 years

Child health research retreat empowered microbiome scientists in underrepresented countries

Machine learning predicts success of genome editing

Blood cancer manipulates immune cells in order to survive

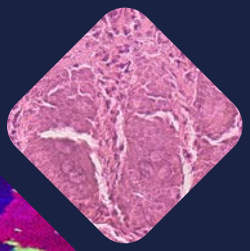
Single-cell atlas of the whole human lung released

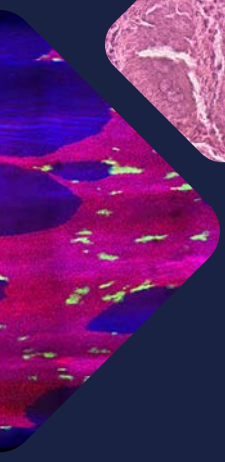
GASPACHO tool reveals susceptibility to COVID-19

JAN

NOTCH1 gene variants inhibit oesophageal cancer growth

'Collateral damage' from normal cell function plays a role in cancer





FEB

Professor Matt Hurles appointed Director of the Sanger Institute

MAR

Equality, Diversity and Inclusion champions recognised at Best Practice Award

Sanger Institute hosted the 2023 Research Institute Technician Symposium for technicians and technical experts

Research rewrites the rules of CRISPR activation

APR

Sanger Institute cancer drug discovery spin-out company, Mosaic Therapeutics, launched

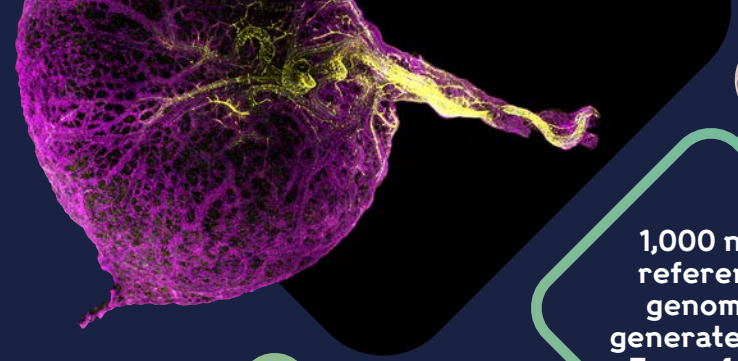
MAY

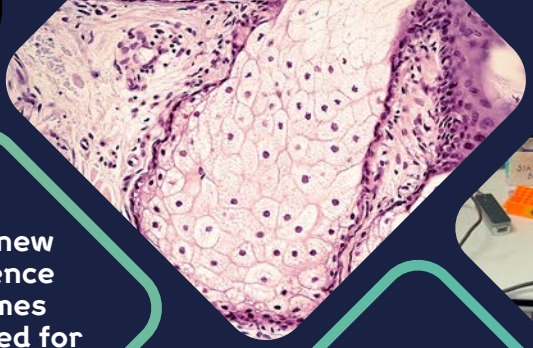
First draft human pangenome reference published

Jon Teague became the Director of COSMIC associate research programme

JUN

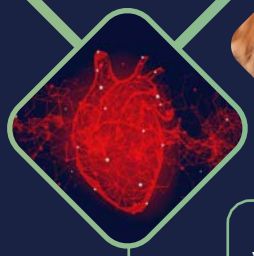
Genetic secrets of icefish survival discovered





JUL

Heart Cell Atlas 2.0 published



Genomics shows importance of vaccination

Sanger Institute awarded UK Science Council Employer Champion status

Gorillas' evolutionary history explored

AUG

Genomics reveals yolk sac's vital roles in early human development

Key human bladder cancer gene pinpointed in cats and dogs

Diversity of human Y chromosome revealed

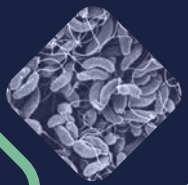
SEP

1,000 new reference genomes generated for Tree of Life

Professor Saheer Gharbia joins the Sanger Institute's Genomic Surveillance Unit

Genes powering Yemen cholera outbreak revealed

Genetic recycling explains explosive fish diversity

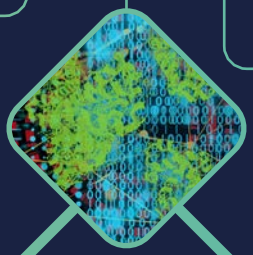


OCT

Contagious cancers in cockles explored with genomics

First genomic surveillance of antibiotic-resistant *Klebsiella* bacteria in Ghana

Psoriasis not caused spontaneous DNA changes in skin cells



Generative and Synthetic Genomics research programme launched

NOV

Pocket-sized DNA sequencers track malaria drug resistance in near real-time

Roots of cancer risk from gene editing discovered

Quotient Therapeutics launched to translate the Sanger Institute's cancer research into new treatments

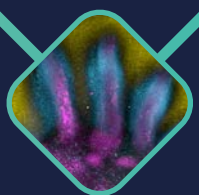
UK Biobank opens up data of 500,000 whole human genomes

DEC

Impact of 12,776 DNA changes to gene associated with neurodevelopmental disorders studied

CellHint tool unifies single-cell data

First spatial map of human limb development published



At a Glance

Cellular

191 organoid models banked

4,260 flow cytometry experiments

453 spatial slides stained and processed

220 cell lines differentiated into neural progenitors

108 CRISPR screens

DNA sequencing

An average of **10,724bn** DNA bases a day were read

Every **12.1 mins,** the equivalent of one gold-standard (30x) human genome was read

1,463 species were sequenced*

Every week, the equivalent of **822** gold-standard human genomes were read

Compute

29,000 approx. total number of compute cores

16,000 approx. high-performance compute cores in Sanger HPC clusters

13,000 approx. cores as part of Sanger's private cloud environments

123PB approx. useable storage

**Full year figure (Nov 2022-2023)*

Year in Numbers

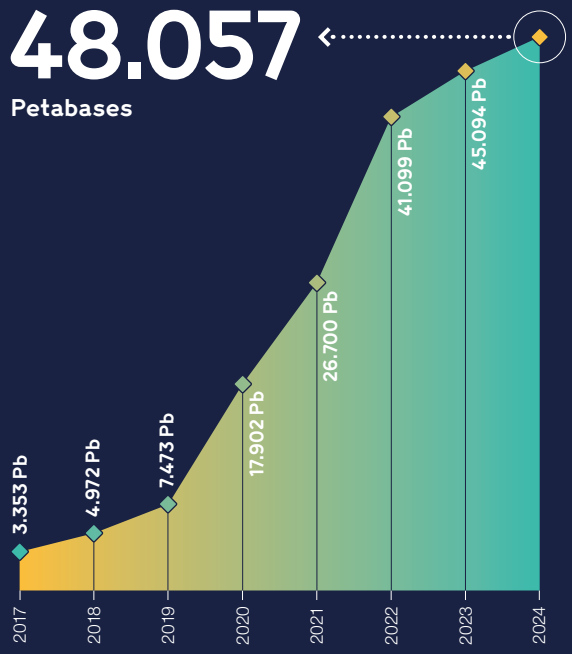
449 research articles and reviews published in 2023

1,484,072 citations of our papers and reviews to date

10,080 published articles and reviews to date

Sanger Institute authored papers and reviews are **4.13** times more cited than the world average*

*Data for the last 5 years, 2019-2023. Data correct as at 10th April 2024



Cumulative total of DNA sequenced by the Sanger Institute

Our research in numbers

Where Sanger staff are from*

How much DNA was sequenced?

Sanger staff are made up of **73** different nationalities

*As of December 2023

Our work

We develop and deliver cutting-edge data generation and analysis

Cancer, Ageing and Somatic Mutation

We study the genetic changes in normal tissues to better understand their causes and consequences on ageing and disease. We conduct large-scale cellular experiments to discover how mutations affect cancer development.

Page 12

Human Genetics

We combine population-scale genetics and cell-based studies with clinical data to identify and study severe developmental disorders. We study the biology of health and disease in the immune system and blood cells through large-scale cell-based experiments.

Page 20

Cellular Genetics

We map cells in the human body at scale by combining single-cell genomic profiling, 3D imaging and computational methods. We investigate the dynamic changes that occur within cells, tissues, organs and organisms during development, health, disease and ageing.

Page 16

Parasites and Microbes

We study the genomics and evolution of disease-causing organisms and the human microbiome. We build networks at scale to help monitor infectious diseases and the effects of health policies worldwide, identifying the drivers of drug, vaccine and insecticide resistance to guide health planning.

Page 24

Tree of Life

We are building the library of life. We produce high-quality reference genomes to explore the evolution, function and interactions of life on Earth. We seek to aid conservation and biodiversity work and provide the underpinnings of a new way of doing biology.

Page 28


wellcome
sanger
institute

Cancer, Ageing and Somatic Mutation

We study the genetic changes in normal tissues to better understand their causes and consequences on ageing and disease. We conduct large-scale cellular experiments to discover how mutations affect cancer development.

In this section

Our findings show it is good to have multiple animal models for bladder cancer biology representing different driver genes to capture distinct patient populations.

Dr Louise van der Weyden
Senior staff scientist at the Wellcome Sanger Institute

1

Human cancer genes found in cats and dogs

Sanger Institute scientists have identified key genes responsible for spontaneous cancer in animals. Their discovery could spur on targeted therapies for both animals and humans.

In humans, muscle-invasive bladder cancer (MIBC) is highly aggressive and associated with a poor prognosis. Treatment options are limited. Previous research has linked roughly 60 genes to human bladder cancer – making targeted drug development difficult.

Cats and dogs develop bladder cancer with similarities to human MIBC, offering a valuable alternative approach for studying the disease. Researchers from the Wellcome Sanger Institute and the University of Guelph in Canada sequenced the DNA of cancer tumours in cats and dogs and compared the findings with data from human bladder cancers. Their study is the most extensive genetic sequencing of canine bladder cancer and the first-ever DNA sequencing of feline bladder cancer.

The team's analysis found that three genes – *TP53, FAT1* and *NRAS* – were mutated in both human MIBC and in cat bladder cancer. In addition, two genes – *ARID1A* and *KDM6A* – were mutated in both human and dog bladder cancer. These cancer-related genes, conserved across multiple species, are likely to be biologically relevant for disease development.

This research highlights how studying tumours from cats and dogs that naturally develop bladder cancer can generate insights into a particularly aggressive type of human bladder cancer. Using genomics, the researchers have uncovered a unique view into the hidden similarities and differences of cancer between species. This will help scientists to drill down to the key cancer-driving genes and prioritise drug targets for patients.

Reference
Wong K. *et al. Genome Biology* 2023; **24:** 191.

2

Why Singapore has less skin cancer than UK

Researchers have discovered differences in the skin's ability to protect itself from ultraviolet (UV) radiation between UK and Singaporean populations.

Keratinocyte skin cancer, including basal cell and squamous cell carcinomas, is the most common type of skin cancer worldwide. It is driven by a person's lifetime exposure to UV radiation, which mutates a cell's DNA and causes disease. Despite Singapore having three times the UV strength compared to the UK, keratinocyte cancer rates are 17 times higher in the UK.

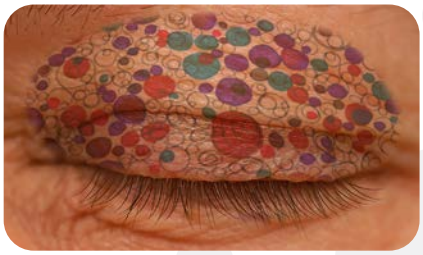
Researchers from the Sanger Institute and the Skin Research Institute of Singapore collected healthy eyelid tissue samples from patients undergoing routine surgery. To compare the mutational landscapes between populations, they sequenced DNA from the skin cells, focusing on 74 genes commonly mutated in cancers.

Analysing over 400 samples, they found that cells in Northern European skin types in the UK accumulated four times as many cancer-associated DNA changes on average, including mutations in known cancer genes such as *TP53*. By 60 years of age, nearly every cell in the UK donor skin had a mutation in a cancer-associated gene. Copy number variations – where areas of the genome had been lost or duplicated – were also more prevalent in Northern European skin types in the UK (13 per cent) compared with Singapore (1 per cent).

By analysing the patterns of DNA changes, known as mutational signatures, the team found that most of the DNA mutations in UK skin are caused by UV, whereas in Singapore mutations are due to ageing processes. The genetic features in the UK skin samples shared similarities with cancer and help explain the high keratinocyte rates.

The results show the importance of comparing populations with varying cancer rates to understand cancer evolution, susceptibility and protective mechanisms against disease.

Reference
King C. *et al. Nature Genetics* 2023; **55:** 1440-47.



400+
skin samples studied

17x
more keratinocyte cancers in the UK than Singapore

3x
stronger average UV in Singapore than the UK

3

Exploring how cockles catch cancer

The genomes of transmissible cancers in cockles — marine cancers that can spread through the water — have been sequenced, unearthing new insights into how these cancers have moved through animal populations for hundreds, possibly thousands, of years.

The common cockle (*Cerastoderma edule*) inhabits the coasts of Europe and north-west Africa and is often harvested for food. It is susceptible to a contagious form of cancer, bivalve transmissible neoplasia (BTN). The disease is spread by living cells, passing from one cockle to the next through seawater. BTN causes a disease similar to leukaemia and is usually lethal. It poses an ecological threat to coastal environments and commercial aquaculture, yet until now it has remained largely unexplored.

To study the origins and evolution of BTN, researchers from the Wellcome Sanger Institute and the Universidade de Santiago de Compostela in Spain first produced a high-quality reference genome for *C. edule*.

The scientists also collected 7,000 cockles at 36 locations from 11 countries. They used the reference sequence to catalogue the genomic variation in 61 BTN tumours.

Combining histopathology, cytogenetics and sequencing of whole genomes and transcriptomes, their study illuminates the evolutionary history of marine leukaemias. The findings suggest that BTN cancers emerged centuries or even millennia ago.

The team also discovered that BTN tumour genomes are highly unstable. The number and size of chromosomes varied markedly, both between and within tumours. Some cells contained as few as 11 chromosomes and others up to 354. The normal number in cockle cells is 38.

This degree of chromosomal instability is far greater than any seen in human tumours. It suggests that a stable genome is not needed for the long-term survival of these transmissible cancers.

Understanding how BTN cells tolerate this instability could help inform new approaches to cancer treatments.

Reference
Bruzos A.L. *et al. Nature Cancer* 2023; **4:** 1575-91.

7,000
cockles from 36 locations in 11 countries were studied

61
BTN tumours catalogued

38
chromosomes in normal cockle genome

11-354
chromosomes in cockle genomes with BTN

4

Why some child cancer therapies last longer

Genomics has shown why some children with leukaemia have a longer remission than others after receiving cutting-edge CAR T-cell therapy. A research collaboration combined expertise in novel immune therapy design and state-of-the-art computational analysis to identify a genetic signature that indicates which CAR T-cells will be the most effective in the long term.

In recent years, chimeric antigen receptors (CAR) T-cells – genetically engineered immune cells designed to target an individual's leukaemia – have become an established treatment option for children with relapsed or incurable rare forms of the disease. A key factor in the treatment's effectiveness is the durability of CAR T-cells in the body. However, little has been known about what makes these cells persist or not.

A research team from Great Ormond Street Hospital, the Wellcome Sanger Institute and the UCL Great Ormond Street Institute of Child Health worked with 10 families for up to five years after their child's CAR T-cell treatment to study why some CAR T-cells stay in the body.

Using high-throughput single-cell gene expression and T-cell receptor sequencing, the team systematically examined the characteristics of individual T-cells. They studied T-cells from the treatment infusion and sequential samples from blood and bone marrow.

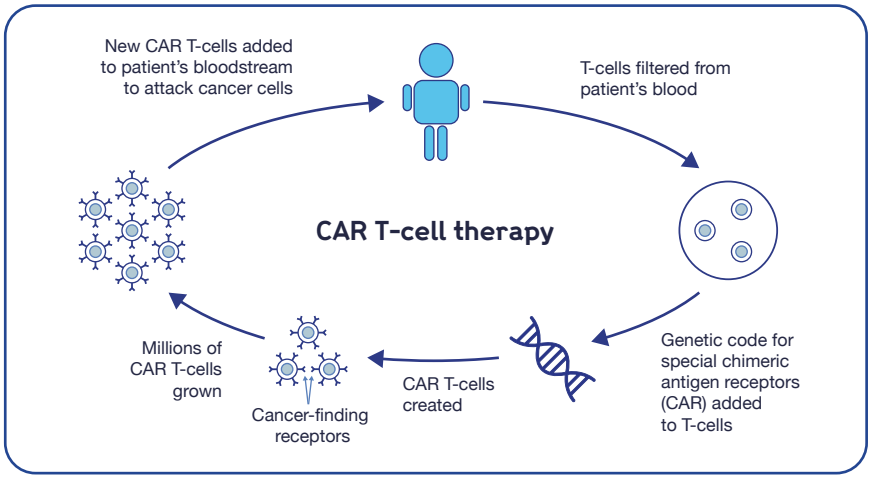
They found that long-lived CAR T-cells developed a distinct set of receptors on their surface and a distinct pattern of gene activity.

This 'signature' was consistent in all children with a prolonged treatment response, and in various T-cell subtypes. It was also observed in two adults with leukaemia who achieved decade-long remissions through a different CAR T-cell product.

Expanding their study to encompass diverse T-cells from healthy and diseased tissues, the researchers found that the signature was unique to long-lived CAR T-cells. This implies that the signature not only serves as a marker for persistent cells but may also be integral to their anti-cancer functions.

The researchers ultimately aim to understand if there is a way to spot, or even create, CAR T-cells with inherent longevity, potentially improving this therapy.

Reference
Anderson N.D. *et al. Nature Medicine* 2023; **29:** 1700-09.



Profile

Dr Yvette Hooks
Senior Technical Specialist in the Cancer Support Team

Yvette has worked as a histologist at the Sanger Institute for 18 years and currently applies her skills to prepare samples for laser capture micro-dissection. While most of the tissues are from patients with cancer, she has also prepared samples from tiger, giraffe, naked mole rat and other species.

"I really enjoy contributing to Sanger's amazing cancer research and hearing colleagues say that their studies couldn't happen without my histology support."

5

Genetic code of rare kidney cancer reveals drug target

Researchers have uncovered the genetic drivers of reninoma, a rare form of kidney cancer. The team has also revealed a new drug target that could serve as an alternative treatment when surgery is not possible.

Reninoma is one of the rarest tumours, with just 100 cases documented worldwide to date. Although it can usually be cured with surgery, some instances of the disease can cause severe hypertension, and in others the cancer can metastasise and spread. There are no existing drug treatments for reninoma; management involves surgery alone. Until now, the genetic changes that drive this cancer have been unknown.

In a collaborative study between the Wellcome Sanger Institute, Great Ormond Street Hospital and The Royal Free Hospital, researchers examined two cancer samples — from a young adult and a child — with advanced genomic techniques, including whole genome and single nuclear RNA sequencing.

They found that a specific error in the genetic code of a known cancer gene, *NOTCH1*, is behind the development of this rare cancer. Re-analysis of previously published reninoma genetic data verified their finding.

Their work suggests that using existing drugs that target *NOTCH1* (NOTCH1 inhibitors) could be a potential treatment for reninoma patients when surgery is not a viable option. Such NOTCH1 inhibitors are currently used to treat other types of cancer.

While systematic efforts to study cancer genomes are largely complete, this work demonstrates that efforts to map and understand the genomics of rare tumours are still needed to complete the compendium of the human cancer genome.

Reference
Treger T.D. *et al. Nature Communications* 2023; **14:** 5826.

6

Cancer risk from gene therapy affected by cell competition

Gene therapy holds immense potential to cure some genetic conditions. However, in some trials of gene therapy for sickle-cell disease, a small number of patients developed blood cancer. To understand why, Wellcome Sanger Institute researchers and their collaborators analysed genetic changes before and after gene therapy. They uncovered "cell competition" among blood stem cells – a phenomenon that could influence cancer risk.

Gene therapy involves modifying a patient's own stem cells outside the body, correcting the faulty gene responsible for the condition and re-introducing the edited cells into the body. However, a small number of patients with sicke-cell disease have developed blood cancer following gene therapy, though the precise cause is not clear.

To investigate, the team sequenced the genomes of thousands of blood stem cells from six people who underwent gene therapy. They examined cancer-related genetic changes, comparing blood stem cells before and after the procedure.

The analysis revealed a post-treatment accumulation of cells carrying genetic mutations that affect blood cell growth. These DNA mutations have been previously seen at higher levels in elderly individuals and in those with certain blood cancers.

Importantly, the researchers showed that the gene therapy treatment itself is not the likely cause of new DNA mutations in blood stem cells. Instead, the process of genetically modifying the stem cells outside the body and re-transplanting them makes blood stem cells that already have these mutations more prominent, amplifying their influence on the blood and immune systems.

While the relationship between these findings and the risk of blood cancers is not yet fully understood, the work indicates some preventative measures to optimise gene therapy. It also underscores the need for long-term monitoring of genetic changes in stem cells for patients undergoing gene therapy.

Reference

Spencer Chapman M. *et al. Nature Medicine* 2023; **29:** 3175-83.

Profile

Dr Jo Fowler
Senior Staff Scientist in the Jones research group

Jo studies the lining of the oesophagus to explore how genetic and environmental changes affect cell growth and behaviour. The Jones team uses high depth sequencing and a new method developed in house for growing primary tissue into epithelioids: 3D culture models that can be maintained for up to a year without passaging.

"Normal epithelia is a bit of a battleground with a patchwork of mutant cells. I've heard this described like 'a game of clones' - warring families of cells, vying for space."

To explore this cellular battle, the Jones team uses gene editing to study the effects of mutations within epithelioid cultures. They also change the environment with UV light or oxidative stress, and analyse genomes, gene activity and regulation in individual cells to see how the cells respond. In addition, live-imaging techniques provide real-time data on the competition between the cells.

"Our aim is to understand how these cells compete in this microscopic landscape and how that relates to cancer, or cancer risk. If we can understand that, there might be places to intervene and prevent disease."

Cellular Genetics

We map cells in the human body at scale by combining single-cell genomic profiling, 3D imaging and computational methods. We investigate the dynamic changes that occur within cells, tissues, organs and organisms during development, health, disease and ageing.

In this section

This is the first time that the yolk sac has been profiled at a single cell level, giving us an incredible amount of information on how this primary organ works in the first stages of human development.

Issac Goh
Visiting scientist at the Wellcome Sanger Institute

1

Multiple organ functions of the yolk sac revealed

How the human yolk sac supports early embryonic development and the prenatal immune system has been mapped by researchers from the Sanger Institute, Newcastle University and the Cambridge Stem Cell Institute. They charted how it provides the functions of the liver, bone marrow and kidneys, before these organs are formed, and produces immune and red blood cells.

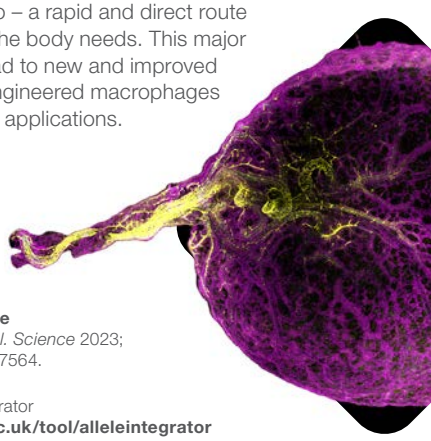
During the first weeks of pregnancy, the yolk sac develops within the uterus, providing nutritional and metabolic support. To map the gene activity, state and functions of individual cells in the sac, the research team used single-cell sequencing techniques on 10 samples spanning four to eight weeks post conception.

The results were integrated with existing datasets to analyse 169,798 cells. Two-dimensional and 3D imaging techniques were used to provide spatial context and validate the single-cell data.

Comparing the data with those from rabbits and mice showed important differences, with implications for future studies. Unlike in mice, the human yolk sac endoderm (the innermost cells) produces coagulation proteins and blood cell growth factors. In addition, the human yolk sac is the dominant source of early red blood cell production, whereas the liver also plays a role in mice.

This is the first time that these multiple functions, delivered by the yolk sac acting as a temporary organ outside the embryo, have been seen.

The researchers also uncovered an accelerated way of producing macrophages, an important immune system component. This macrophage development pathway appears unique to the early embryo – a rapid and direct route to get the cells the body needs. This major finding could lead to new and improved production of engineered macrophages with therapeutic applications.



Reference
Goh I. *et al. Science* 2023; **381:** eadd7564.

alleleIntegrator
sanger.ac.uk/tool/alleleintegrator

2

Tool reveals how people respond to COVID-19

A global research team has developed a new statistical tool that allowed them to capture dynamic gene activity changes during the immune response. It also enables researchers to identify genes and molecular pathways associated with disease risk that have previously been too complex to detect.

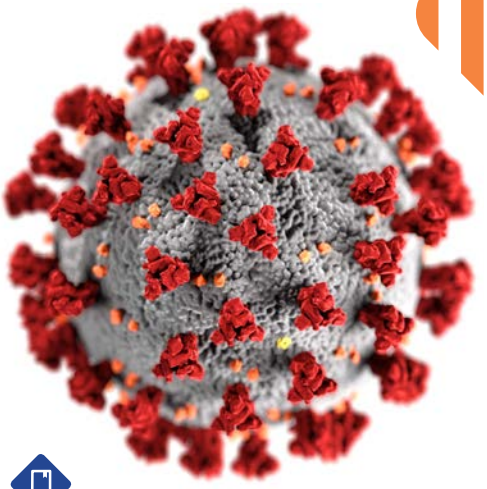
There is a wide variation in how people respond to COVID-19 and other infections, from mild symptoms to severe illness. Part of this may be due to genetic differences that regulate gene activity in the immune response. To explore disease variation, researchers from the Sanger Institute, the National Center for Child Health and Development in Japan and Tel Aviv University studied expression quantitative trait loci (eQTLs). These DNA differences affect whether it is active in a cell, or not, and to what level.

The team measured gene activity in tens of thousands of fibroblast cells – that form connective tissues – from 68 donors, using single-cell RNA-sequencing methods. The cells were triggered by immune stimuli in the laboratory, and gene activity was measured over time. They analysed the data using a new tool, GAuSsian Processes for Association mapping leveraging Cell HeterOgeneity (GASPACHO), generating a unique single-cell map of gene activity in the innate immune response.

Using GASPACHO, the team discovered 1,275 eQTLs. Many of these were near locations in the genome known to be related to immune disease susceptibility.

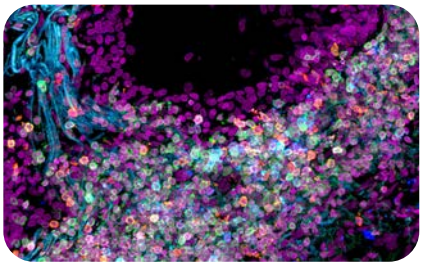
Using additional experimental techniques and datasets, the team found an association between a particular genome location and COVID-19 severity. This location reduces the effects of the *OAS1* gene, impairing viral clearance from cells.

GASPACHO provides a unique framework for uncovering genetic variants that control gene activity in a cell. It will now be used on additional datasets to uncover further genetic mechanisms underlying disease, potentially revealing new drug targets.



Reference
Kumasaka N. *et al. Nature Genetics* 2023; **55:** 1066-75.

3



Single-cell atlas of the whole human lung

Researchers at the Sanger Institute, Helmholtz Munich, University Medical Center Groningen and their collaborators have published the largest and most comprehensive single-cell map of the human lung. The atlas reveals rare cell types and highlights cellular differences between healthy people. Their work also found common cell states between lung fibrosis, cancer and COVID-19, offering new ways of understanding lung disease.

Lung research has benefited greatly from recent single-cell studies that show which genes are active in individual cells. However, the number of samples and individuals included per study has limited research so far. To address this gap, the team of researchers used advanced machine learning to pool and integrate datasets from every major single-cell RNA-sequencing lung study published, spanning 2.4 million cells from 486 individuals.

Combining the datasets allowed the researchers to uncover new details about the cells within them. This process included annotations of rare and previously undescribed cell types, as well as new cell states. They used the unprecedented complexity of the atlas to identify sets of genes that are associated with factors such as location in the lung, age, sex, BMI and smoking status.

The researchers also took datasets from 10 different lung conditions and projected these onto the healthy data to understand disease states. They discovered that different lung diseases shared common immune cell states. For example, a subset of macrophages shared similar gene activity in lung fibrosis, cancer and COVID-19. Such cells could play a similar role in scar formation in the lung in each disease and could be potential therapeutic targets.

The lung atlas is publicly available together with an online platform for automated mapping of new data. Taken together, the Human Lung Cell Atlas is a universal reference for single-cell lung research that promises to accelerate future studies into pulmonary health and disease.

Reference
Sikkema L. *et al. Nature Medicine* 2023; **29:** 1563-77.

Profile

Kerstin Meyer
Principal Staff Scientist working on the Human Cell Atlas within the Cellular Genetics programme

Kerstin leads the Human Cell Atlas projects characterising the cellular landscape of the lung and respiratory system in human development and over the life span. The research also explores the effects of diseases such as asthma and COVID-19.

The Human Cell Atlas aims to map out all of the hundreds of different types of cells in the human body using single-cell sequencing technology. This approach allows researchers to quantify which genes are active inside individual cells, and when, giving unprecedented insights into their functions within tissues and organs.

Read more at sangerinstitute.blog

4

‘Mini placentas’ give new insights into pregnancy

As part of the Human Cell Atlas initiative that is mapping every cell type in the human body, researchers have revealed what happens in the early stages of placental development, a process crucial for a successful pregnancy. Their findings have created an in-depth picture of how the placenta develops and communicates with the uterus.

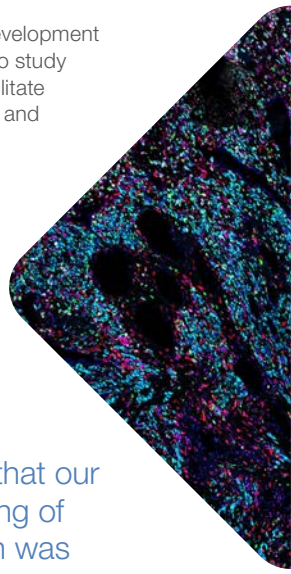
The formation and embedding of the placenta into the uterus – known as placentation – is crucial for nurturing and protecting the foetus during pregnancy. The placental trophoblastmucosal layer of the uterus transforms maternal arteries to supply oxygen and nutrients to the foetus. Defects in this process underlie common disorders such as pre-eclampsia. Despite pregnancy disorders causing illness and death worldwide, many of the processes involved are not fully understood.

To explore normal and disordered placentation at a molecular level, researchers from the Sanger Institute, University of Cambridge, the Friedrich Miescher Institute for Biomedical Research, the European Molecular Biology Laboratory and others applied single-cell genomics and spatial transcriptomics technologies to a rare historical set of samples.

The team uncovered the full trajectory of placental and trophoblast development in unprecedented detail. Their findings suggest what could go wrong in disease and describe the involvement of multiple populations of cells, including maternal immune and vascular cells. The techniques also allowed the researchers to understand how trophoblast cells communicate with the maternal environment around them.

The team compared their results to placental trophoblast organoids, sometimes called ‘mini-placentas’, grown in the laboratory. Most of the cells identified in the tissue samples were also seen in the organoids. However, some later populations of trophoblasts were not seen in the laboratory, suggesting that they only form in the uterus after receiving signals from maternal cells.

This research will help in the development of effective laboratory models to study placental development and facilitate new ways to diagnose, prevent and treat pregnancy disorders.



Reference
Arutyunyan A. *et al. Nature* 2023; **616:** 143-51.

> Our research shows that our previous understanding of placental implantation was incomplete, and that the maternal uterine cells release communication signals to encourage placental growth.

Dr Roser Vento-Tormo
Group Leader at the
Wellcome Sanger Institute

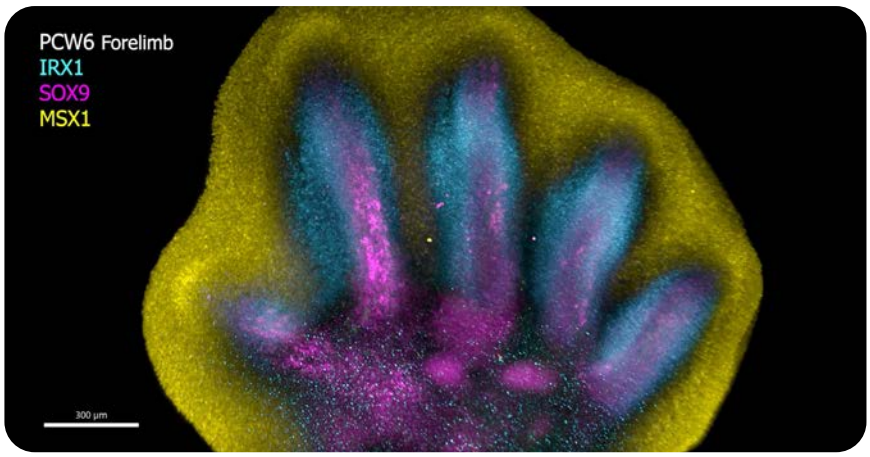
5

First spatial map of human limb development

Human fingers and toes do not grow outward; instead, they form within a larger foundational bud. When the intervening cells recede, the digits beneath are revealed. This is one of many processes captured for the first time by Sanger Institute scientists as part of a spatial cell atlas of the entire developing human limb. The map, which is resolved in time and space, reveals vital insights into limb formation and explains syndromes found at birth.

In early human development, limbs initially emerge as undifferentiated cell pouches on the sides of the body, without a specific shape or function. But after eight weeks of development, they are well differentiated, anatomically complex and recognisable as limbs, complete with fingers or toes. This requires a very rapid and precise orchestration of cells. Small disturbances to this process can have a downstream effect. This is why variations in the limbs are among the most frequently reported syndromes at birth, affecting approximately one in 500 births globally.

While limb development has been extensively studied in mice and chickens, how well these animals reflect human limb development has been unclear.



As part of the Human Cell Atlas initiative to map every cell type in the human body, the team analysed patterns of gene activity in tissues between 5-9 weeks of development. Using cutting-edge single-cell and spatial transcriptomics, the scientists found extensive diversification of cells from a few progenitors to a myriad of separate cell states, including several new cell populations.

The team identified specific programmes of gene activity, which are activated at certain times and in specific areas, that shape the forming limbs. Further investigation revealed which genes, when disrupted, are associated with specific limb syndromes such as short fingers or toes, or extra ones.

The team also confirmed that many aspects of limb development are shared between humans and mice. The atlas provides an openly available resource that captures the intricate processes governing the limbs’ rapid development during the early stages of development.

Reference
Bao Z. *et al. Nature* 2023.
DOI: 10.1038/s41586-023-06806-x

6

Heart cell atlas reveals origins of heartbeat

Researchers have produced the most detailed and comprehensive map of the human heart to date, including the specialised tissue of the cardiac conduction system – where the heartbeat originates. Their study also developed a new tool that uncovers the effects of drugs on heart rate.

Researchers at the Wellcome Sanger Institute and the National Heart and Lung Institute at Imperial College London have assembled a human Heart Cell Atlas. Cardiovascular diseases are the leading cause of death globally, and the atlas paves the way for therapies to boost cardiac health and develop targeted treatments for arrhythmias.

Charting eight heart regions, the team's researchers found 75 different cell states including the cells of the cardiac conduction system – responsible for the heartbeat – not previously understood in such detail in humans. Using spatial transcriptomics to map where cells sit within a tissue, the scientists were able to understand how these cells communicate with each other for the first time.

The team also presented Drug2cell, a computational tool that can predict drug targets and side-effects. It leverages single-cell profiles and the 19 million drug-target interactions in the EMBL-EMBL's European Bioinformatics Institute database.

Unexpectedly, this tool identified that pacemaker cells express the target of certain medications, such as GLP1 drugs. These drugs are used for diabetes and weight loss and are known to raise heart rate. This study suggests that the increase in heart rate might be partly due to a direct action of these drugs on pacemaker cells, a finding the team also showed in an experimental stem cell model of pacemaker cells.

The study also found a close relationship between conduction system cells and glial cells. Glial cells are part of the nervous system, mostly found in the brain and explored very little in the heart. The research suggests that glial cells are in physical contact with conduction system cells and may play an important supporting role: communicating with the pacemaker cells, guiding nerve endings in their direction and supporting their release of glutamate, a neurotransmitter.

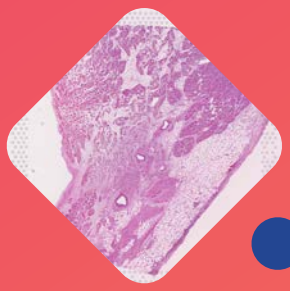
Another key finding was an immune structure on the heart's outer surface. This contains plasma cells, which release antibodies to prevent infection from the lungs.

The team also identified a cellular niche in healthy hearts that produces a hormone, currently used as an early warning sign of heart failure. This opens the door to new therapeutic opportunities to prevent disease progression by targeting specific cell types.

The atlas provides new clarity on cardiac electro-anatomy and immunology and a suite of computational approaches that can be applied to other tissues and organs.

Reference
Kanemaru K. *et al. Nature* 2023; **619:** 801-10.

Drug2cell
sanger.ac.uk/technology/drug2cell/



Profile

James Cranley, PhD student and trainee cardiologist, and Kazumasa Kanemaru, Postdoctoral Fellow, in the Teichmann research group.

James and Kazumasa worked together to help build the heart cell atlas – version 2.

James was drawn to the project after the first Heart Cell Atlas was published. At the time, he was treating patients with cardiac arrhythmias and other types of heart rhythm disturbances.

Kazumasa is a qualified doctor and trained as an immunologist in Japan. Before he joined the Sanger Institute he was researching the function of single molecules in mouse dermatitis models.

Their complementary skill sets helped to uncover new insights into the development and function of the human heart. James' knowledge of cardiology supplemented Kazumasa's molecular biology experience to enhance their understanding of the niches of cells they uncovered. Together, they discovered one of these previously unknown, unclarified niches, the sinoatrial node, responsible for the regular and coordinated electrical activation of the heart, which is composed of pacemaker cells, fibroblasts and glial cells.

Read more at sangerinstitute.blog

Human Genetics

We combine population-scale genetics and cell-based studies with clinical data to identify and study severe developmental disorders. We study the biology of health and disease in the immune system and blood cells through large-scale cell-based experiments.

In this section

13,500
families across the UK and Ireland studied

1

5,500 children helped by major genomics study

A genomics study of 13,500 families across the UK and Ireland has transformed the lives of 5,500 children by providing crucial diagnoses of rare genetic conditions. Sanger Institute scientists collaborated with NHS clinicians to carry out the Deciphering Developmental Disorders (DDD) study. By applying genomic research to healthcare, and sharing the information with diagnostic centres worldwide, this research may help revolutionise patient care and improve rare disease diagnosis rates globally.

Rare genetic conditions collectively affect up to 1 in 17 people, but they are challenging to study owing to their complexity. However, an accurate diagnosis is vital to guide treatment and improve patients' quality of life.

Addressing this challenge, the researchers conducted large-scale sequencing and analysis of thousands of children with undiagnosed developmental disorders, and their parents.

In total, the team have provided diagnoses for 5,500 children, identified 800 genes related to these rare conditions and discovered 60 new genetic diseases. The families that have received a diagnosis are better able to understand their child's situation, feel less isolated and become empowered to work together to create support networks to share experiences.

Whilst most of the conditions were caused by new genetic changes that were not inherited from parents, the scientists also identified prenatal risk factors such as premature birth. They also highlighted differences in diagnosis rates among different populations – for example, lower rates in non-Europeans – underscoring the importance of including diverse research participants.

The work demonstrates the value of applying large-scale genomics to medicine and healthcare. As the genomic technologies and techniques employed by the DDD study move into routine healthcare, more insights into the genetics underpinning rare developmental disorders will continue to be discovered, further enhancing clinical diagnoses.

Reference
Wright C.F. *et al. New England Journal of Medicine* 2023; **388:** 1559-71.

2

Improving gene editing with machine learning

Sanger Institute scientists working with the University of Tartu, Estonia have developed a new machine learning tool that increases the effectiveness of 'prime editing' — an advanced genome editing technique. This tool could help power the treatment of genetic diseases such as cystic fibrosis by correcting harmful variations in people's genes.

Prime editing was developed as an evolution of the pioneering CRISPR-Cas9 technique that uses the enzyme Cas-9 as 'molecular scissors' to precisely remove, add or replace sections of DNA. It is more efficient, accurate and safer than previous methods of genome editing. Yet prime editing tools have limited clinical use due to varying levels of success in their edits.

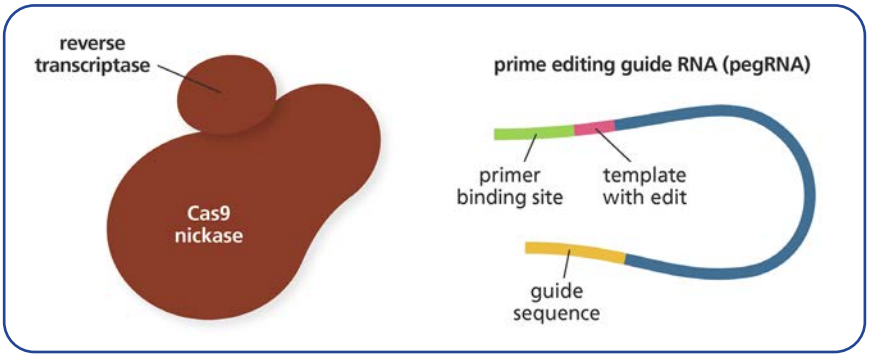
The collaboration applied machine learning to understand the factors that affect prime editing success. First, the team created 3,604 DNA sequences of varying lengths and inserted them into different human cells using prime editors. Next, the researchers trained the machine learning algorithm using the success of each prime editor in different scenarios to make predictions.

It revealed that prime editing effectiveness was influenced by DNA sequence length and the repair mechanism used. Based on this, researchers can now design prime editors tailored to each genetic condition. This reduces the 'trial and error' approach of traditional genome editing and potentially accelerates the transfer of these treatments into the clinic.

This study demonstrates the value of combining machine learning with genomics to tackle challenges in healthcare, and that improved prime editing may generate faster and more effective genetic disease treatments. It could also guide new interventions for previously untreatable genetic conditions. The Sanger Institute will continue to apply the technology to a range of human genetic diseases to discover if and how they may be treated with prime editing.

Reference
Koeppel J. *et al. Nature Biotechnology* 2023; **41:** 1446-56.

> The potential of prime editing to improve human health is vast, but first we need to understand the easiest, most efficient and safest ways to make these edits. It's all about understanding the rules of the game, which the data and tool resulting from this study will help us to do.
>
> **Dr Leo Parts**
> Group Leader at the
> Wellcome Sanger Institute



3

Rare disease study reveals treatment target

Sanger researchers have shed new light on KPTN-related disorder (KRD). The team have created new models to study the disease and the molecular pathways that cause it. Their work places KRD in the context of a family of disorders with a common underlying mechanism. The promising treatment of related conditions means that there could soon be an effective medication for KRD.

KPTN-related disorder (KRD) is a rare genetic disease characterised by macrocephaly – enlarged skull and brain, developmental delay, intellectual disability, seizures and anxiety. Sanger Institute researchers have previously identified genetic variants in the *KPTN* gene that underlie KRD.

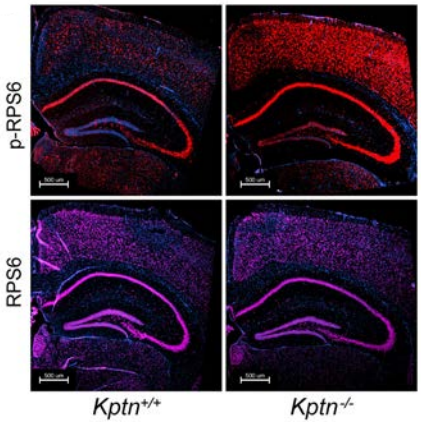
To understand how changes in the *KPTN* gene cause KRD, the team created two new models of the disease. The first was a genetically engineered mouse lacking a functioning *KPTN* gene. The second used genetically engineered stem cells to mimic those of KRD patients, which the team grew into nerve cells in the laboratory.

The mice showed KRD symptoms including macrocephaly, behavioural changes and cognitive deficits. Molecular and structural analyses also highlighted further brain differences to healthy mice that resulted from unusual postnatal brain development. These results provide key insights into the disease and identify the critical time window for potential treatment.

Analyses in both mice and human cells demonstrated that the *KPTN* gene is involved in regulating the mTOR signalling pathway. mTOR is an enzyme that coordinates cell growth and metabolism with environmental inputs. Disruptions to mTOR signalling are associated with cancer, diabetes, epilepsy and developmental delay. The team's work not only implicates mTOR signalling in KRD, but also other brain-overgrowth conditions.

Importantly, the faulty mTOR signalling in mice was corrected by treatment with an mTOR-inhibiting drug. Given that mTOR inhibitors successfully treat other conditions, this may accelerate development of a treatment for KRD.

Reference
Levitin M.O. *et al. Brain* 2023; **146:** 4766-83.



Profile

Dr Laura Fachal
Senior Staff Scientist in the Anderson research group

Laura studies the genetics of Inflammatory Bowel Disease (IBD) using genome wide association studies (GWAS) and laboratory experiments. Along with the rest of the team, her aim is to build a complete genetic map for complex immune-mediated disease to help identify new drug targets.

"I like that we have both a computational team and a laboratory team together. I can input into experimental design because that feeds through into the types of data we get at the end. I can appreciate how the data are produced, which is vital for analysis."

Read more at sangerinstitute.blog

4

Genetics rules out causes of psoriasis

Sanger scientists and collaborators have shown that psoriasis is not caused or spread by genetic changes in skin cells. The results will steer research to explore other reasons for the development of psoriasis and may aid researchers in searching for new treatments.

Psoriasis is a chronic, inflammatory autoimmune disease that results in itchy or flaky skin patches. Scientists do not know what causes psoriasis, despite it affecting 125 million worldwide.

One theory is that somatic mutations – genetic changes to non-reproductive cells – cause psoriasis. These can result from DNA replication mistakes, chemical exposure and other environmental factors. Often the changes are harmless, but if they give the cell a survival advantage, they become driver mutations and spread.

To test this theory, scientists sequenced the DNA of skin samples from 111 people with psoriasis, comparing patches of psoriatic and unaffected skin. Their genetic analyses revealed no significant differences between the affected and healthy skin samples, suggesting that psoriasis is not caused by somatic mutations. However, the research uncovered four driver mutations in both psoriatic and non-psoriatic skin. This insight enhances the scientific community's broader understanding of skin mutations.

The researchers also found a mutational signature – a unique pattern of change to DNA – linked to psoriasis medications called psoralens. These chemicals were previously used in sunscreen, which may explain why the signatures were also found in people never treated with psoralens. Further research is needed to understand the impact of these mutations and environmental exposure.

By ruling out a causal role for somatic mutations in psoriasis, this study will help focus research into other possible causes and treatments. The results also demonstrate that research into somatic mutations, when combined with environmental data, provides a powerful way to understand human health and disease.

Reference
Olafsson S. *et al. Nature Genetics* 2023; **55:** 1892-1900.

5

Health impact of 12,776 changes to one gene charted

Researchers from the Sanger Institute and the University of Cambridge have made significant strides in understanding the functional impact of genetic changes within the *DDX3X* gene. The gene is linked to neurodevelopmental disorders and certain cancers. The scientists created of an extensive map illustrating how all possible genetic changes in the *DDX3X* gene can affect health.

DDX3X-related neurodevelopmental disorder is typically associated with intellectual disability, developmental delays and seizures. Changes in the *DDX3X* gene have also been linked to cancer, although the molecular mechanism behind this have been unclear. Despite the prevalence of genetic sequencing to diagnose rare disease understanding the relevance of most variants in disease-associated genes remains difficult.

Focussing on *DDX3X*, the team used saturation gene editing to directly study 12,776 genetic changes in the gene. They made the genetic changes to cells grown in the laboratory and studied the resulting effects. A quarter of these alterations disrupted the normal functioning of the DDX3X protein in the cells.

To understand the effects of the genetic alterations beyond the laboratory, they compared the experimental data with health data from UK Biobank and from databases of genetic changes seen in people with neurodevelopmental disorders and cancer.

The results are an extensive map, illustrating how all possible genetic changes in the *DDX3X* gene can affect health. While their method incorporated machine learning, it outperformed solely AI-based methods and revealed the significance of 90 per cent of previously unexplained genetic changes.

The researchers showed the genetic changes in cancer prevent the DDX3X protein from functioning correctly, providing crucial insights for developing new cancer treatments targeting the gene.

These new findings, freely available, offer immediate diagnostic applications for DDX3X-related neurodevelopmental disorders.

Sanger Institute researchers are now applying saturation gene editing at scale to many other genes relevant for neurodevelopmental disorders and cancer, collaborating globally through the Atlas of Variant Effects Alliance. The approach offers a promising avenue for unlocking hidden insights within our genetic code.

Reference
Radford E.J. *et al. Nature Communications* 2023; **14:** 7702.

6

How parental relatedness affects common, complex diseases

Wellcome Sanger Institute scientists have explored the link between consanguinity – unions between two blood-related individuals who are second cousins or closer – and the risk of developing several complex genetic diseases, including type 2 diabetes, using a new approach that removes confounding sociocultural factors.

Consanguinity is practised globally to varying degrees. Such a union between blood-related individuals increases the fraction of an individual's genome that is inherited identically from both parents, a phenomenon termed autozygosity. While it is well established that consanguinity increases the risk of rare, single-gene disorders, its impact on common diseases is not well understood.

Researchers from the Sanger Institute and their collaborators at Queen Mary University of London analysed the genomic data of diverse groups to investigate the relationship between autozygosity and the prevalence of common diseases.

Researchers analysed genomic data to describe different patterns of consanguinity in distinct populations, including 23,978 British individuals of Pakistani and Bangladeshi descent from the Genes & Health cohort and 397,184 individuals of European or South Asian descent from the UK Biobank cohort.

To study the link between autozygosity and common disease risk without the influence of external variables, such as religiosity, education or diet, researchers selected only individuals from the Genes & Health and UK Biobank cohorts whose parents were inferred to be first cousins based on genetic data. This new restriction method ensured any links observed were biological in cause.

Among the 61 complex genetic diseases they examined, the researchers identified 12 associated with increased autozygosity resulting from consanguinity, including type 2 diabetes, asthma and post-traumatic stress disorder (PTSD). They suggest consanguinity may account for approximately 10 per cent of type 2 diabetes cases among British Pakistanis and around 3 per cent among British Bangladeshis.

The researchers worked closely with the Genes & Health Community Advisory Board during the study. They highlight that any health risks of consanguinity should be balanced with the positive social benefits of the practice, as well as being considered alongside other, more substantial, modifiable risk factors, such as exercise, smoking and body mass index.

The research sheds light on the complex factors influencing health outcomes within British Pakistani and Bangladeshi communities. It also underscores the importance of understanding the genetic and environmental factors contributing to common diseases in diverse populations, ultimately guiding more targeted healthcare interventions.

Reference
Malawsky D.S. *et al. Cell* 2023; **186:** 4514-27.e14

Profile

Haerin Jang
PhD student in the Davenport research group

Haerin is studying the genetics of Systemic Lupus Erythematosus (known as SLE, or lupus) which affects 50,000 people in the UK. Her goal is to understand why there are such large differences between how different people are affected by the disease.

"In total, I am analysing data from 750,000 individual cells. I'm writing code, developing computational tools and using statistical models to understand and interpret all these data."

"From these data, we can stratify the patients into groups with different disease severity, for example, to see how their genetics plays a role. It may not be solvable – the datasets are complex – but we hope to pinpoint some of the regions in the genome that affect cell type specific gene expression and influence the course of disease."

The data and analyses from her work will be open for other researchers, through Open Targets (the Institute's pre-competitive public-private research partnership), creating a valuable resource for new advances into understanding the condition.

Read more at sangerinstitute.blog

Parasites and Microbes

We study the genomics and evolution of disease-causing organisms and the human microbiome. We build networks at scale to help monitor infectious diseases and the effects of health policies worldwide, identifying the drivers of drug, vaccine and insecticide resistance to guide health planning.

In this section

1

Genomics shows value of mass vaccination

Researchers have uncovered two new cholera substrains in a displaced refugee population in southern Bangladesh. The team confirmed a pre-emptive mass vaccination campaign of over one million refugees was successful in preventing an epidemic. They also showed that the cholera bacterium targeted in this campaign was the high-risk pandemic strain responsible for global outbreaks.

There are an estimated 1.3 to 4 million cases of cholera annually, caused by *Vibrio cholerae*. If left untreated, infection is quickly lethal; the disease is also responsible for an estimated 21,000–143,000 deaths worldwide each year. Access to safe water and sanitation is critical to prevent and control the transmission of cholera, meaning that less affluent communities are more exposed. In Bangladesh, cholera is endemic, with an estimated 66 million people at risk.

A study, led by the Wellcome Sanger Institute and the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), used whole genome sequencing to analyse and track cholera strains in the Rohingya Refugee population in Southern Bangladesh. By analysing 223 cholera samples, collected between 2017 and 2019, the researchers identified that the main strain circulating was 7PET – one that has already caused millions of deaths worldwide in a pandemic that began in the 1960s.

The work uncovered the presence of two new sublineages of 7PET, one with a global presence and another found in Asia and the Middle East. These two cholera sublineages also affect people differently, and the team suggests that variations in the bacterium's DNA could explain these differences.

The study shows that genomic surveillance of cholera during a mass vaccination campaign can help public officials make informed decisions about disease control. In this case, the genomic data gave important information about the strains in circulation, confirming that the vaccination campaign was vital and justified in targeting a more virulent form of the bacterium.

Reference

Taylor-Brown, A. *et al. Nature Communications* 2023; **14:** 3773.

2

Bacterial drug resistance tracked in Ghana

The first genomic surveillance of *Klebsiella* bacteria in Ghana has shown that multi-drug resistant pathogens are only found in hospital settings. Scientists from the Wellcome Sanger Institute, Oslo University Hospitals, the University for Development Studies, Ghana and collaborators used a One Health approach to understand the spread of antibiotic resistance. The insights could help inform control measures for *Klebsiella pneumoniae*.

Klebsiella pneumoniae is a major human pathogen with the ability to cause a wide range of infections, including pneumonia, meningitis, wound and soft tissue infections, urinary tract infections and neonatal sepsis. Antibiotic resistance in *K. pneumoniae* has been steadily increasing.

Some strains of the bacteria are able to produce extended-spectrum beta-lactamases (ESBLs), enzymes that enable them to become resistant to certain types of antibiotics, including penicillin. Other strains of *K. pneumoniae* have become resistant to carbapenems, a broad-spectrum antibiotic used for infections that are not responding to other treatments.

While there has been genomic surveillance of *Klebsiella* in European countries, there are many countries where data are still needed. The researchers sequenced the genomes of 573 *Klebsiella* samples collected from 78 clinical, environmental and animal sources in and around the city of Tamale, Ghana.

Among the *Klebsiella* isolates sequenced, researchers found that *K. pneumoniae* made up two-thirds. The team identified two strains that are resistant to carbapenems and multiple strains containing ESBL genes. This level of antibiotic resistance was similar to that seen in Italy. The antibiotic resistance was rare in the environmental samples, suggesting that the drug-resistant *K. pneumoniae* strains are less successful outside of hospitals and do not outcompete other, less dangerous forms of the bacteria.

The researchers suggest that the clinical use of antibiotics drove the increase in antibiotic resistance. This insight could help inform public health measures focused on reducing the spread of heavily antibiotic-resistant bacteria in hospitals and highlights the importance of genomic surveillance.

Reference
Calland J.K. *et al. The Lancet Microbe* 2023; **4:** e943-52.

The genomes of 573 *Klebsiella* samples from 78 clinical, environmental and animal sources in and around Tamale, Ghana were sequenced

3

Networks of syphilis spread and antibiotic resistance found

In a pioneering approach to sexually transmitted infection (STI) surveillance, Sanger Institute scientists applied genomics to standard epidemiological data to gain insights into the spread of syphilis among populations in England.

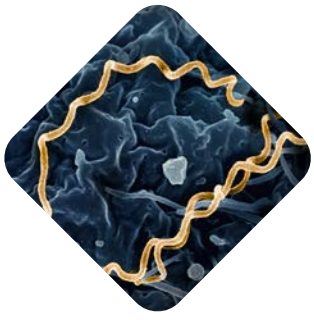
The rates of syphilis, a sexually transmitted infection caused by the bacterium *Treponema pallidum* have been dramatically increasing in England over the past 10 years. To understand transmission dynamics of the bacterium, researchers at the Sanger Institute and the UK Health Security Agency sequenced 237 whole genome samples of *T. pallidum*.

The researchers linked national patient demographic, geospatial and behavioural metadata to whole *T. pallidum* genome sequences previously generated from patient samples collected from across England between 2012 and 2018. They performed detailed phylogenomic analyses, comparing the bacterial genomes from different individuals.

The team identified single letter changes in DNA sequences to distinguish one sublineage of *T. pallidum* from another – a level of detail never seen before. They built intricate networks of syphilis spread, delineated by geography, sexual preference and other factors. These inferences were verified by epidemiological data. Their analysis revealed two dominant *T. pallidum* sublineages in England, present in different populations. This information could allow public health authorities to map distinct sexual networks during ongoing outbreaks, and inform control strategies.

The study's analysis of *T. pallidum* genetic diversity also uncovered significant drug resistance, particularly to macrolides, a common class of antibiotics used to treat STIs. This evidence can guide public health policies in safeguarding antibiotics.

This research highlights the potential for genomics in enhancing understanding and informing strategies for STI surveillance, prevention and treatment. It also underscores the importance of collaborative efforts between researchers and public health agencies.



Reference
Beale M.A. *et al. The Lancet Microbe* 2023; **4:** e770-80.

Profile

Dr Oumie Kuyateh is a Sanger Excellence Fellow and a Junior Research Fellow at the University of Cambridge.

Growing up in The Gambia, Oumie moved to the UK to study biochemistry at University College London before achieving her Masters and PhD at the University of Edinburgh.

Oumie's research is focused on understanding the microbiomes of African children – both in health and sickness – and how factors such as pollution, antibiotics and vaccines, affect them. The hope is that this will inform the design of diagnostic tests for respiratory diseases such as pneumonia.

"Sanger is a great place to be, and I think one of the best things about it is the focus on collaboration."

Read more at sangerinstitute.blog

4

Life cycle of human blood fluke mapped

***Schistosoma mansoni* is an important but neglected flatworm that affects millions of people worldwide. It has a complex life cycle, moving between environments and hosts. A team at the Sanger Institute undertook the first comprehensive sex- and stage-specific analysis of the parasite's gene activity, from eggs to sexually naïve worms.**

Schistosomiasis is a neglected tropical disease that causes abdominal pain and diarrhoea, hypertension in abdominal blood vessels, liver granulomas and fibrosis. In 2021, over 75 million people in 51 countries received preventative treatment.

The parasite has a complex life cycle, with sexual reproduction in a mammalian host, periods living in freshwater and asexual replication in snails.

The Sanger Institute team previously generated a high-quality reference genome for *S. mansoni* to underpin further genetic and evolutionary study. The complete *S. mansoni* life cycle was maintained at the Sanger Institute, and the team previously characterised active genes at various life-cycle stages. They also used single-cell RNA sequencing to provide intricate details of body plans and cellular functions of the blood fluke.

However, significant variation exists across published datasets, including the use of different parasite strains, pooling of male and female parasites and differing sequencing technologies. This makes it challenging to understand gene activity changes across the life cycle or to compare data sets.

To better understand the parasite, the team used RNA sequencing to study the active genes at eight different life stages. The stages span eggs, free-living water-borne forms and from both hosts. They generated deep RNA-sequencing data for 75 data sets, using a single approach to increase the accuracy comparisons.

The data are freely accessible via a user-friendly tool to explore gene activity. It will support future studies into parasite development and aid drug-target discovery.

Reference
Buddenborg S.K. *et al. Scientific Data* 2023; **10:** 775.

Tool to visualise and explore gene expression
https://lifecycle.schisto.xyz/

75+ million
people in 51 countries received preventative treatment in 2021

5

E. coli's protective coat offers treatment target

A multi-centre team led by the Sanger Institute has mapped the evolutionary timeline and population distribution of *Escherichia coli's* protective outer capsule, which is responsible for the bacterium's virulence. Targeting this layer could help treat infections.

The recent rise of hyper-virulent and multi-drug resistant *E. coli* strains means there is an urgent need to develop effective strategies to prevent and treat infections. The team focused on a subset of *E. coli* with a K1 capsule that cause severe diseases such as bloodstream or kidney infections and meningitis in newborns.

The researchers mapped the evolution, prevalence and distribution of *E. coli* with K1. Using high-resolution population genomics, whole genome sequencing and computational tools, they analysed 5,065 clinical samples from different countries and time periods, including the pre-antibiotic.

Their study found that 25 per cent of all current *E. coli* bacteria responsible for blood infections contain the genetic information needed to develop the K1 capsule. They also found that the K1 capsule evolved approximately 500 years earlier than previously thought, highlighting the importance of the capsule for the bacterium's survival. Understanding the evolutionary history and current presence of K1 in *E. coli* populations will allow researchers to better monitor and predict the emergence of virulent clones.

The team also used enzymes from bacteriophages – viruses that 'infect and kill' bacteria – to remove *E. coli's* capsule. This resulted in the bacteria becoming vulnerable to the human immune system in a laboratory system. Their work suggests that targeting the capsule could be a way to treat *E. coli* infection without antibiotics.

> These studies will enable us to reconstruct the evolutionary history of successful bacterial lineages and pinpoint changes in their genetic make-up that can lead to their ability to spread and cause disease. Such knowledge is ultimately providing the basis for designing future interventions and therapies.
>
> **Professor Jukka Corander**
> Wellcome Sanger Institute and the University of Oslo

Reference
Arredondo-Alonso S. *et al. Nature Communications* 2023; **14:** 3294.

6

Genes powering Yemen cholera outbreak found

Scientists have identified the source of antibiotic resistance in bacteria driving the ongoing Yemen cholera epidemic.

The Yemen cholera outbreak, the largest in modern history, has affected over 2.5 million people and resulted in at least 4,000 deaths since 2016. Initially, macrolide antibiotics were widely used to treat the disease, but by 2018 these frontline treatments were starting to fail.

To uncover the reasons behind the growing drug resistance, researchers from the Sanger Institute, University of Toronto, Institut Pasteur, Sana'a University and their collaborators analysed the DNA from 260 samples of *Vibrio cholerae* — the bacterium responsible for the disease. The samples were from the Yemen epidemic, collected between 2016 and 2019.

The team found that a type of *V. cholerae* containing multidrug-resistant genetic elements took over as the main pathogen during the Yemen outbreak. They identified the presence of a new plasmid — a small, circular DNA molecule — in all epidemic *V. cholerae* samples tested since November 2018. This plasmid included genes encoding resistance to multiple antibiotics, including macrolides.

The multidrug-resistant plasmid was also found in local, endemic *V. cholerae* strains unrelated to the outbreak, suggesting that these different strains were able to exchange plasmids with antibiotic-resistance capabilities. It is likely that extensive antibiotic use drove this process.

The new plasmid's stability within *V. cholerae* is concerning in the context of future cholera outbreaks.

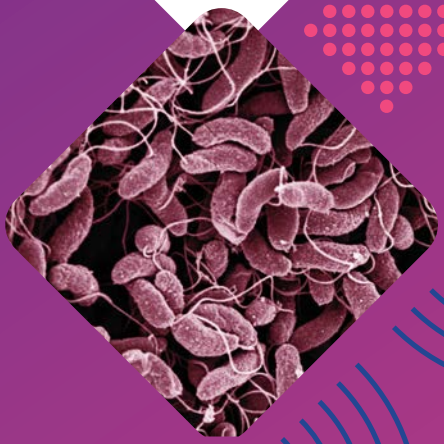
The study underscores the adaptability of the cholera pathogen and the need for further research into bacterial genome evolution and emergence of multidrug-resistant strains.

The collaborative consortium of authors organised an online symposium in October 2022 to showcase the various expertise of each research group and to foster collective thinking to take this initiative forward.

Recordings of these talks can be accessed online at: **https://sites.google.com/view/yemencholeragenome/home**

Reference
Lassalle F. *et al. Nature Microbiology* 2023; **8:** 1787-98.

2.5+ million
people have been affected by the Yemen cholera outbreak



4,000+
deaths have occurred since 2016

Profile

Florent Lassalle
Principal Bioinformatician and leader of the Parasites and Microbes Informatics team at the Wellcome Sanger Institute

Florent's team of software developers and bioinformaticians support the pipeline development needed to deliver the high-throughput analysis of genomes required by the Sanger Institute's parasites and microbes researchers.

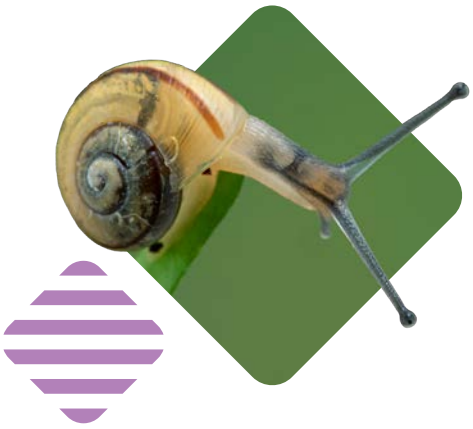
One aspect of this work is using pangeome approaches to tackle the scale of the datasets being generated. Florent leads the development of several pipelines that allow the Institute to scale up its delivery of core research services such as genome annotation, read mapping and metagenome analysis.

His team also explore bacterial genome diversity and adaptation through phylogenetic methods to model the processes of gene flow within microbial populations and reconstruct the history of diversification of their pangenomes.

Tree of Life

We are building the library of life. We produce high-quality reference genomes to explore the evolution, function and interactions of life on Earth. We seek to aid conservation and biodiversity work and provide the underpinnings of a new way of doing biology.

In this section



1

How snails earn their stripes

Sanger Institute scientists, alongside collaborators from the University of Nottingham and the Netherlands, have created a detailed genetic map of the Grove snail. They uncovered core genetic mechanisms underlying the snails' diverse shell patterns. The findings enhance our understanding of snail biology and set the stage for wider insights into colour variation within species across the animal kingdom, which is often linked to key evolutionary adaptations.

Animal species often show significant colour diversity, yet the study of snails, with their wide array of shell colours and patterns, has remained relatively overlooked. This research focuses on the Grove snail, *Cepaea nemoralis*, which lives in diverse habitats across Europe and North America. Contrary to the traditional categorisation of their shells as yellow, pink or brown, recent research has shown they exhibit colours on a continuum, and the banding patterns range from none to five.

Prior research identified nine genetic locations that influence these shell features, forming a distinctive 'supergene'. In this genetic arrangement, linked genes are inherited together, reducing the chances of recombination and likely preserving evolutionary traits.

However, the repetitive and complex nature of snail DNA has made it difficult to put together its full genome sequence. Building on recent advances in biotechnology, for rapid and cost-effective genotyping, Sanger Institute scientists have created the first high-density linkage map for the Grove snail. This map pinpoints two genes that control shell colour and banding and highlights the chromosomal region containing the supergene.

These advancements in genome mapping provide more than just insights into colour variation in the Grove snail. They can be widely applied to research on population genetics and evolutionary biology, potentially highlighting links between colour variation and environmental factors, such as climate change. As genetic mapping and genome assembly technologies evolve for diverse and complex species, the scientific understanding of the genetics underlying the repeated evolution of colour variation, inheritance and the emergence of supergenes will deepen.

Reference
Johansen M. *et al. Heredity* 2023; **131:** 327-37.

2

Uncovering the secrets of life in extreme cold

An international team of scientists have sequenced the genomes of 24 Antarctic fish species for the first time. These fish, called Notothenioids, provide a useful model for studying the genomics of extremophiles: creatures that thrive in extreme conditions. The genomes will offer new insights into how the fish thrive in the sub-zero temperatures of the Southern Ocean.

The study focused on a subgroup called 'icefish', which are the only vertebrates that lack red blood cells and therefore have no oxygen-carrying haemoglobin proteins.

Scientists previously struggled to sequence icefish due to their long and repetitive DNA.

However, this team leveraged the Sanger Institute's advanced long-read sequencing expertise to explore icefish evolution.

The team found that cold-resistant Notothenioids diverged from other species just 10.7 million years ago, with a rapid increase in species five million years ago. There were significant genomic changes that facilitated the icefish's survival in extreme cold, including a doubled genome size from an increase in transposons – elements that can copy themselves and may introduce new functions.

Icefish developed antifreeze glycoproteins that prevent ice crystal formation in their blood. Further, their distinctive oxygen transport system resulted in thinner, faster-flowing blood, albeit with a reduced oxygen-carrying capacity compared to other fish. To compensate, icefish evolved innovative and evolutionarily costly physiological adaptations, such as slow metabolisms and large hearts.

This pioneering research enhances our understanding of biodiversity and extremophile biology and highlights the Institute's expertise in using advanced genomic technologies to study complex species. Since icefish are vulnerable to climate change, these findings underscore the importance of the need to study the impact of temperature changes on diverse ecosystems.

Reference
Bista I. *et al. Nature Communications* 2023; **14:** 3412.

3

New 'Tree of Sex' database for evolutionary reproduction

Wellcome Sanger Institute scientists have launched a revamped genomics database, the Tree of Sex, which provides new opportunities for understanding life's diverse reproductive strategies, and how these can evolve over time. This public repository builds on the Sanger Institute's capacity to leverage large-scale genomic technologies.

An earlier version of the Tree of Sex launched 13 years ago, which captured the reproductive strategies of over 22,000 eukaryotic species. Since then, Sanger researchers have produced hundreds of high-quality reference genomes, which provide a wealth of new and relevant information for the study of reproductive evolution.

The reproductive systems and sex determination methods for eukaryotic species are vast and varied – even for closely related species. These range from hermaphroditic organisms such as snails, which have both male and female reproductive organs to those that undergo sex changes; for instance, young male clownfish can transform into females once they reach a large size. In contrast, the Amazon Molly Fish has evolved away from male reproductive organs and reproduces asexually.

The advantage of this database is its capacity to collate information on the diverse reproductive traits of various species within the context of phylogenetics – the evolutionary relationships between organisms. This allows scientists to correlate changes in mating systems over time with key genomic features, such as chromosome number and ploidy level (number of chromosome sets in a cell), as well as the life history of a species.

By integrating the existing Tree of Sex database with new genomic data, the researchers are developing a curated repository to explore questions about reproductive evolution, including studies into genetic diversity, extinction risk and species creation.

The team has designed this centralised repository to be sustainable, scalable and universally accessible. Sanger's unique position provides stable funding and cutting-edge genomics infrastructure, ensuring the database's continued maintenance, updates and adaptation to future advances in the field of evolutionary reproduction.

Tree of Sex project website
https://treeofsex.sanger.ac.uk

Profile

Andrew Varley
Enabling Platforms Team Leader, Tree of Life programme

Andrew leads a team of software developers, who develop and maintain bespoke programmes that Tree of Life researchers use to process and track incoming samples and ensure that the resulting genomic data are assigned correctly.

The team has written more than 250,000 lines of code so far, over a wide range of technologies and frameworks, deployed on dozens of servers, and continuously process data and metadata.

Find out more at sangerinstitute.blog

The people who will be out there collecting these organisms … will be preserving them in such a way that we can generate a baseline understanding of their genomes. And forever more there will be a record of what that organism's DNA looked like.

Dr Mara Lawniczak
Group Leader at the
Wellcome Sanger Institute

4

Discovering new ocean species to preserve life

The Sanger Institute has joined the Ocean Census, the largest global collaboration of scientists discovering the secrets of ocean life and protecting biodiversity. The programme aims to identify at least 100,000 new marine species in its first decade, with Sanger researchers playing a pivotal role in analysing their genomes.

As the largest ecosystem on Earth, the ocean sustains all life by generating oxygen, absorbing carbon dioxide and providing a food source for billions. It also enables researchers to drive advances in medical research and understand the origins of life. Despite its significance, the ocean's biodiversity, which is concentrated in hotspots, remains largely undiscovered and vulnerable.

Scientists estimate approximately 2 million marine species exist, yet only 10 per cent of these have been identified. This undiscovered life may hold essential information for preserving the future of life on Earth.

For the past two centuries, traditional taxonomy has only identified around 2,000 new species annually. However, faced with impending climate and biodiversity crises, researchers need a faster approach. The Ocean Census is achieving this by applying technological advancements in areas such as digital imaging, sequencing and machine learning.

The scientists will use divers, submarines and deep-sea robots to explore biodiversity hotspots across the ocean depths. Ocean Census's biodiversity centres across the globe, including the Sanger Institute, will conduct high-resolution imaging and DNA sequencing. This information will be shared in a public repository to help inform conservation efforts.

Understanding ocean biodiversity is fundamental to addressing the climate emergency, sustainable food production and new medical and biotechnology discoveries. The Ocean Census warns that fast action is needed within the next decade to change the climate trajectory for future millennia. Ocean life must be understood and protected to secure the future of life on Earth.

www.oceancensus.org

5

Genomics helps endangered species

Wellcome Sanger Institute scientists have produced a reference genome of the critically endangered British pine hoverfly. Conservationists are using the genome sequence in a coordinated programme to save this important pollinator from extinction in the UK. The reference genome, which is publicly available through the Ensembl browser, also lays the foundation for researchers to apply genomics to conserve species that help sustain life on Earth.

The pine hoverfly (*Blera fallax*) is the most endangered insect in Britain, found only at the edge of the Scottish Cairngorms forest. After several failed attempts at reintroduction, conservationists set up a staggered release and controlled breeding programme in 2018, scaling up to introducing 6,000 larvae annually. Now almost a decade later, the pine hoverfly and its larvae, the so-called rat-tailed maggots, are thriving.

Although the number of pine hoverflies in Scotland is now flourishing, the genetic health of the population is uncertain due to the high risk of inbreeding. For the breeding programme to succeed, researchers must use detailed genomic studies to monitor the genetic diversity of the population at each generation. Addressing this, Sanger Institute scientists have produced the first high-quality reference genome for the species, which enables conservationists to select the healthiest mating pairs.

The research team will also compare the population's genetic diversity to that of larger, healthier populations, such as those in Scandinavia. This will help guide the genetic diversity of the British population, and if the diversity needs to increase, it might be possible to relocate flies from Scandinavia. The reference genome can also be used to help determine if Scandinavian flies would thrive in British forests.

This cutting-edge research showcases how genomics can be applied to conservation projects. The reference genomes produced at the Sanger Institute will play a vital role in global initiatives that aim to ensure the genetic health and viability of endangered populations and safeguard biodiversity.

Reference
Taylor H.R. *et al. Wellcome Open Research* 2023; **8:** 89.

6

1,000 reference genomes generated

Wellcome Sanger Institute scientists have generated 1,000 high-quality reference genomes for diverse eukaryotic species (those with a cell nucleus and membrane), across the tree of life. The data are part of global, collaborative projects which together aim to sequence the DNA of all species on Earth and hold promise for advances in conservation, drug discovery and evolutionary studies.

Teams at the Sanger Institute have combined advances in sequencing technologies with their laboratory, bioinformatics and genome assembly expertise. They have built unique systems to process and analyse DNA and accurately determine an organism's genome sequence for the first time. They have now produced over 1,000 genomes, which represents nearly half of all high-quality reference genomes produced to date worldwide. Their work is part of collaborations that involve naturalists, taxonomists, laboratory specialists and computer scientists at organisations across the globe.

The teams published these genomes openly in the European Nucleotide Archive (ENA) for use by researchers worldwide. In line with their goal of capturing biodiversity, the team sequenced genomes that span 141 different taxonomic orders across more than 20 phyla.

Of the 1,000 genomes, 390 are Lepidoptera species (moths and butterflies), enabling in-depth analyses of the genomes of these closely related and biologically important species.

The 1,000th genome, European Mistletoe, was the largest ever assembled and which posed technical challenges with its 94 billion DNA base pairs (30 times the size of the human genome) and its large chromosomes.

Other notable genomes included the common toad, which needed extensive manual error correction owing to its complexity. The Chalkhill blue butterfly had the most chromosomes (90), while silverweed had the highest ploidy count at four copies of its genome per cell.

The team continues to sequence thousands more genomes and leads on the Darwin Tree of Life Project, which aims to sequence 70,000 species in Britain and Ireland. All the sequences generated contribute to the global Earth BioGenome Project, whose mission is to produce reference genomes for all 1.2 million eukaryotic species on Earth. Their work will underpin biological studies for the future and advance the new scientific field of biodiversity genomics, which will lead to improvements across diverse areas including biomedicines, climate adaptation and evolutionary biology.

Profile

Dr Cibele Sotero-Caio
Genomic Data Curator in the Tree of Life Informatics Infrastructure team

Cibele looks after the Sanger Institute's Genomes on a Tree (or GoaT) database, which will hold data for the hundreds of thousand of species that are having their genomes sequenced for the first time.

Free and open, GoaT tracks projects around the world, pulling in information about which eukaryotic species are being sequenced, and what their genomes look like. It's used for both planning and research and is an increasingly rich source of knowledge about the world's genomes.

"I'm drawn to biology and try to define patterns – but observing and understanding exceptions is one of the most cool things about being a biologist. Because of my work here, I've seen so many exceptions to things that I thought were true. And that is actually what I enjoy the most; daily data analysis is never boring."

Find out more at sangerinstitute.blog

Our approach

We empower bold thinking to take scientific risks that benefit research worldwide

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 34

Impact

We strive to carry out the most insightful research and help scientists and doctors to use our discoveries to develop new techniques, tests and therapies.

Page 36

Collaboration

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

Page 38

Innovation

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

Page 40

Culture

The diversity in skills and knowledge that we all bring combine to make the Institute the thriving ideas factory that it is. We support our colleagues to reach their full potential and to help each other thrive in their work. We encourage everyone to benefit from the wide range of creativity and expertise at the Institute by valuing each other's differences in thought, background and perspective.

Page 42

Influencing Policy

We advise all levels of government both in the UK and across the world on the role, impact and importance of genomic and life science research.

Page 44

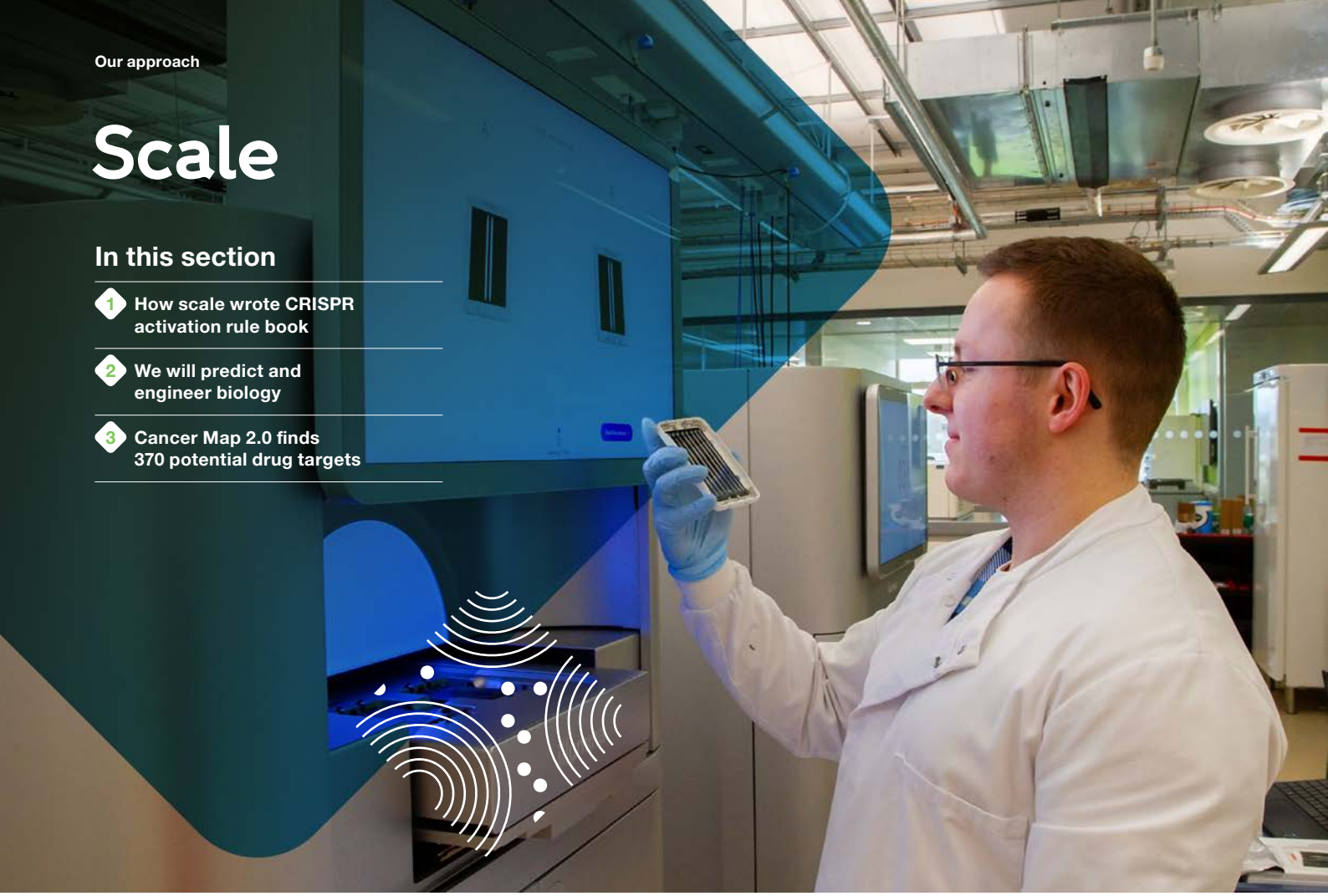
30 years of Impact

For 30 years, we have pioneered using cutting-edge technologies at scale to open new fields of discovery in science and health

Page 46

Scale

In this section



1

How scale wrote CRISPR activation rule book

By inserting a marker gene at thousands of sites across the genome of a human stem cell line and observing the effects of CRISPR activation (CRISPRa) in a range of cell states, Sanger scientists and collaborators have written the rules of how to use this technology most effectively.

Researchers in Sanger's cellular genetics teams applied massively parallel, high-throughput integration of a minimal, barcoded reporter gene at thousands of points in the genome of a human stem cell line. By activating the cells with CRISPRa, observing the results and determining gene position with DNA sequencing, the researchers could identify locations where activation was successful. The stem cell line used differentiates into neurons, allowing the team to also gather information on CRISPRa efficiency in different cell types.

The team uncovered multiple features that impact CRISPRa efficiency, including expression level, chromatin status, cell state and gene location. They found that bivalent genes – key developmental regulator genes that have both repressing and activating marks in the same region – can be robustly activated by CRISPRa. They also discovered that CRISPRa could achieve the same overexpression levels that are necessary to drive significant changes in cell state and cause them to differentiate.

While most genes can be activated by CRISPRa, the researchers found that not every cell responded to the same extent. For example, genes that contain H3K9me3-repressing regulators showed greater variation in response.

These data demonstrate for the first time that CRISPRa is generally applicable across chromatin states and cell types and highlights the factors that impact the degree of gene activation. Understanding these factors will aid the design of future CRISPRa screens, which are used to look for genes involved in genetic diseases.

Reference

Wu Q. *et al. Molecular Cell* 2023; **83**: P1125-39.

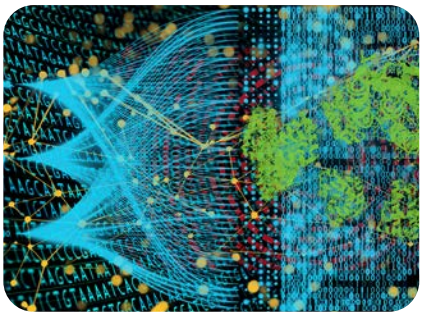
Profile

Qianxin Wu, the first author of this study, works jointly between the Cellular Genetics research programme and the Cellular and Gene Editing R&D team. A highly entrepreneurial scientist, with a passion for problem solving, Qianxin is at the forefront of innovation, working on CRISPR, genome editing and single-cell technologies.

Find out more at sangerinstitute.blog

We believe that the transformation of biology into a programmable engineering science will be the most important technological revolution of this century, and that Generative and Synthetic Genomics will open up unprecedented possibilities for industry, agriculture, the environment and medicine.

Professor Ben Lehner
Head of the Generative and Synthetic Genomics programme, Wellcome Sanger Institute



2

We will predict and engineer biology

The Sanger Institute has launched a new research programme that combines large-scale genomic data generation with machine learning to predict the impacts of mutations and engineer new biological systems. It is underpinned by the Institute's world-leading capabilities in large-scale genomic data generation and analysis.

Genomics and molecular biology have enabled researchers to extensively describe and understand biological systems in health and disease. But scientists struggle to predict how biological systems respond to variations in DNA sequence which means engineering genomes to produce much-needed materials and therapies remains difficult.

The problem lies in understanding how DNA sequences determine the properties and regulation of proteins. To deliver this knowledge, the Generative and Synthetic Genomics programme combines Sanger's world-leading capabilities in large-scale genomic data generation and analysis with machine learning. The results will transform genomics, allowing researchers to make accurate predictions about the effects of individual, and combinations of, DNA changes on functional biology. These predictions will then be used to design the properties, activities, regulation and expression of proteins from scratch.

The programme is initially focusing on the individual proteins that build our bodies, and how they are controlled. To achieve this, the teams aim to understand, predict and engineer the effects of editing every nucleotide – the building blocks of DNA – in a genome.

The findings will make it much easier to engineer proteins as therapeutics and for clean biotechnology. The work will also lay the foundations for generating models for engineering gene pathways and entire cells and tissues for medical and biotechnological applications.

Finally, the programme seeks to develop technologies to write and edit genomes at scale and speed. This will provide understanding of how genomes work and allow scientists to engineer, in a considered and responsible manner, the genomes of simpler organisms such as yeast and bacteria to make useful products in non-polluting ways.

3

Cancer Map 2.0 finds 370 potential drug targets

Using machine learning to analyse genetic and genomic data from 930 cancer cell lines, researchers from the Sanger Institute, Open Targets and collaborators sifted through thousands of potential anti-cancer targets to identify 370 high-priority candidates.

Currently most types of cancer treatment target both healthy and cancerous cells, causing harsh side-effects, and some cancer types have no effective therapies at all. To avoid unwanted collateral damage and effectively treat resistant tumours, precision medicines based on the exact genetic mutations driving the specific cancer are needed.

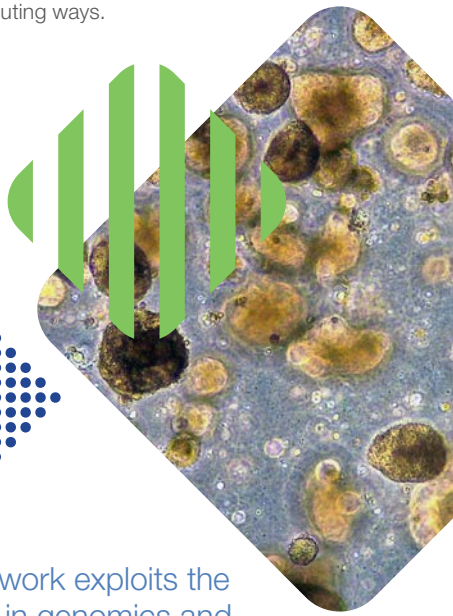
To identify potential targets for precision drug treatment, the Sanger Institute, GSK, EMBL-EBI, Open Targets and collaborators created the Cancer Dependency Map by studying more than 300 cancer cell models covering 30 cancer types. Using CRISPR gene editing to systematically disrupt the functions of nearly 20,000 genes, the partnership identified 600 potential drug targets. The results, published in 2019, helped to guide and accelerate the development of targeted cancer treatments.

In early 2024, the partnership released the results of the second generation of the Map, which used machine learning to interrogate a greatly increased data set of 930 cancer cell lines. The new Map not only identified 370 priority drug targets across 27 cancer types but also linked these targets to clinical markers to identify those patients who would benefit the most.

In addition, the collaboration explored how the dependency-marker pairs it had found fitted into known networks of molecular interactions within cells to provide clues as to how cell biology is disrupted by cancer, and which targets might yield the most effective therapies. This knowledge provides enhanced possibilities for accelerating drug development, driving the development of personalised, precision medicine.

Reference
Pacini C. *et al. Cancer Cell* 2024. DOI: **10.1016/j.ccell.2023.12.016**

These data can be accessed at: **https://depmap.sanger.ac.uk**



[Our] work exploits the latest in genomics and computational biology to understand how we can best target cancer cells. This will help drug developers focus on the highest value targets to bring new medicines to patients more quickly.

Dr Mathew Garnett
Group Leader in Cancer, Ageing and Somatic Mutation, Wellcome Sanger Institute

Impact

In this section

1

Genome ‘web’ captures human diversity

Twenty years after the first human reference genome was published, the Sanger Institute supported the delivery of the first human pangenome that more accurately captures global genomic diversity, enabling greater understanding of differences in disease severity and treatment response.

The first reference human genome was a pivotal moment in global science. It powered a new wave of discovery in health and disease and established the principles of open-access science. However, this reference genome has important limitations.

A key problem with the original reference genome is that it was created from just 20 individuals from one location in North America (and nearly 70 per cent came from just one person). Another issue is that the reference genome is haploid, detailing the DNA code of just one chromosome from a cell’s diploid pair.

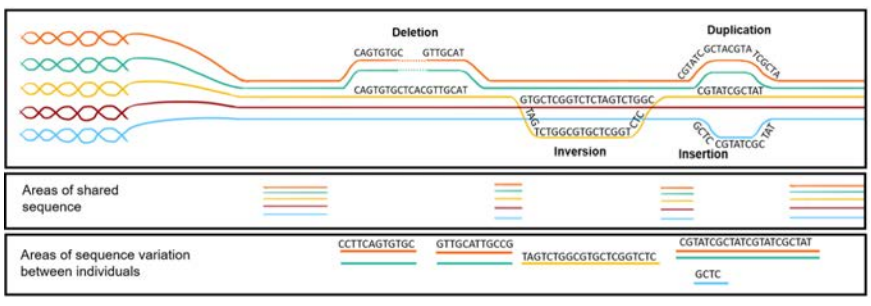
To create a more representative reference genome, the Human Pangenome Reference Consortium read 47 full diploid genomes from diverse individuals. To do this, the Consortium needed a high-throughput, scalable system to generate the diploid genome sequences with exceptional accuracy and coverage – and minimal manual intervention. The Sanger Institute helped to identify the best combination of genome sequencing and assembly approaches by coordinating and assessing an international ‘bake off’ of techniques.

The resulting, mostly automated, system produced diploid reference genomes that were more than 99 per cent accurate and 99 per cent complete. The resulting pangenome is a web of DNA sequences with different paths through it. Some regions of the genome have many routes – where sequence diverges between people. Others have one path – where sequence is conserved.

The sequences also capture the full range of genetic diversity: from single base DNA changes, through copy number variations to structural rearrangements. These differences have important effects on gene function and can now be accurately studied.

References

Liao W-W. *et al. Nature* 2023; **617:** 312-24.
and
Jarvis E.D. *et al. Nature* 2022; **611:** 519-31.



2

Sparking a CHAIN reaction in child health

Connections forged between African researchers and Sanger Institute scientists are building genomic capacity in childhood microbiome research in low- and middle-income countries.

To date, the majority of research on the human microbiome has focused on high-income countries, particularly Europe and North America. To address this power imbalance, Sanger researchers and partners in Kenya, Pakistan and the US held a retreat at the Wellcome Genome Campus.

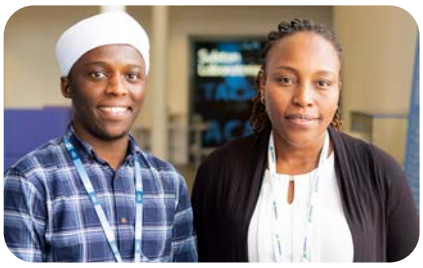
Supported by Wellcome Connecting Science and the Gates Foundation, the event brought together more than 40 leaders in gut microbiome research, funding and policy from more than 20 countries. They discussed how to accelerate their studies to improve the health of children in Asian, Latin American and African countries.

The retreat built on relationships forged between the Sanger Institute's Lawley laboratory and the Childhood Acute Illness and Nutrition Network (CHAIN), based at KEMRI Wellcome Trust in Nairobi. One of the organisers, Caroline Tigoli, met Dr Trevor Lawley through her work at CHAIN, and he is now co-supervising her DPhil scholarship, and another attendee, a joint Sanger Institute and University of Cambridge PhD student Bonface Gichuki, also connected with Dr Lawley through CHAIN.

The event enabled early career scientists from countries where microbiome research is underrepresented to network with established researchers from around the world. As well as sharing knowledge, the attendees agreed that mentoring and career development, establishing international networks and collaboration were key to improving research capacity.

As a result, a group has been formed to develop an institutional model for microbiome innovation in global child health and to share best practices. As part of this work a dedicated slack channel has been set up to enable ongoing open dialogue between all interested parties.

Reference
Browne H. *et al. Nature* 2024; **625:** 237-40.



Bonface Gichuki and Caroline Tigoli, who are studying in the Lawley laboratory and who contributed to the retreat

> One of the challenges for researchers in Africa is access, whether this is to collaborators, expertise, funding or even the supplies required to conduct sequencing projects. Collaborations like CHAIN are a really important catalyst for genomic research in Africa.
>
> **Bonface Gichuki**
> PhD student, Wellcome Sanger Institute

3

A cool way to reduce carbon

An initiative to improve access to vital biological samples and reduce sample duplication is speeding delivery of the Sanger Institute's science and reducing its carbon footprint.

The Sanger Institute is home to 28 million samples across 300 libraries, stored in 60 freezers. Some samples date back to the inception of the Institute, 30 years ago. Within this frozen archive are key biological samples and important resources for future experiments. Locating the relevant resource can be challenging, and the freezers' energy needs are significant.

To speed access to vital samples and significantly reduce the Institute's carbon footprint, the Sanger's Legacy Assessment Project is analysing and rationalising the amount of samples the Institute retains. The initiative is a collaboration between teams in Informatics and Digital Solutions, Legal and Governance, Scientific Programmes and Scientific Operations. The work, which is led by the Institute's BioResource team, has revealed that the Institute's long-term needs can be accommodated by approximately 10 freezers.

Work is ongoing to dispose of the 50 freezers' worth of duplicate or no-longer needed samples safely. In total, 14 of the current fleet of freezers are sufficiently energy-efficient and will be kept, and the rest will be removed. This will result in a saving of approximately 225,615kWh of electricity per year.

225,615kWh
of electricity per year will be saved

50
freezers' worth of samples will be removed

Collaboration

In this section

1

Building networks and capacity around the world

To drive the next wave of discovery and innovation, the Sanger Institute's Information and Digital Solutions Team have created the Biodata Developers' (BioDev) Network. Its aim is to create skilled communities of knowledge and support centred around cutting-edge computational approaches through mentoring, training and best practices.

BioDev Network

In 2023, the BioDev Network was established by the Sanger Institute to foster new technology capabilities across organisations, countries and continents. It empowers bioinformaticians, data engineers, software developers and IT specialists to create and share ground-breaking solutions in:

- Machine learning and AI to analyse vast volumes of genomic, clinical and visual data.
- Bioinformatics workflow solutions to deliver genomic data generation and analysis at scale.
- Federated access to ensure that genomic and clinical data are appropriately and safely accessible for the global research community.

To ensure that the benefits of genomic discovery and application are shared equally, the BioDev Network focuses on:

- Global communities.
- Fellowships, Mentorships, Training and Networks.

The Network's open, globally accessible communities provide training at all levels and interests through workshops, seminars, hackathons and forums. These events are free to everyone to provide opportunities for continuous learning and knowledge sharing. The communities also organise symposia for professionals, experts and enthusiasts to explore future trends, innovations and challenges.

The BioDev Network Future Innovators Programme – launched in 2024 – seeks to build capacity worldwide and ensure that knowledge is shared without barriers. The eight-month course combines virtual learning with on-site experience to provide mentorship, skill development and networking opportunities for participants from under-represented groups in the global south.

Participants receive:

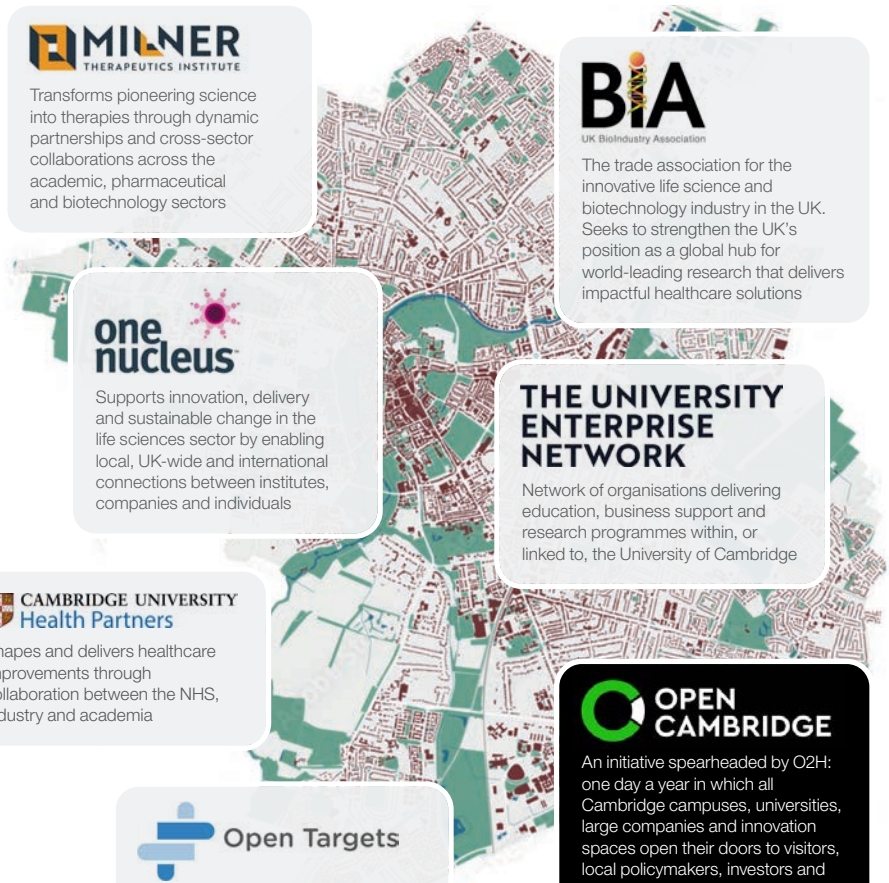
- Hands-on project experience within an informatics team.
- Virtual roundtable discussions and expert sessions covering time management, agile development and presentation skills.
- On-campus experience including networking opportunities with senior leaders and a chance to present at a genomics conference.

2

Cambridge connections

Innovation ecosystem at Cambridge – adding value to the local ecosystem with our unique genomics and biodata focus.

The Sanger Institute plays a unique role in Cambridge's innovation culture by supporting and strengthening the following collaborations to deliver clinical and societal benefit.



MILNER THERAPEUTICS INSTITUTE

Transforms pioneering science into therapies through dynamic partnerships and cross-sector collaborations across the academic, pharmaceutical and biotechnology sectors

BIA UK BioIndustry Association

The trade association for the innovative life science and biotechnology industry in the UK. Seeks to strengthen the UK's position as a global hub for world-leading research that delivers impactful healthcare solutions

one nucleus

Supports innovation, delivery and sustainable change in the life sciences sector by enabling local, UK-wide and international connections between institutes, companies and individuals

THE UNIVERSITY ENTERPRISE NETWORK

Network of organisations delivering education, business support and research programmes within, or linked to, the University of Cambridge

CAMBRIDGE&

Seeks to attract the highest-quality organisations and individuals in the knowledge-intensive industries from around the world to the Cambridge research ecosystem

CAMBRIDGE UNIVERSITY Health Partners

Shapes and delivers healthcare improvements through collaboration between the NHS, industry and academia

Innovate Cambridge

Boosts innovation in health and social care, agriculture and energy across the Greater Cambridge area by bringing together leaders from industry, investment, research institutes, local government and the University of Cambridge

OPEN CAMBRIDGE

An initiative spearheaded by O2H: one day a year in which all Cambridge campuses, universities, large companies and innovation spaces open their doors to visitors, local policymakers, investors and the local community to showcase their work and technology

Open Targets

Precompetitive public-private partnership that uses human genetics and genomics to identify and prioritise targets for drug development

3

UK-South Africa collaboration will power infectious disease surveillance

The Wellcome Sanger Institute has partnered with Stellenbosch University to coordinate their global genomic surveillance work. The partnership will identify emerging threats from respiratory viruses, mosquito- and water-borne diseases and others with pandemic potential.

The Sanger Institute and Stellenbosch University were key contributors to the global genomic surveillance of COVID-19 during the pandemic, each identifying important variants that altered public health policy in near real-time. As a result of their hard-won expertise in sequencing and analysing the changes taking place in the SARS-CoV-2 virus' genome at the national population level, the Sanger Institute founded the Genomic Surveillance Unit (GSU), and Stellenbosch University created the Centre for Epidemic Response and Innovation (CERI).

In 2024, the Genomic Surveillance Unit and the Centre for Epidemic Response and Innovation agreed to coordinate their efforts to deepen their collective knowledge of how infectious diseases evolve and spread and close the information gaps that threaten public health. The partnership will build upon both organisations' decades-long experience of malaria, HIV and other diseases, as well as COVID-19.

The close relationship will allow teams in the UK and South Africa to share resources, coordinate strategies and powerfully support partners in disease surveillance around the world. In particular, these activities align with the mission of the World Health Organization's International Pathogen Surveillance Network.

As part of this strengthened relationship, Professor Tulio de Oliveria, who founded and leads the Centre for Epidemic Response and Innovation, has joined the Genomic Surveillance Unit as Deputy Director.

"We are in a good position now to respond effectively to epidemics in our own regions and support genomic surveillance across the world."

Professor Tulio de Oliveira
Director of the Centre for Epidemic Response and Innovation at Stellenbosch University and Deputy Director of the Genomic Surveillance Unit at the Wellcome Sanger Institute

GSU GENOMIC SURVEILLANCE UNIT

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

CERI Centre for Epidemic Response and Innovation

Innovation

In this section

1

Clinically approved tool predicts blood cancer risk

Predict Blood is a validated online calculator developed by Sanger scientists that uses genomic data to identify patients at risk of developing aggressive cancers such as acute myeloid leukaemia. The software, using data analysed by the Cancer, Ageing and Somatic Mutation programme, can help NHS doctors to determine which patients might need active clinical monitoring.

Approximately 30,000 people in the UK have myeloproliferative neoplasm (MPN) blood cancer. While most patients can be managed for years, some will develop more aggressive blood cancers that require active treatment. Predict Blood, registered with the UK Medicines and Healthcare products Regulatory Authority (MHRA), combines clinical and genomic information to help doctors make personally tailored disease predictions to guide patient management.

Until recently, the only classification system for MPNs was one from the 1950s, which had challenges and divided the conditions into three clinical types. To better understand the biological factors at work in MPNs, Sanger scientists worked with researchers from Wellcome-MRC Cambridge Stem Cell Institute and the University of Cambridge to study 69 cancer genes from more than 2,000 patients. The team identified eight new genetic subtypes that were clinically different.

To inform prognoses for new patients, the team developed an algorithm that combines clinical and genomic data to place individuals into one of the eight new subtypes. The method outperformed previous classification schemes and gave patient-specific predictions, rather than simply classifying patients into broad risk categories.

Sanger scientists, with funding from the Institute's Genomics Innovation Office, developed the Predict Blood website based on their algorithm. For this, the Institute drew on the Winton Institute's knowledge of communicating risk to ensure that that the website's results and descriptions were transparent and clear.

The Sanger Institute is looking for external partners, including commercial entities, to maximise access to the tool globally, enabling long-term sustainability.

Reference
Grinfield J. *et al. New England Journal of Medicine* 2018; **378:** 1416-30.

https://blood.predict.nhs.uk

2

Sanger spin outs prove their worth

In the arenas of genomic clinical diagnosis, microbiome-mediated medication and cancer drug development, Sanger Institute science is delivering real-world impact through three innovative companies.

The Sanger Institute specialises in blue-sky, large-scale genomic science that produces discoveries and methodologies that could revolutionise research and clinical practice. To secure the funding necessary to deliver meaningful impact, spin-out companies are formed.

Congenica offers clinical decision support software to researchers, clinicians and biopharmaceutical companies around the world. Based on pioneering research from the Sanger Institute and the UK NHS, its platform analyses the entire human genome by collecting trillions of data points to interpret and understand them in detail. It is the exclusive Clinical Decision Support partner for the NHS Genomic Medicine Service.

In 2023, it formed strategic partnerships with myTomorrows to enable doctors worldwide to find pre-approval treatments and active clinical trials for treatment-resistant patients, and with NoorDX to deliver genomic sample-to-report services across the Middle East. In addition, it secured a new two-year contract with the Hong Kong Genome Project.

Microbiotica utilises painstaking microbiome research to replicate and understand the human gut microbiome in health and disease. The company's microbiome platform contains the most diverse human gut microbiome collection worldwide, coupled with a Reference Genome Database covering the global human microbiome, advanced informatics and machine learning.

In partnership with MSD, one of Microbiotica's live biotherapeutic products, MB097, is undergoing a phase 1b clinical trial in melanoma patients with primary resistance to anti-PD-1-containing immunotherapies.

Mosaic Therapeutics is based on a decade of research into cancer cell genetic vulnerabilities. It combines large datasets, experimental approaches and advanced computational methods to identify and develop targeted cancer therapies for biomarker-stratified patient populations. In 2023, it raised £22.5 million Series A funding.

Congenica

microbiotica

MOSAIC TX

3

Using a lifetime's DNA changes to make medicines

Quotient Therapeutics is a new company that will build on the Sanger Institute's deep expertise and insights into DNA variation present in the body's trillions of cells to identify new approaches to treating disease.

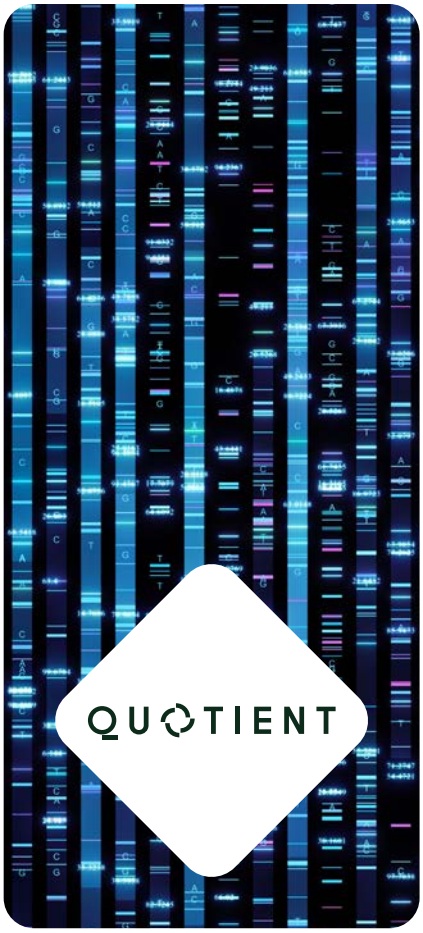
Using techniques developed at the Sanger Institute, the Quotient Somatic Genomics platform will study the genetic variation our bodies build up over time at the cellular level to develop novel therapeutics informed by new links between genes and disease.

The company will investigate how random genetic changes that human cells accumulate over a person's lifetime, and the selective pressure they may be subjected to, result in trillions of unique genomes in the body. Some of these genetic changes will make a cell more resistant or more vulnerable to disease, while others will cause disease directly. Quotient will study this natural library of DNA variations in every tissue to find those that are neutral, beneficial, harmful or disease-causing, and then use that knowledge to develop new therapeutics.

The company is a collaboration between Flagship Pioneering, the Sanger Institute and the University of Texas Southwestern. In their first major investment outside of the US, Flagship Pioneering has made an initial commitment of £41 million to advance the development of the company's platform and to pursue a pipeline of medicines across a wide range of therapeutic areas, including immune disease, cardiometabolic disease, infectious disease, oncology, neurodegenerative disease, rare disease and ageing.

> To think that the concepts and technologies that we have developed could translate into patient benefit has been a source of great excitement.

Dr Iñigo Martincorena
Group Leader in the Cancer, Ageing and Somatic Mutation programme, Wellcome Sanger Institute



Culture

In this section

1

Sanger is a UK Science Council Employer Champion

The Institute's commitment to professional development and registration for its staff to drive the quality and practice of its science has been recognised.

The Employer Champion Programme award is granted by the UK's Science Council, which is the guardian of professional standards for scientists and science technicians through their professional registration awards. These standards aim to ensure best practice and integrity in the science workforce.

In a ceremony in July 2023, the Institute was given the status by Helen Gordon, Science Council CEO, and praised for its work on Sanger's Technician Commitment programme. The Institute joined the Technician Commitment initiative in 2018 to develop and recognise its highly skilled technical employees, who are vital for the organisation's achievements. To further enhance this work, in 2019, the Institute created a Technician Commitment manager role – the first of its kind in the UK.

Since then, the initiative has grown in both size and ambition in its efforts to support technical staff at the Sanger Institute. With more than 700 members and four working groups, Sanger's Technician Commitment looks at a wide range of topics such as visibility, recognition, career progression and development support.

700 members

4 Working Groups

1 Technician Commitment Manager

This accomplishment was made possible through the encouragement of senior leaders, scientific managers and, importantly, the ownership of the Sanger Technical staff.

Burcu Bronner-Anar
Technician Commitment Manager at the Sanger Institute

2

Sanger supports postdocs returning from extended leave

Postdoctoral research contracts are typically 3-5 years long, making periods of extended leave difficult. As part of the Institute's work with Athena SWAN and the Technician Commitment, the Equality Diversity and Inclusion team have developed a scheme that partners returners with technical staff to recover lost productivity.

Postdoctoral researchers play a key role in the Sanger Institute's science, delivering specific time-bound research projects to advance understanding. However, the nature of their contracts means that any extended leave – for example due to parental or carer responsibilities – significantly affects their ability to deliver impactful outputs and could hinder their careers. Many family friendly benefits are already available; postdocs taking maternity leave are granted a contract extension of equivalent duration to their absence, for example.

However, to broaden the support available for all returners needing to regain research momentum, the Sanger Institute has additionally created the Research Support Scheme.

Originally developed in response to the impact of the COVID-19 pandemic, the scheme pairs returning postdocs with a research assistant. The postdocs are given a vital 'capacity boost' to restart and deliver their science more quickly and efficiently. The technical experts benefit from training and mentoring in new areas of research, offering career development opportunities.

The scheme was a great success, empowering both partners.

A technical specialist reported: "I now better understand the research process, experimental design and how decisions are made. By using very different methods to the ones I use in my everyday job, I also got a lot more experience, which was great to build my CV."

Spurred by this, the Institute has extended the scheme to all returning postdocs as part of its flexible package of support.

I have no words to say how incredible the support was because I had someone who I could trust to conduct my research. I could take a step back from my work too, which helped me think about and plan a publication. It also helped me develop my skills as a mentor.

Suruchi Pacharne
Technical Specialist and former postdoctoral researcher in the Cancer, Ageing and Somatic Mutation Programme

3

Building equitable research collaborations

The Sanger Institute has introduced comprehensive guidelines for equitable research collaborations, aimed at ensuring mutual benefits for all parties involved. Developed through extensive consultation with collaborating researchers, Sanger research teams, funders and ethics experts around the world, the guidance lays the foundations for sustainable long-term research partnerships.

Recognising the importance of embedding equity into all the Institute's research, the Policy and Advocacy Team undertook a two-year initiative to develop guidelines that would offer clarity for Sanger researchers when establishing scientific collaborations. In partnership with experts and stakeholders worldwide, the entire life cycle of a research project was meticulously explored.

A comprehensive literature review revealed key issues spanning:

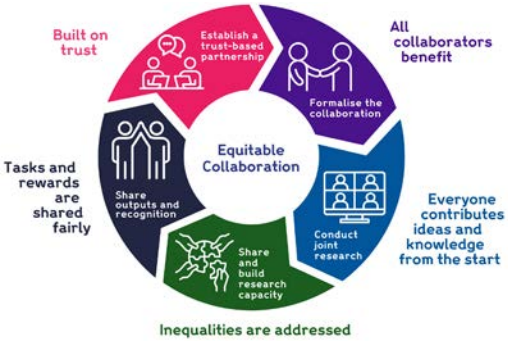
- project inception
- funding requests
- contract negotiations
- research implementation
- capacity building
- sharing outputs and recognition.

The team then worked with leading ethics researchers and social scientists to develop a set of draft principles. These were then deliberated upon in two roundtable discussions, in partnership with the Centre for Science and Policy (CSaP) at the University of Cambridge.

The first roundtable was conducted virtually to enable greater participation from genomic scientists in low- and middle-income countries. They highlighted the need to engage all partners at the project's inception to ensure alignment with the needs of collaborating researchers and communities to maximise impact and benefits. In addition, they stressed the need to share research capacity between collaborating organisations as the project progresses.

The second roundtable gathered ethicists, funding bodies and UK-based scientists with established collaborations in low- and middle-income countries to further refine the guidelines. The resulting in-depth guidance and quick-reference summary is now available to all Sanger researchers.

The Equitable Collaborations lifecycle



Influencing Policy

In this section

1

Rejoining Horizon Europe boosts Sanger's superpower

The Sanger Institute is delighted that the UK has rejoined Horizon Europe, enabling our scientists to use their greatest power: collaboration.

Since Brexit, the Sanger Institute has warned of the damage that leaving Horizon Europe would bring to UK science and urged the Government to negotiate terms to rejoin the European Commission's flagship funding framework. Now that these warnings have been heeded, the Institute looks forward to leading and contributing to the next wave of genomic discovery.

Collaboration lies at the heart of the Institute's science. Almost all Sanger's research is based on partnerships that cross national, continental and political boundaries.

Partnership and collaboration are vital as we attempt to tackle scientific challenges that are beyond the scale of any one organisation or nation. For the Institute, rejoining Horizon Europe was less about the money and more about what the funding enables – large collaborative scientific projects, the true superpower of science.

The Sanger Institute has led and contributed to many world-leading and fruitful collaborations across Europe, including the Human Cell Atlas, PERSIST-SEQ, Biodiversity Genomics Europe and The European Commission Joint Action 'Towards the European Health Data Space', and we look forward to many more to come.

> Participation in Horizon Europe will catalyse and fund the kinds of collaborations that deliver transformative genomics that benefit health.
>
> **Dr Sarion Bowers**
> Head of Policy, Wellcome Sanger Institute

[Proposals to cut net immigration] threaten our science, but more importantly are causing distress and alarm to individuals and their families. They brought their skills and talents to this country and now face enormous uncertainty about their status and the costs they face.

Dr Sarion Bowers
Head of Policy, Wellcome Sanger Institute

2

Keeping the UK open for science

The Sanger Institute's world-leading science is founded on the creativity and talent of its staff, drawn from the brightest minds from around the globe. The Institute is a leading opponent of moves to impede researcher immigration, which would make the UK a less appealing destination for the next generation of genome scientists.

The Institute has worked closely with fellow institutes, universities and research organisations to ensure that genomic science is not damaged by measures to reduce immigration. From proposed increases to visa application costs to salary-based limits on eligibility, the Institute has provided clear and unequivocal advice.

We work hard to attract and retain specialist staff to deliver our transformative science. We develop local talent through apprenticeships, a fully funded PhD programme, dedicated Postdoctoral Excellence Fellowships and Fellowships for those returning to science. But we also need to recruit internationally to generate our cutting-edge research.

We employ talented staff from more than 70 countries. But proposed changes to immigration requirements have fuelled concerns among our staff that they and their families may no longer be able to remain in the UK. Some have even asked whether they should make arrangements to leave the UK.

The Institute is proud to offer competitive salary packages, but proposed increases in salary thresholds will make many of our research positions inaccessible to scientists outside of the UK. In addition, a sizeable number of our researchers join us via the Skilled Worker visa route, and clarity is needed as to whether concessions for recent graduates, people with a PhD relevant to the job and for Shortage Occupations will continue to apply, and on what basis.

The Government says that it is committed to attracting and retaining top researchers and innovators. In partnership with its collaborators, the Sanger Institute's policy team continues to hold the Government to its promise.

3

Sanger supports the MESSAGE

The Sanger Institute has joined with organisations across the UK research sector to support a first-of-its-kind sex and gender policy for biomedical, health and care studies. The policy will cement UK research's position as a world leader in rigorous, sustainable science and will provide the most effective evidence to improve outcomes.

Sex and gender play fundamental roles in individual and population health. These factors influence:

- the medical conditions people develop
- the symptoms they experience
- the treatments and quality of care they receive
- their disease progression and their overall outcomes.

However, the majority of research studies have been conducted primarily on men, male cells or male animals, biasing diagnostic criteria and health outcomes.

To ensure that vital insights dependent on sex and gender are not missed requires research to build consideration of sex and gender into every stage: from study design and recruitment to data analysis and transparent reporting of results. Yet unlike other high-income countries – notably Canada, the United States and European nations under Horizon Europe – the UK currently has no standard, unified guidance for researchers about how adequately to consider sex dimensions in cell and animal studies and sex and gender dimensions in human studies.

To address this, the Sanger Institute has signed up to the sex and gender policy co-designed by The George Institute for Global Health's Medical Science Sex and Gender Equity (MESSAGE) project for high-quality, reproducible and inclusive biomedical, health and care research. The project, funded by the Wellcome Trust, collaborated with research funders, regulators, researchers, patient and public groups, academic publishers and the Department of Health and Social Care to create an in-depth policy framework for the UK biomedical research sector. As a signatory, the Sanger Institute will adapt and integrate these best-practice recommendations into its research.

MESSAGE
Medical Science Sex and Gender Equity

The Institute is committed to providing the research community with the guidance, skills and tools to ensure that future research meets the needs of all people, no matter their sex or gender. We are developing an organisational strategy on representative research to ensure that our research benefits everyone, including those of underrepresented groups.

Hayley Clissold
Policy Lead, Wellcome Sanger Institute

30 years of Impact

In this section

1

Providing reference genomes to explore biology

Over the past 30 years, the Sanger Institute has generated reference genomes for many of the most important organisms in life sciences research: from disease-causing bacteria, viruses and parasites to model organisms and key species in diverse ecosystems.

Originally founded to help deliver the first human reference genome, the Wellcome Sanger Institute contributed one third of the total DNA sequence. As the only UK partner in the global consortium, the Institute played a leading role in enabling a huge shift in understanding of the biological basis of human health and disease and establishing the global practice of open data sharing.

Building on our scientists' hard-won expertise, the Sanger Institute went on to provide the reference genome sequences of most of the widely used model organisms in the life sciences – yeast, *E. coli*, *C. elegans*, mouse, pig and zebrafish. The organisms have provided deep understanding, how cells and tissues function and have been used extensively to develop diagnostics, therapeutics and biotechnology.

In parallel, the Institute has collaborated with partners worldwide to systematically generate reference genomes for many viruses, parasites and bacteria responsible for human infectious disease. From cholera to malaria, from dysentery to MRSA and from leprosy to tuberculosis, using genomic surveillance to study the transmission and evolution of these diseases has revealed new insights and therapies and guided public health vaccination and drug treatment campaigns.

More recently, the Institute has embarked on its most ambitious project so far – supporting the delivery of high-quality reference genomes for all known eukaryotic species in Britain and Ireland. The Darwin Tree of Life project is providing reference genomes that will help guide conservation efforts and enable greater understanding of how ecosystems adapt to climate change across the UK and the world.

2

Tackling cancer and ageing with genomics

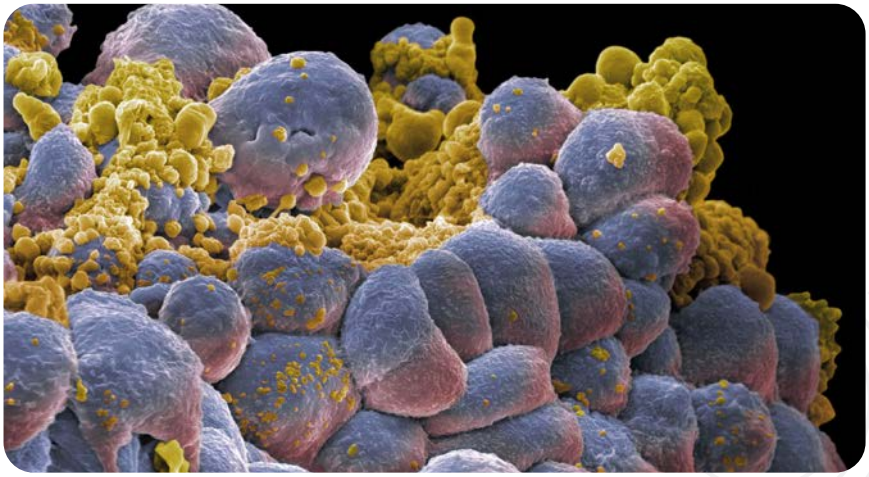
Diseases of ageing and cancer are often the result of the DNA changes that have accumulated in a person's genome over their lifetime. For the past 24 years, the Sanger Institute has identified these changes, and their environmental or cellular causes, in ever-greater detail.

In 2000, the Institute established its Cancer Genome Project to identify those genes whose variants contributed to uncontrolled cell growth. Since this time, Sanger scientists and collaborators have systematically read the DNA of tens of thousands of cancer genomes, contributing to the 800 cancer-driving genes identified to date.

The Institute's researchers have developed new laboratory and computational techniques to identify and understand the mutational processes underlying cancer development. This knowledge has led to the successful development of new cancer treatments and improved diagnostics, provided new tools for early detection and offered insights into preventative approaches for those at risk. These discoveries inform current clinical practice through online tools that combine genomic and clinical data to guide healthcare professionals' cancer treatment decisions.

New single-cell sequencing and bioinformatics approaches developed within the Cancer, Ageing and Somatic Mutations programme are now providing fine-grained understanding of how complex, long-term health conditions develop. The discovery of 'driver' mutations – similar to those found in cancer – in age-related and complex disease conditions has raised fundamental questions and opportunities in their detection and treatment.

A spin out company, based on these new technologies, has been founded to develop and deliver transformative treatments and diagnostics in immune disease, cardiometabolic disease, infectious disease, oncology, neurodegenerative disease, rare disease and ageing.



3

Exploring the human condition with genomics

From capturing the diversity of human genetic variation, and identifying the roots of developmental disorders to mapping the precise cellular makeup of the body, the Sanger Institute is at the forefront of genomic discovery.

Sanger researchers have made significant contributions to multiple projects that studied how the variations in the human genome contribute to health and disease. The discoveries from projects that sequenced ever-larger sample sizes, including HapMap Project, 1000 Genomes Project, UK10K project and UK BioBank, provided key information on genome and protein function that has powered subsequent exploration of human evolution, migration and disease through genome variation.

Through contributions to the UK collaborative Wellcome Trust Case Control Consortium using genome-wide association studies, Sanger researchers discovered numerous genes implicated in common diseases.

In 2010, Sanger researchers embarked on the landmark Deciphering Developmental Disorders (DDD) study to identify the gene variants contributing to rare childhood diseases. So far, 70 previously unidentified genes whose variants contribute to severe inherited developmental disorders have been uncovered, directly providing genetic diagnoses to more than 5,000 families.

Through the Prenatal Assessment of Genomes and Exomes (PAGE) study, new inherited gene variants were identified during prenatal ultrasound screening. These data contribute to the DECIPHER database – co-founded by the Sanger Institute – which underpins a wide range of rare disease research and informs clinical practice.

Sanger Institute scientists co-founded and co-lead the Human Cell Atlas (HCA), an international consortium that uses single-cell transcriptome sequencing to create a comprehensive catalogue of cell types (and their functions) in the human body. So far, the Institute has contributed to a wide range of cellular maps detailing the composition and functions of the heart, lung, placenta and limbs in the adult body and during development. These atlases promise to help transform medical research and healthcare worldwide.


DELL

We embed equality and mutual benefit at the heart of our collaborations

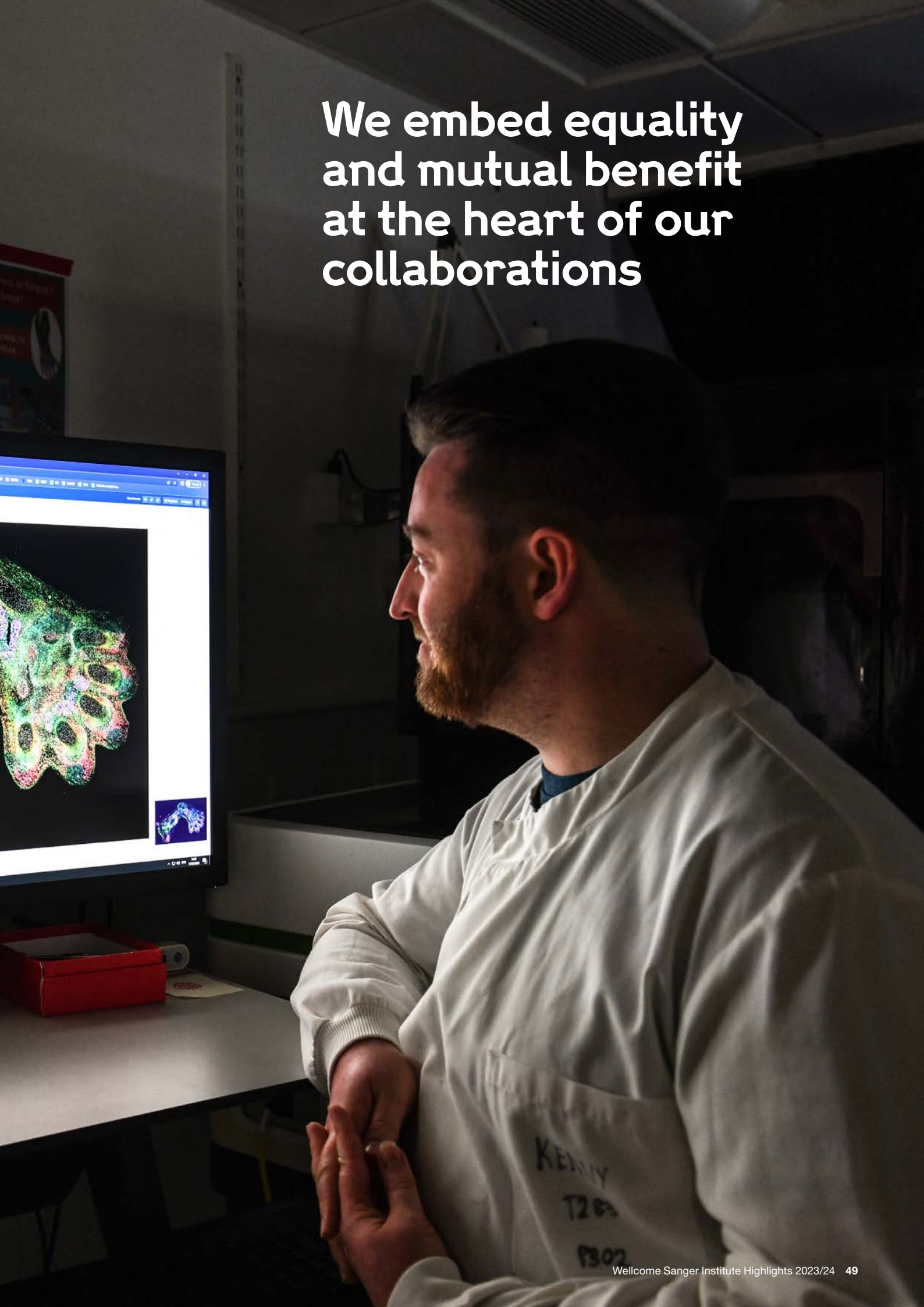


Image Credits

All images in this Annual Highlights document belong to the Wellcome Sanger Institute and have been produced by commissioned photographers (David Lavene, Greg Moss, Phil Mynott and Dan Ross) or by Sanger Institute employees, or have been sourced from AdobeStock, except where stated below:

Cover	Scanning electron microscope image of *Trichuris muris* – Dave Goulding / Wellcome Sanger Institute

2023 Timeline

Page 6	Immune cells and red blood cells – Annie Cavanagh, CC BY-NC 4.0 Lung cells – Nathan Richoz, University of Cambridge Icefish – Pete Bucktrout, British Antarctic Survey
Page 7	Yolk sac – Megumi Inoue, Institut de la Vision Scanning electron microscope image of cholera bacteria – Dave Goulding / Wellcome Sanger Institute Developing human limb bud – Alain Chédotal and Raphaël Blain, Inserm

Cancer, Ageing and Somatic Mutation

Page 15	Red blood cells and sickled cell – Janice Haney Carr / CDC PHIL

Cellular Genetics

Page 16	Yolk sac – Megumi Inoue, Institut de la Vision
Page 17	SARS-CoV-2 virus – Alissa Eckert, MSMI, Dan Higgins, MAMS Lung cells – Nathan Richoz, University of Cambridge
Page 18	Developing human limb bud – Alain Chédotal and Raphaël Blain, Inserm

Human Genetics

Page 21	Prime editing – Laura Olivares Boldú, Wellcome Connecting Science
	Mouse brain fluorescence slides – *Brain* 2023; **146:** 4766–4783

Parasites and Microbes

Page 24	Child vaccination – icddr,b
Page 26	Scanning electron microscope image of *S. mansoni* – Dave Goulding / Wellcome Sanger Institute Scanning electron microscope image of *E. coli* – David Gregory and Debbie Marshall, Wellcome Images, CC-BY-4.0 Scanning electron microscope image of *E. coli* – Dave Goulding / Wellcome Sanger Institute

Tree of Life

Page 29	Icefish – Pete Bucktrout, British Antarctic Survey

Impact

Page 36	Diagram of pangenome – Megan Gozzard, Wellcome Sanger Institute

Innovation

Page 40	Scanning electron microscope image of red blood cells – Annie Cavanagh, CC BY-NC 4.0

30 years of Impact

Page 46	*C. elegans* – ZEISS Microscopy, Creative Commons Attribution 2.0 Scanning electron microscope image of *E. coli* – David Gregory and Debbie Marshall, Wellcome Images, CC-BY-4.0 Yeast cells – Mogana Das Murtey and Patchamuthu Ramasamy, Creative Commons Attribution 3.0
Page 47	Scanning electron microscope image of cancer cells – Annie Cavanagh, CC BY-NC 4.0

Wellcome Sanger Institute Highlights 2023/24

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 Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA.

First published by the Wellcome Sanger Institute, 2024.

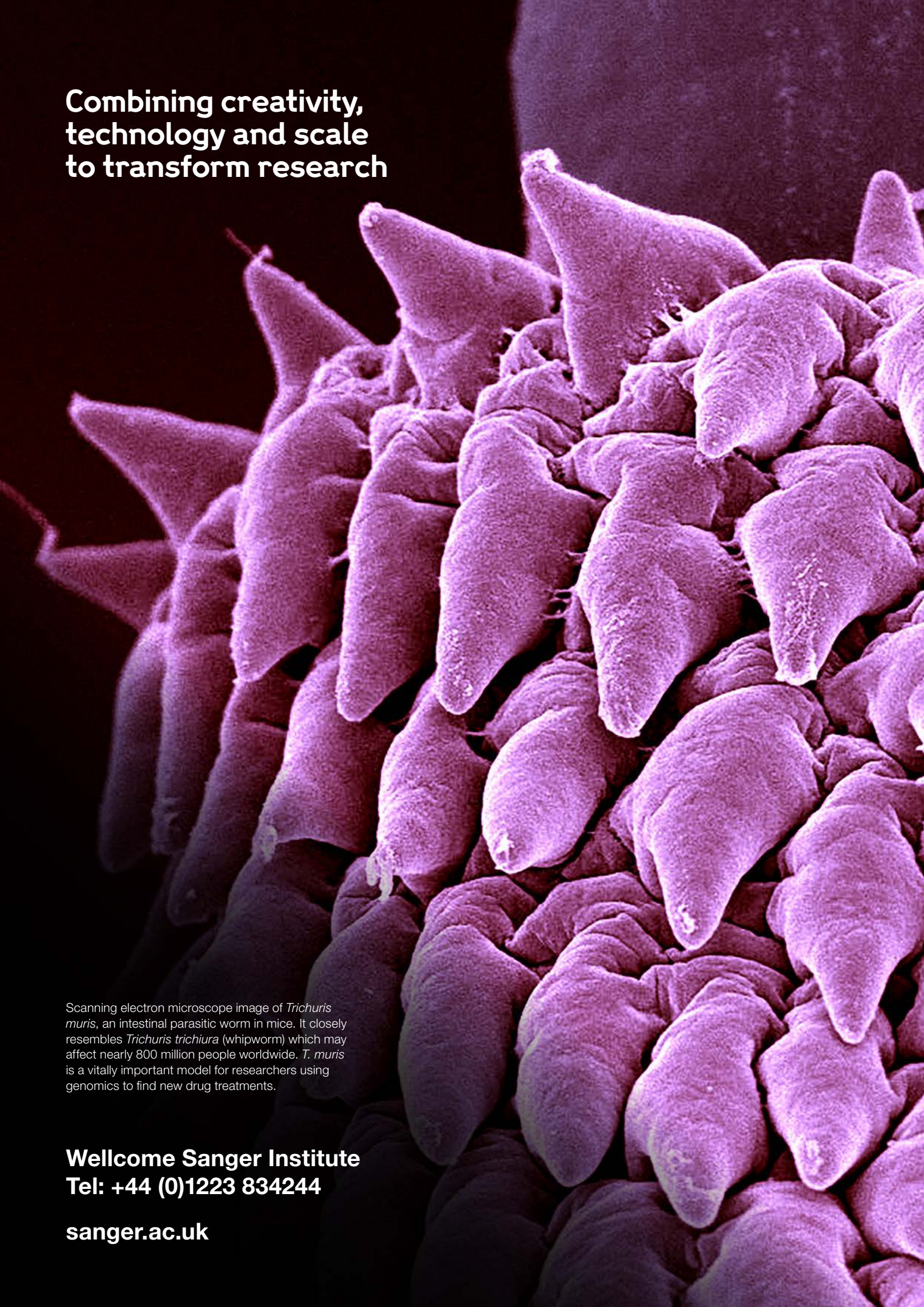
Printed by Kingfisher Press
on FSC® certified paper.

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

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

Designed and produced by **MadeNoise**

Combining creativity, technology and scale to transform research



Scanning electron microscope image of *Trichuris muris*, an intestinal parasitic worm in mice. It closely resembles *Trichuris trichiura* (whipworm) which may affect nearly 800 million people worldwide. *T. muris* is a vitally important model for researchers using genomics to find new drug treatments.

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

sanger.ac.uk