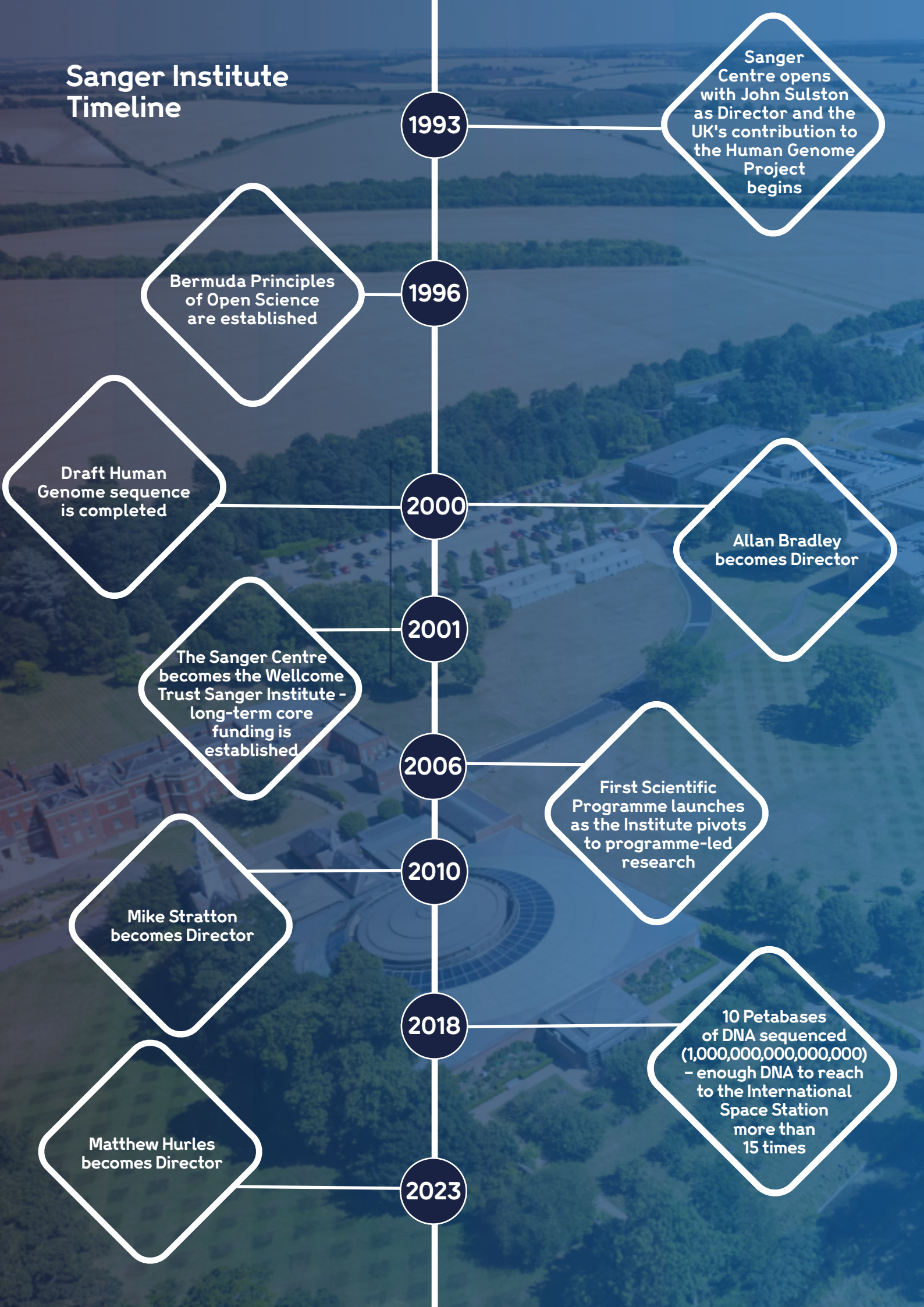
Impact Report 1993–2023

Wellcome Sanger Institute

wellcome sanger institute

Sanger Institute Timeline



1993

Sanger Centre opens with John Sulston as Director and the UK's contribution to the Human Genome Project begins

1996

Bermuda Principles of Open Science are established

2000

Draft Human Genome sequence is completed

Allan Bradley becomes Director

2001

The Sanger Centre becomes the Wellcome Trust Sanger Institute - long-term core funding is established

2006

First Scientific Programme launches as the Institute pivots to programme-led research

2010

Mike Stratton becomes Director

2018

10 Petabases of DNA sequenced (1,000,000,000,000,000) – enough DNA to reach to the International Space Station more than 15 times

2023

Matthew Hurles becomes Director

Introduction

Thirty years of world-changing science

In 1992 when the Wellcome Trust first agreed to fund the creation of the Sanger Centre to deliver the UK's contribution to the Human Genome Project, few could have envisaged the impact the organisation would have on science worldwide.

In partnership with researchers around the globe, the Sanger Institute has changed the paradigm of life sciences research. The Institute helped to establish the principles of free, open-access data sharing and has delivered a wide range of foundational genomic and biological resources.

The science conducted by the Wellcome Sanger Institute opens up new fields of discovery, powers equitable global research, and delivers insights for clinical, societal and conservational benefit.

This report provides a brief insight into some of the outcomes of the work by our scientists, technical experts and staff over the past 30 years. We look forward to the next 30.

Professor Matthew Hurles, Director, Wellcome Sanger Institute

Contents

Mission and Vision

We seek to understand life in all its forms to change how the world thinks about biology and human health by conducting ambitious, world-leading science at a scale few can match

The Wellcome Sanger Institute is a world leader in genome research that delivers insights into human, evolutionary and pathogen biology.

Our aim is to provide insights into human, parasite and microbe evolution, cellular growth and activity, the processes that underlie mutation and tumour formation, and the diversity of complex life around the world.

To drive the next wave of global research, we promote, support and openly share genomic knowledge and practices to benefit academic research, medical practice and society.

In order to achieve our vision and deliver on our mission to "apply and explore genomic technologies at scale to advance understanding of biology and improve health", we commit to:

Introduction to the Impact Framework

To enable assessment of the Sanger Institute's impact, we have developed the impact framework to guide reporting and strategic thinking.

In measuring impact, we differentiate between inputs, outputs, outcomes and impact. Outcomes can be, and often are, used as proxy measurements for impact, because of the similar emphasis on downstream effects.

The framework focuses on the impact on the outside world across six key areas of our activities relating to Research, Innovation and Learning and Engagement:

1. Narratives of research
2. Bibliometric analysis of research publications
3. Data resources generated
4. Incubating future leaders in genome research
5. Spinouts
6. Knowledge exchange.

These six areas cannot capture every part of our work. They provide insight into key areas where we seek to evaluate our work against our mission and vision.

The Sanger Institute's main research focus and activity is on delivering fundamental discoveries to power global science. We evaluate our impact both quantitatively (through bibliometric analysis) and qualitatively (through Narratives of Research). Research outputs are also shared through data resources. The Sanger Institute has a long-standing tradition of nurturing and training the future leaders in genomics research.

This impact report seeks to provide a record of the Institute's achievements over the 30 years of its existence.

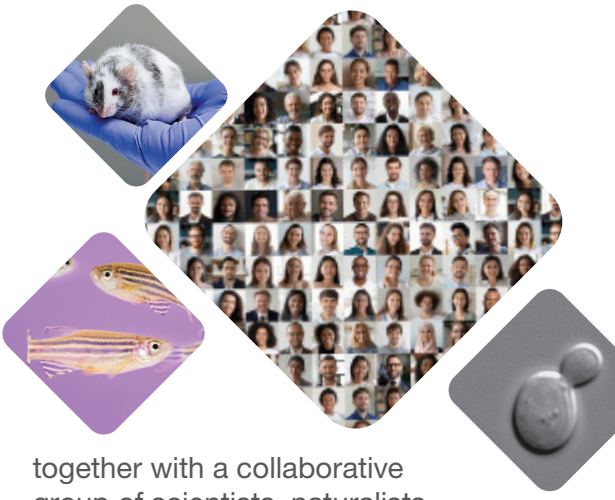
The Wellcome Sanger Institute Impact Framework

Research impact:

Narratives outlining the key research contributions of the Wellcome Sanger Institute 1993–2023

The following case studies highlight some of the key achievements of the Sanger Institute over the past 30 years.

Providing fundamental resources for understanding biology: reference genomes



The Wellcome Sanger Institute was the only UK partner in the global Human Genome Sequencing Consortium that completed the first draft of the complete human genome, contributing one third of the total sequence. The reference human genome sequence is considered a major milestone in the history of biology and in the advancement of knowledge generally, enabling a huge shift in understanding of the biological basis of essentially all human diseases and permanently transforming global practice in open data sharing.

The Sanger Institute went on to provide the reference genome sequences of most of the widely used model organisms informing on human biology – yeast, *E.coli* bacteria, *C. elegans* nematode worm, mouse and zebrafish, which have all been essential experimental tools for understanding biology and extensively used in translating research into successful development of therapeutics, biotechnology and genetic engineering. Their use has had major impact on human health and economic impacts for the biotechnology and the pharmaceutical industry.

Infectious disease remains a major global cause of death, and the Sanger Institute, with others, have made systematic efforts to provide the reference genome sequences of all human infectious disease pathogens. This has provided new avenues to their management, prevention and treatment.

Providing a reference genome sequence – a blueprint - for all species on Earth was once a distant aspiration since the discovery of the structure of DNA. The Sanger Institute, together with a collaborative group of scientists, naturalists and sample collectors, are spearheading efforts to deliver high quality reference genomes of all known eukaryotic species of the British Isles, by far the largest such initiative currently in the world. It leads the way in providing the basis for deeper understanding of every form of life, with wide ranging practical applications from providing new insights into ecosystems to help adapt to climate change, to advancing the field of synthetic biology to reveal the mechanisms needed to produce a medicinal compounds found in plants, and many unforeseen applications.

Powering cancer diagnosis and treatment through genomics

All cancers are caused by somatic mutations, changes that occur over the course of an individual's lifetime in the copies of the human genome present in essentially every cell of the body. The Sanger Institute established its Cancer Genome Project in 2000 to identify the "cancer genes", which, when mutated, cause normal cells to become cancer cells.

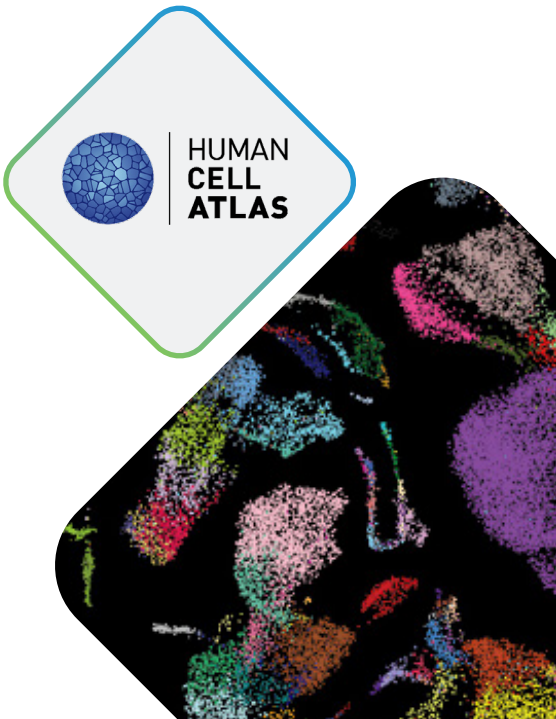
Sanger scientists and collaborators have systematically sequenced tens of thousands of cancer genomes, contributing to the 800 genes identified to date as playing a central role in the genetic and functional analysis of cancer cells. The discoveries, including the identification of the mutational processes underlying cancer development, have led to the development of successful new drugs for cancer treatment, new diagnostics, new approaches for early detection, new approaches for monitoring cancer burden and recurrence, and new insights for cancer prevention, informing current clinical practice. Cancer genome sequencing has also had substantial economic impact with many biotechnology companies created and pharmaceutical companies basing their drug strategies on the discoveries made.

Exploring the full natural variation in human genomes

Sanger researchers 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 sequencing of whole human genomes from 250,000 people in UK BioBank, provided information on genome and protein function that has been the basis for subsequent exploration of human disease through genome variation. It also influenced multiple waves of subsequent human genetics research including into human evolution, the spread of homo sapiens from Africa, the existence of other humanoid species and many other paths of enquiry.

Mapping all the cells and their functions in the human body

Cells are the basic units of life, yet our knowledge of cell types and their classification has been limited until recently. Co-founded and co-led by Sanger Institute scientists, the Human Cell Atlas (HCA) is an international consortium producing a comprehensive catalogue of cell types in the normal human body using single cell transcriptome sequencing. With the ultimate goal of mapping all 37 trillion cells of a human body, it is bringing together researchers across multiple scientific disciplines, influencing technology development and providing atlases of the cells to help transform medical research and healthcare worldwide



Identifying the roots of inherited genetic diseases for clinical use

The completion of the Human Genome project provided the framework for systematically investigating how variations in the human genome sequence contribute to susceptibility of inherited diseases.

Sanger researchers undertook systematic sequencing of the X chromosome and found many mutated genes causing learning disability.

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 a landmark Deciphering Developmental Disorders (DDD) Study to identify the mutated genes contributing to rare childhood diseases, which identified more than 70 new mutated genes contributing to inherited severe developmental disorders and directly providing genetic diagnoses to more than 5,000 families in the study. Through the Prenatal Assessment of Genomes and Exomes (PAGE) study, new inherited mutated genes were identified during prenatal ultrasound screening. This data contributes to DECIPHER database supporting rare disease research and continues to inform clinical practice today.

Diagnoses given to more than 5,000 familes with inherited severe developmental disorders

PAGE
Prenatal Assessment of Genomes and Exome

Variations in 70 genes contributing to rare childhood diseases identified

Understanding the genomic basis of ageing and disease

Much of the research has focused on the detection of somatic mutations in cancer genomes, with little knowledge of the mutation rates and mutational processes present in many normal human cell types. Technical obstacles have long stood in the way of understanding the landscape of differences between different cell and tissue types.

The Sanger Institute has led in establishing the foundations of this new domain of research, having developed multiple DNA sequencing approaches allowing detection of somatic mutations in normal tissues and discovering unexpected “driver” mutations, previously found only in the genomes of cancer and premalignant tumours, in some normal tissues. These have spurred fundamental new biological questions about normal human biology and highlighted opportunities to modify risk of cancer and other diseases, including the establishment of a new spin out company to follow through on therapeutic opportunities in this area.

Understanding the effect of every possible variation in the human genome

Artificially generating DNA mutations to study the effects of the engineered genes, proteins or strains of organisms, has been used to understand mechanisms of exactly how the changes in DNA sequence cause disease. Technical limitations has meant that a comprehensive understanding of all possible variants of human genes has been out of reach until very recently.

By undertaking mutagenesis at scale and computational analyses, Sanger researchers have been spearheading efforts to discover dependencies of every cancer cell in a patient to develop new therapies. Using new methodology, referred to as Multiplex Assays of Variant Effect (MAVE), Sanger Institute researchers are contributing to international efforts in generating comprehensive reference maps for every protein encoding gene and regulatory element in the human genome by creating Atlas of Variant Effect maps.

Using large-scale mutagenesis to test predictive models and study all aspects of DNA variation and protein function is setting the foundations for a new era of bioengineering, enabling scientists to design and produce new proteins and small molecules to bring in a new wave of therapeutics based on a refined understanding not possible before.

Tracking infectious diseases with genomic surveillance

By their nature infectious diseases are transmitted between human beings and/or other species. Understanding the nature and patterns of historical and ongoing spread provides insight into their underlying biology, identifies new strains that may be emerging and highlights opportunities for intervention. Sanger scientists have collaborated with investigators in many countries to understand the patterns of spread of a range of infectious diseases, including malaria, cholera, methicillin resistant staphylococcus (MRSA).

These studies provided fundamental scientific insights into pathogenic microorganism evolution with a degree of resolution in time and space that had not been previously possible.

Genomic tracking and surveillance of pathogens has informed multiple national public health control strategies for a number of infectious disease-causing microorganisms in the UK and globally. Recent application of using the acquired knowledge and capability of genomic surveillance at scale includes

contribution to provide 20 per cent of the world's total SARS-CoV-2 sequences during the Covid-19 pandemic, providing near-real time information for the UK Health Security Agency on the spread of the virus.

Discover more

The report on **Research impact** has more insight into selected Sanger Institute's scientific contributions over the past 30 years. Intentionally brief, these serve as a record of some of the most influential of the Institute's programmes of research and discoveries. Much of their impact lies in providing fundamental knowledge that has altered prevailing understanding of biology and the way biology research is conducted. It is difficult to quantify the impact of such contributions, which change what scientists think and what scientists do, as they underpin numerous advances made by others and continue enabling new developments in diagnostics and therapeutics.

Bibliometric Analysis of Research Publications

We communicate our research findings to the global scientific community through research articles and reviews that are published in peer-reviewed journals.

To quantitatively analyse our research publications and reviews, and to monitor our outputs over time, we undertake bibliometric analysis. This analysis complements the Narratives of Research by providing quantitative indicators of how others have been influenced by Sanger Institute research.

To help us maintain our position as a world-leading genomics research centre we use citation metrics to measure the impact that the Institute's foundational discoveries arising from Sanger Institute research have across the wider scientific community.

Sanger Institute's research is highly collaborative, with researchers contributing to multiple consortia across the globe. Despite a relatively small number of faculty, Sanger Institute produces several hundred publications each year. Citation metrics reflect Sanger Institute's global influence: the research articles are highly cited and referred to in further research conducted in countries across the globe.

Read the full analysis of Sanger Institute publications here.

Since 1996

Sanger Institute scientists have published **9,937** primary articles and reviews

These publications have been cited **1,388,662** times in further research

The research has been cited **4.28** times more than the global average (FWCI metric)

Our research publications are cited **140** times on average

On average **26%** of research articles and reviews are among the top 5% most cited in the world

Data Resources

We seek to continuously provide new insights into the biology of living organisms by producing data resources, annotating and analysing the genomic data and making it accessible.

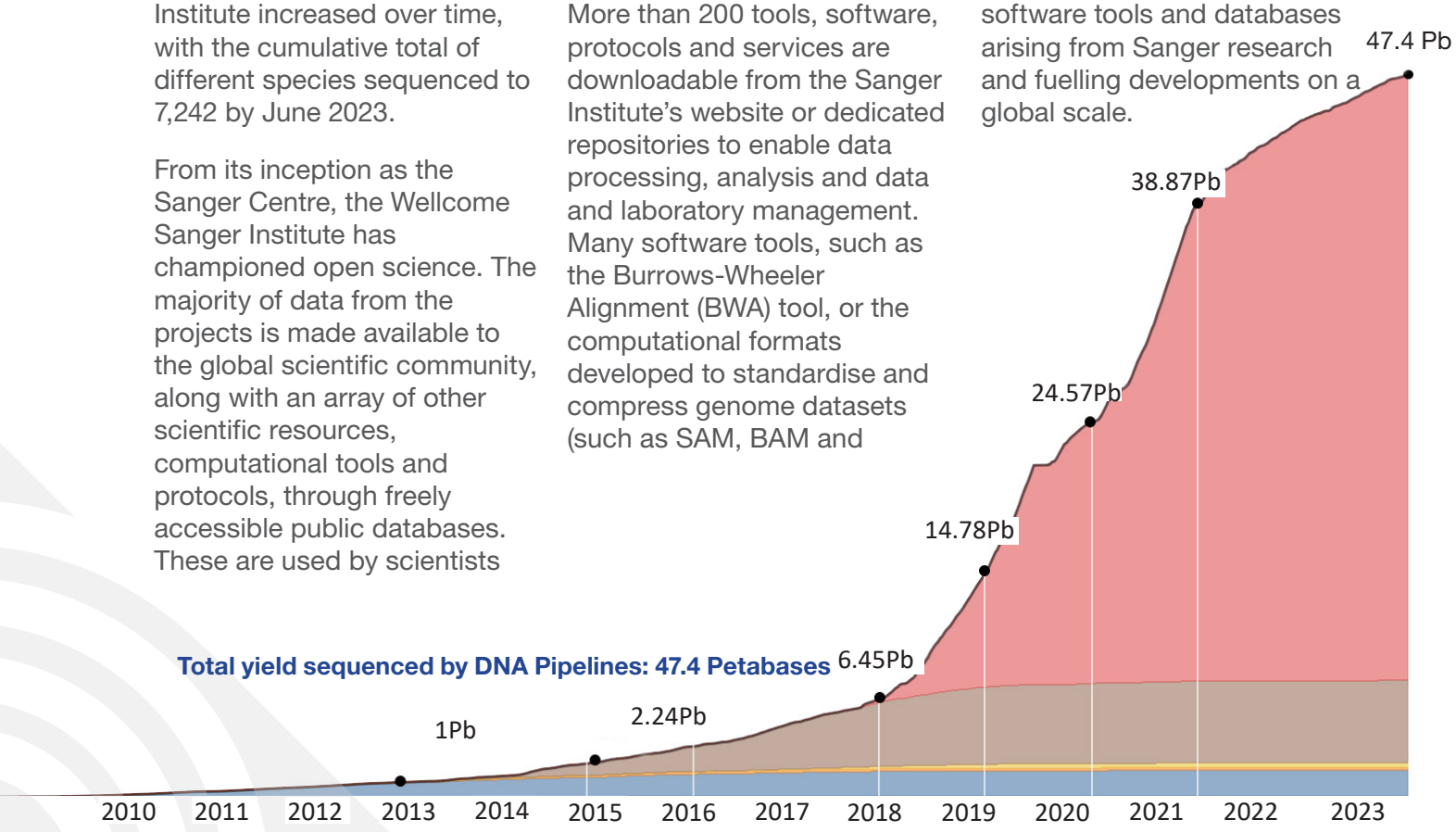
Genome sequencing at the Institute has generated a cumulative yield of over 46.6 Petabases of sequencing data by June 2023 see graph below.

The diversity of the species sequenced at the Sanger Institute increased over time, with the cumulative total of different species sequenced to 7,242 by June 2023.

From its inception as the Sanger Centre, the Wellcome Sanger Institute has championed open science. The majority of data from the projects is made available to the global scientific community, along with an array of other scientific resources, computational tools and protocols, through freely accessible public databases. These are used by scientists and industry worldwide and provide a comprehensive record of sequencing information, covering raw sequencing data, sequence assembly information and functional annotation.

More than 200 tools, software, protocols and services are downloadable from the Sanger Institute's website or dedicated repositories to enable data processing, analysis and data and laboratory management. Many software tools, such as the Burrows-Wheeler Alignment (BWA) tool, or the computational formats developed to standardise and compress genome datasets (such as SAM, BAM and CRAM) have been adopted as the industry standard for genomic data compression, storage and transfer and are still used today.

Read more here about the variety of data resources, software tools and databases arising from Sanger research and fuelling developments on a global scale.



Total yield sequenced by DNA Pipelines: 47.4 Petabases

Total sequencing contribution by the Sanger Institute, 2010-2023.

Incubating Future Leaders in Genomic Research:

PhD and MPhil programmes

We seek to train and inspire the next generation of genomic research leaders by providing an exceptional intellectual environment and infrastructure.

The Sanger Institute nurtures the next generation of genomic leaders with the goal of producing a constant flow of motivated scientists who have the necessary core skills, experience and outlook to become future leaders in genomic research. The strength of the Sanger Institute lies in the expertise of its diverse and inclusive community.

The Sanger Institute's faculty research groups lead thematic scientific research programmes, and the Associate Research Programme and partnership programmes diversify, enrich and strengthen the science portfolio.

Over 700 Technicians and Technical experts comprise a nearly half of the staff at the Institute. These key individuals and teams provide the practical application of knowledge, including hands-on support, directly contributing to our discovery science, teaching and learning, and enterprise activities.

At any one time over 100 postdoctoral fellows are embedded within the scientific programmes and research teams and make significant contributions to the Institute's scientific output. By September 2023, over 1,000 postdoctorale fellows have trained at the Sanger Institute.

The Sanger Institute has hosted and trained PhD students since it was founded in 1993, the Graduate Programme hosts 12 PhD students per year on a 4-year PhD programme and two fellowships per year through the University of Cambridge Wellcome PhD Programme for Clinicians. By September 2023, 375 PhD students and 39 MPhil students have studied at the Sanger Institute.

In addition, a one-year MPhil programme funds three MPhil students per year. All successful Sanger Institute MPhil and PhD students graduate with a degree from the University of Cambridge.

Sanger PhD students have maintained an average four-year submission rate of 89% over the past 12 years (93% in 2021/22), far exceeding the 70% threshold expected by the Research Councils. The data from the Sanger Institute's PhD and MPhil programmes on the career outcomes indicates that it consistently meets its goal of nurturing the next generation of leaders in genome research.

For more in-depth look at the PhD and MPhil programmes and their outcomes, click here.

Genomic Innovation:

Spinouts

We look to maximise the impact of our research to deliver benefits to medicine, society and the scientific community through developing new products and services

The UK is a leader in exploitation of genomics. It attracts more investment and is home to more genomics companies than any other country in Europe. This growing industrial ecosystem stems from early and visionary charitable investment in genomics that established, in the UK, foundational resources and capabilities to sequence, analyse and share the first human genome.

Thirty years on, the Sanger Institute still has a key role to play in this maturing ecosystem by producing discovery research at a scale that enables elucidation of mechanisms and factors affecting the development of disease, thus contributing insights that can change medicine.

The Institute achieves its impact on science and society through a variety of mechanisms including publications, knowledge and tools dissemination, influencing policies and training genomic scientists. The Innovation team adds to these capabilities aiming to increase the uptake and utility of our research.

Various methods are used to translate genomics and biodata; these include

- licensing of intellectual property (IP)
- collaboration with industrial partners
- formation of spin-outs.

The innovation vision strives to derive maximum impact from technologies or resources developed at the Institute, and adopts a broad definition of impact:

- on society – by working with industry partners and investors to apply our research to solve real world challenges;
- on scientists – by providing opportunities to translate their research and make a difference and by providing alternative career pathways;
- on science – by using commercial routes when appropriate to efficiently disseminate our tools and technologies and therefore enable the research of others.

Multiple new businesses have been spun out of Sanger Institute's research, including:

- Kymab
- Congenica
- Microbiotica
- Mosaic Therapeutics.

Read more on the spinouts here.

Knowledge Exchange:

The impact of training in genomic science

We share our knowledge and encourage meaningful dialogues within an ethos of open science across the Institute and beyond through data sharing, collaborations, conferences and training courses

The Wellcome Connecting Science has made dedicated contributions to capability building globally through training in technical skills and methodologies, creation of educational materials and schools outreach activities.

The programme was established at Guy's and St Thomas' Hospital, London in 1988 and moved to the Wellcome Genome Campus in 1998. Focusing primarily on the audiences of research scientists and healthcare professionals, it aims to improve the knowledge base, skills and careers in genome science.

The training in genomics is delivered in three ways:

- in-person training courses held on the Wellcome Genome Campus at the Hinxton Hall Conference Centre
- in-person/hybrid conferences to enable researchers worldwide to participate without the need to travel
- in-person training courses held at regional hubs in Latin America, Asia and Africa (Global Training), as well as online courses.

Between 2014-2019, 49 in-person conferences, 50 unique in-person courses on the Genome Campus and 18 unique courses in multiple LMIC regional hubs were held.

One of key ambitions of the programme is capacity building in low and middle income countries.

The courses and conferences have a global reach and are sought out by a large number of participants in over 100 countries, and the programme has reached more than 100,000 research and healthcare professionals across the globe during the 30 years of its existence.

Read more on the impact of the courses and conferences here.

Wellcome Sanger Institute Impact Report - 1993-2023

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, 2023.

We apply and explore genomic technologies at scale to advance understanding of biology and improve health

Wellcome Sanger Institute
sanger.ac.uk