


human
mouse
zebrafish
bacteria

10

Annual Review



The Wellcome Trust Sanger Institute is one of the world's leading genome centres. Since its inception in 1993, the Institute has played a key role in the development of genomic science, from sequencing of reference genomes to understanding the role of genomic variation in health and disease. Through its ability to conduct research at scale, it is able to engage in bold and long-term exploratory projects that are designed to influence and empower medical science globally. Institute research findings, generated through its own research programmes and through its leading role in international consortia, are being used to develop new diagnostics and treatments for human disease.

The Wellcome Trust is a global charitable foundation dedicated to achieving extraordinary improvements in human and animal health. It supports the brightest minds in biomedical research and the medical humanities. The Trust's breadth of support includes public engagement, education and the application of research to improve health. It is independent of both political and commercial interests.

Cover design: a graphical representation of the range of research carried out at the Institute. The four main organism types studied are human, mouse, zebrafish and pathogens (here represented by *Mycobacterium*). Behind each organism name is part of a DNA polymerase gene specific for that organism. DNA polymerase, present in all species, is the enzyme that replicates DNA at cell division and thus is intimately involved in evolution. The Institute was the leader in each of these genome sequence efforts. The background image is of *Clostridium difficile*. Using high-throughput whole-genome sequencing, Institute researchers are able to track *C. difficile* transmission and genome evolution within hospitals. Background image courtesy of Dave Goulding, Genome Research Limited.

04 ↘ A year in review

10 ↘ Human genetics

26 ↘ Cancer genetics and genomics

32 ↘ Mouse and zebrafish genetics

50 ↘ Pathogen genomics

62 ↘ Bioinformatics

72 ↘ Core facilities

80 ↘ The Institute, Education and Communication

The Institute's world-class resources in high-throughput sequencing and analysis enable our researchers in human, cancer, pathogen and model organism studies to work synergistically to deliver key insights into health issues across the globe.

» A year in review Mike Stratton

Welcome to the Wellcome Trust Sanger Institute's Annual Review for 2009/10.

During 2010, the Wellcome Trust Sanger Institute has restated and consolidated its mission of advancing understanding of human and pathogen biology through the study of genome sequences in order to improve human health. The major theme of our research continues to be the study of differences between genomes, naturally occurring or engineered, and the consequences of this variation. This underpins the work carried out in all our scientific areas of Human genetics, Cancer genetics and genomics, Pathogen genomics, Mouse and zebrafish genetics and Bioinformatics.

It has been a year of many changes at the Institute, some bringing a sense of progress and optimism, others much sorrow. The death of Leena Peltonen from cancer cast a dark shadow. Leena brought energy, spirit and boundless enthusiasm for human genetics to the Institute, together with her personal vibrancy and colour. In her short time with us, she wrought changes that will leave a lasting legacy, most notably in the nurturing of a new generation of young Faculty members in Human genetics.



Leena Peltonen (1952–2010)

Professor Leena Peltonen, Head of Human genetics at the Wellcome Trust Sanger Institute, died from cancer in March 2010 at her home in Finland.

Leena was one of the leading human geneticists of her era. After her medical and research training in Finland, she held positions at Rutgers Medical School, New Jersey, the University of California, Los Angeles, the University of Helsinki and Finland's National Public Health Institute. As well as being Head of Human genetics at the Sanger Institute, Leena maintained her positions in Finland and was a visiting professor at the Broad Institute in Cambridge, USA.

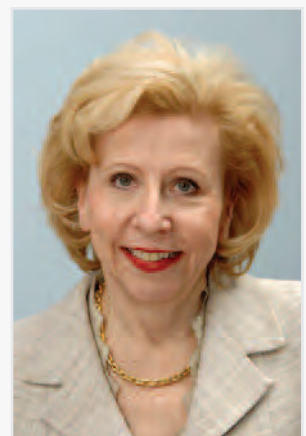
Leena was a visionary geneticist and champion of population genetics and public health, firmly supporting public health strategies alongside clinical interventions to

improve human health. She was highly influential in European medical science, working tirelessly to bring together human genetics activities across Europe.

Her knowledge and wisdom were called upon for national and international committees on population genetics, clinical genetics and ethics of genetics research. Leena served on the Boards of the European and American Societies of Human Genetics and a dozen other scientific organisations.

In her last six months, she was honoured with the Carter medal from the Clinical Genetics Society and was made an Academician of Science, the highest honour in her home country.

As well as her scientific achievements, Leena will be widely remembered as an inspirational mentor and leader, whose disarming smile, charisma and undeniable sense of style left a permanent impression on all who met her.



Professor Leena Peltonen's vision and charisma were an inspiration to many researchers: both for our Faculty and the more than 70 PhD students whose research she guided and nurtured.

There were a number of changes in management. Allan Bradley stepped down from the directorship, taking on the title of Emeritus Director. Four members of Faculty joined the Institute's Board of Management: Derek Stemple became Acting Head of Mouse and zebrafish genetics, Inês Barroso and Richard Durbin joint Acting Heads of Human genetics and Andy Futreal Head of Cancer genetics and genomics. We also welcomed Emma Millican, previously in the Forensic Science Service, to the Sanger Institute in a new role as Head of DNA pipelines to pull together all our platforms in this area.

In 2010, we submitted to the Wellcome Trust our scientific and corporate plan for the five-year period from 2011–16. This was very well received by the review committee and support for the forthcoming quinquennium was subsequently signed off by the Wellcome Trust Board of Governors. Our

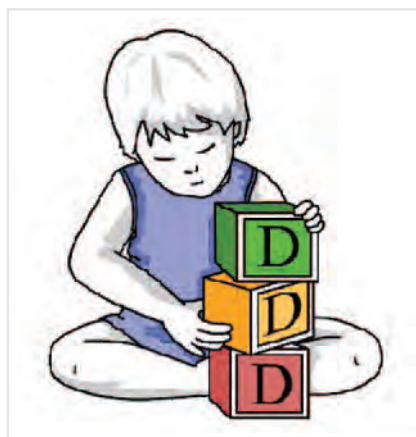
close and enduring relationship with the Wellcome Trust continues to provide the foundation of our scientific culture, allowing us to engage in large-scale and long-term exploratory projects that influence diverse areas of medical science.

Our funding was supplemented by several grants applied for with many UK and international collaborators. One notable example is the hugely ambitious UK10K Strategic Award from the Wellcome Trust to examine the genomes of 4000 healthy individuals and 6000 people with genetic disease. We are also benefitting from two Health Innovation Challenge Fund Awards: for monitoring cancer burden through measurements of cancer-derived DNA in blood samples and for the Deciphering Developmental Disorders (DDD) project, which will examine copy number variation in the genomes of 12 000 children with developmental difficulties.



UK10K

Funded by a Strategic Award from the Wellcome Trust, the UK10K project aims to help uncover disease-causing genes by studying the genetic code of 10 000 people in much finer detail than ever before.



Deciphering Developmental Disorders

The DDD project, funded from the Health Innovation Challenge Fund and Institutional funds, will see the Institute working in close partnership with the UK NHS to improve clinical diagnoses of developmental disorders.

The Institute's research has been much strengthened by the arrival of a number of new Faculty. In Human genetics, Nicole Soranzo joined us to investigate the genetics of lipid levels and other metabolic traits, while Carl Anderson's primary interests lie in genetic susceptibility to the inflammatory bowel diseases Crohn's and ulcerative colitis. Working primarily with mouse models, Darren Logan will be studying the biological basis of instinctive behaviours through genetic approaches, and Vijay Yadav will pursue his interest in bone metabolism.

In the area of Cancer genetics and genomics Ultan McDermott, supported by a Cancer Research UK Fellowship, is exploring how somatic mutations in cancers influence their responses to drugs. Peter Campbell, awarded

a Wellcome Trust Senior Clinical Research Fellowship, is investigating the genomes of many cancer types, notably those derived from white blood cells. George Vassiliou, on a Cancer Research UK Clinician Scientist Fellowship, is modelling the genetics of human leukaemia in mice. Trevor Lawley, who has a MRC New Investigator Research grant, is developing his investigations into the pathogenesis of infection by *Clostridium difficile* and other microorganisms.

The Institute emphatically supports collaborative research and endeavours to foster strong working relationships with others in the UK and internationally. So we were extremely pleased that Sharon Peacock, Professor of Clinical Microbiology at the University of Cambridge, agreed to become an Honorary Faculty member, consolidating her close links with many Sanger Institute scientists.

A number of our Faculty have been honoured by outside bodies. Peter Campbell shared the 2010 Future Leaders in Cancer Research prize from Cancer Research UK, Mike Stratton was awarded the Lila Gruber

Award for Achievements in Cancer Research, and Richard Durbin and Mike Stratton became members of the European Molecular Biology Organization (EMBO).

It has been a fruitful year for publications. We led or contributed to 290 research publications, of which 41 were in *Nature*, *Cell*, *Science*, *NEJM* or *Nature Genetics*. These included the first analyses from the 1000 Genomes Project, which aims to characterise genomic variation in many populations across the globe. There was also a suite of publications specifically exploring genomic copy number alterations in human populations, encompassing a range of scientific questions from the patterns of variation observed to their contributions to disease.

The advent of next-generation DNA sequencing unleashed a torrent of data from cancer genomes, leading to the publication of the first comprehensive catalogues of somatic mutations in individual cancers and providing insights into the underlying processes that generate these genomic changes. Similarly, these technologies allowed our scientists to



Mike Stratton became Director of the Wellcome Trust Sanger Institute in May 2010.

Wellcome Library, London

Mike Stratton joined the Institute in 2000 and became Deputy Director in 2007. He leads the pioneering Cancer Genome Project that has convincingly demonstrated that large-scale sequencing of human cancer genomes is both practicable and generates clinically valuable information. It is now part of a broader international collaboration, the International Cancer Genome Consortium. One of its first discoveries – mutation of *BRAF* in malignant melanoma – has led to the development of highly promising therapies for this deadly and currently untreatable disease.

As Director, Mike is committed to both developing basic genome sciences and fostering their applications for human health.

» A new Director

After a decade at the helm, Allan Bradley has handed on directorship of the Wellcome Trust Sanger Institute to Mike Stratton.

Allan Bradley was recruited from Baylor College of Medicine in the USA to lead the Sanger Institute in 2000, with the challenge of building on the founding directorship of Sir John Sulston. Under Allan's leadership, the Sanger Institute shifted its focus from reference genome sequencing to the use of high-throughput approaches to understand gene function in health and disease. As well as his work with Kymab, Allan retains a research group at the Institute and has the title of Director Emeritus.



Alongside guiding the development of the Kymouse™ model, Allan Bradley will continue his research at the Institute as Director Emeritus.

Wellcome Library, London

From bench to bedside: translating our research

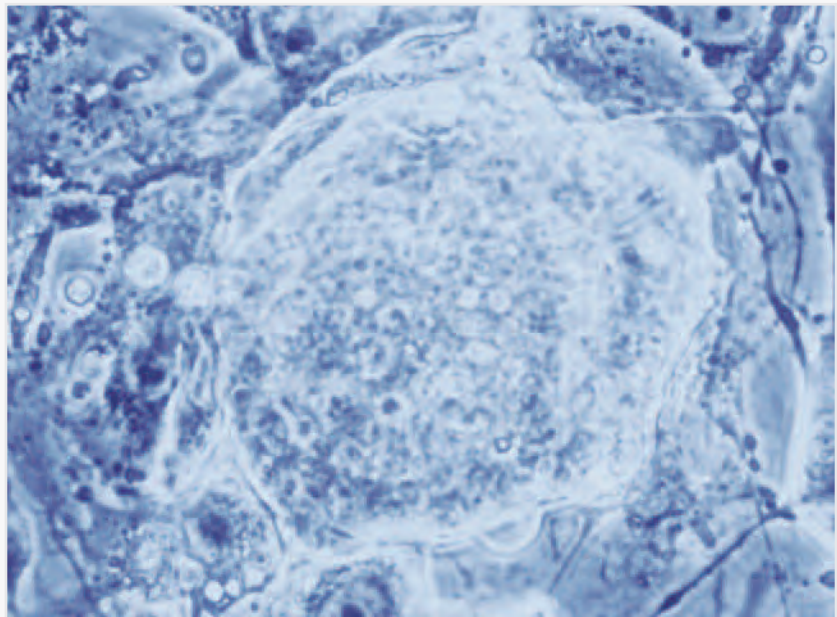
The spin-out of Kymab illustrates one strand of the Sanger Institute's translational strategy.

In 2010, Allan Bradley stepped down as Director of the Sanger Institute to guide the development of the spin-out biopharmaceutical company Kymab. The company received a £20 million injection of funds from the Wellcome Trust's Investment division.

The company's lead technology, based on Allan's pioneering work in chromosome engineering, is the Kymouse™, a strain of mice engineered to carry the entire human B-cell repertoire. In effect, it provides a 'library' of all possible human antibodies, and can act as a starting point for development of new and improved monoclonal antibody-based therapeutics.

The creation of Kymab is evidence of the Sanger Institute's strategic commitment to drive forward the translational potential of its research. While maintaining a strong focus on 'discovery' research, the Institute is also developing a translation strategy to ensure that new knowledge is used to improve human health.

The route of translation will depend on the nature of the advance. For example, the Institute makes sequence information and resources freely available to academic researchers to accelerate research across a wide range of areas. At the same time, initiatives such as the Deciphering



Embryonic stem cell colony. The Institute's flexible translation strategy enables its pioneering work – such as Allan Bradley's Kymouse™ – to be taken forward in the most appropriate way to deliver benefits in human health and research.

Genome Research Limited

Developmental Disorders (DDD) project, being taken forward with support from the Health Innovation Challenge Fund, illustrate the potential for clinical application through non-commercial routes.

In other instances significant additional development funding may be required to deliver a discovery's promise. In these situations, the Institute may seek licensing or partnering agreements with commercial bodies, or consider spin-outs such as Kymab. By applying a pragmatic approach to translation, the Institute is striving to ensure that the health benefits of its research are realised.

study variation between hundreds of individual pathogen genomes at a resolution that was previously unattainable, providing remarkable insights into the ways in which methicillin-resistant *Staphylococcus aureus* (MRSA) and *Salmonella* Typhi have spread around the world.

Our work in model organisms yielded many insights and generated important new directions. *Plasmodium berghei* infects mice and is a model for the human malaria parasite *Plasmodium falciparum*. Examination of engineered strains of *P. berghei* revealed genes that this organism can and cannot live without, the latter potentially being tractable drug targets. Remarkable work in mice has also discovered a way of generating a new class of immune cells that appears to recognise and kill cancer cells.

Our commitment to rapid and open sharing of genomic data is the driver of the Institute's world-leading digital resources. The Institute's websites play a major role in supporting the scientific community around the globe. During 2010, the genome browser Ensembl, which we maintain in collaboration with the European Bioinformatics Institute, had 50 000 visitors per week, with other databases such as COSMIC (cancer) and PFAM (protein families) maintaining leadership in their specific areas.

The internet also provides us with interesting opportunities for innovative ways of publishing and collaborating with other scientists. For example, we have started to publish accounts of our protein families in Wikipedia which others can directly edit, annotate and enrich employing their own expertise and insights.

Technology has continued to change at a blistering pace, nowhere more so than in the generation and storage of DNA sequence. The Institute was the first genome centre to make a major investment in a next-generation sequencing platform in 2007, resulting in a dramatic step change in sequencing capacity. However, in 2010, we already needed to implement a wholesale switchover to the latest version of these sequencing machines and started to explore future directions of these extraordinary technologies by receiving our first third-generation sequencer. The approximately 100-fold increase in data output from these platforms over the past three years has presented daunting challenges for our IT team and the infrastructure they maintain. During 2010, we topped 10 000 000 000 000 000 bytes (10 petabytes) of storage and 10 000 CPU of compute.



Genome Research Limited



Genome Research Limited

Translation has become a stronger theme in the Institute's scientific portfolio, testified to by the funding awards for the Deciphering Developmental Disorders and cancer burden studies and the formation of the Institute's first spinout, Kymab. We hosted our first Industry Open Day to showcase Sanger Institute research to potential translation partners and users of our data and resources. The event was well received and found thought provoking by the many participants from the Pharmaceutical and Biotechnology sectors.

The Institute is committed to engaging in the public debate surrounding genetic research. Our outreach team has developed an extraordinarily diverse set of activities that provide on-site and web-based resources for the public, in particular schoolchildren, to explore genomic science and the wider issues that surround it. Approximately 1500 people visited the Institute this year to sample at first hand the atmosphere of a genome centre, while thousands more logged on via the internet to investigate the implications of direct-to-consumer genetic testing through the YourGenome YouTube channel.

The diversity of the Institute continues to be powered by its science. The 2010 World Cup finals prompted us to investigate how many of the finalists were represented here. It transpired that nationals from 27 of the 32 competing countries work in our organisation, a source of much pride in our cosmopolitan atmosphere and the frontier-free nature of science.

The year was signed off with a revival of the traditional Sanger Institute pantomime, a suitably raucous and disrespectful take on *A Christmas Carol* by Charles Dickens. No-one was spared.

Mike Stratton, Director
February 2011

The Institute will use the power of next- and third-generation sequencing to investigate how genetic variation affects human and pathogen biology, the risk of disease and disease processes.



Dave Syer, Wellcome Trust

The next five years

The latest five-yearly review of the Sanger Institute, completed in 2010, set the stage for an updated strategy to guide the Institute through to 2016.

The Institute operates on a five-yearly cycle of funding from the Wellcome Trust. Following a highly successful five years, the 2010 review led to the approval of £390 million funding for its 2011–16 strategy.

The main theme of the Institute's research over the next five years will be genome variation, whether naturally occurring or artificially engineered, and its consequences in health and disease. We will be exploring this theme in six key areas: human genetics, cancer, mouse and zebrafish genetics, pathogen genomics, malaria and bioinformatics.

The Institute will continue to carry out its science at large scale, with its output based on our three major data production and analysis platforms: the DNA pipelines, the data centre and the mouse and zebrafish facility. These are among the largest in the world and critically underpin our aim to provide biological insights through generating tens of thousands of genome sequences from healthy human beings and those with disease, from cancers, from diverse infectious pathogens and to explore the implications of these through informatics and engineered experimental models.

We will investigate a broad spectrum of diseases both common and rare, ranging from disorders of development, to metabolic abnormalities and cancer, to autoimmune and infectious diseases. We will be assiduously pursuing our

global perspective on human health by developing further partnerships with researchers in Africa and Asia.

While maintaining a focus on discovery and basic science, the Institute will also direct its efforts towards realising health benefits. We will assess rigorously the potential for application of our research findings and will use a suite of mechanisms to foster practical applications.

Genomes are becoming part of mainstream cultural discourse. A central component of our future plans is therefore to engage with a wide range of publics, to kindle interest and develop understanding of the potential and challenges of genome sequences in the fast-moving 21st century.



Dave Syer, Wellcome Trust

The Institute will continue to be a leading provider of engineered mice, zebrafish and embryonic stem cells to the research community.