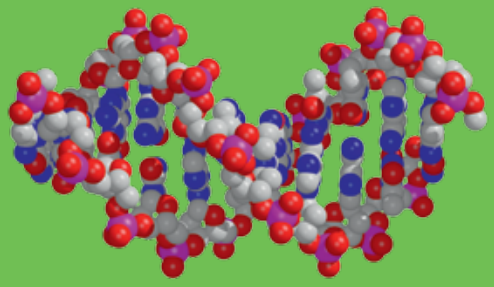
Portfolio Review

Human Genetics 1990–2009

June 2010



wellcome trust

Acknowledgements

The Wellcome Trust would like to thank the many people who generously gave up their time to participate in this review.

The project was led by Liz Allen, Michael Dunn and Claire Vaughan. Key input and support was provided by Dave Carr, Kevin Dolby, Audrey Duncanson, Katherine Littler, Suzi Morris, Annie Sanderson and Jo Scott (landscaping analysis), and Lois Reynolds and Tilli Tansey (Wellcome Trust Expert Group). We also would like to thank David Lynn for his ongoing support to the review.

The views expressed in this report are those of the Wellcome Trust project team – drawing on the evidence compiled during the review. We are indebted to the independent Expert Group, who were pivotal in providing the assessments of the Wellcome Trust's role in supporting human genetics and have informed 'our' speculations for the future. Finally, we would like to thank Professor Francis Collins, who provided valuable input to the development of the timelines.

Contents

Acknowledgements 2

Overview and key findings 4

Landmarks in human genetics 6

1. Introduction and background 8

2. Human genetics research: the global research landscape 9
2.1 Human genetics publication output: 1989–2008 10

3. Looking back: the Wellcome Trust and human genetics 14
3.1 Building research capacity and infrastructure 14
3.1.1 Wellcome Trust Sanger Institute (WTSI) 15
3.1.2 Wellcome Trust Centre for Human Genetics 15
3.1.3 Collaborations, consortia and partnerships 16
3.1.4 Research resources and data 16
3.2 Advancing knowledge and making discoveries 17
3.3 Advancing knowledge and making discoveries: within the field of human genetics 18
3.4 Advancing knowledge and making discoveries: beyond the field of human genetics – 'ripple' effects 19

Case studies 22

4. Looking forward: speculations on the future of human genetics 32
4.1 A need for underpinning research into basic biological mechanisms and genetics 32
4.2 A need for improved phenotypic definition and understanding among populations 34
4.3 A need to secure clinical involvement in human genetics research 34
4.4 The need for high-quality epidemiology and well-powered cohorts and genome-wide association studies 35
4.5 The need to support data sharing and open access to data 36
4.6 The opportunities and threats of personalised genomic medicine 37
4.7 Evolving partnerships with low-income countries 37

Annex A: Methodology 38

Landscape analysis 39

Narrative case studies 41

Wellcome Trust Expert Group on Human Genetics 41

Annex B: Wellcome Trust funding for human genetics 43

Annex C: Wellcome Trust bibliometric analysis for human genetics: 1989–2008 69

Annex D: Human genetics research timeline 73

Overview and key findings

1. Between 1990 and 2009, the Wellcome Trust committed £740m[1] to human-genetics-focused research, accounting for around 10% of its total funding commitment over this time. Over this period the Wellcome Trust has also been an active player in the formulation of international research policy and strategy in the field of human genetics. The drive to maximise the health benefits of research into the human genome remains a core component of the Wellcome Trust's funding strategy today.

2. In 2009, the Wellcome Trust conducted a major retrospective of the key breakthroughs in the field of human genetics over the past 20 years, attempting to identify its role within this. As in any review of the outcomes of research, an intrinsic challenge for a funder – and particularly one supporting basic and fundamental research – is to understand its role among the plethora of influences and actors that are involved in shaping and delivering research. Indeed, attempting to attribute breakthroughs and landmarks in the field solely to an event, researcher and/or funder is to take an improbable view of the way and the timeframes in which science and knowledge progress. To counter this, with the support of a group of independent subject experts, we have attempted to identify where the Wellcome Trust is thought to have played a significant and influential part in the field, but we by no means claim to have been the only influence.

3. We also wanted to use this review in a formative way, to inform future funding strategy; this review therefore brings expert reflections on past developments and critical breakthroughs together with views on remaining challenges and future opportunities for research in human genetics – both for the Wellcome Trust and for all those involved in this rapidly evolving field.

4. In doing this review, we have found that it is not always about how much money is invested; it is about being bold, flexible and agile enough to recognise and act upon the requirements of a particular scientific area at the most opportune points in time and supporting this in the most appropriate ways. It is about working with excellent researchers and, where value can be added, with other partners, to explore new areas and build critical mass.

5. The Wellcome Trust has invested substantially in the field of human genetics over the past 20 years, from its contribution to the Human Genome Project through the Wellcome Trust Sanger Institute to its role in major research consortia and collaborations, such as the Wellcome Trust Case Control Consortium. Through its support the Wellcome Trust has contributed to some of the most important human genetic advances and discoveries (see the **'Landmarks in human genetics' Timeline**). In particular, there are three areas where the Wellcome Trust is thought to have made a significant impact on the field of human genetics:

   - Through its role in providing stable and sustained funding to build research capacity and infrastructure to support human genetics and genomics.
   - Through its researchers, several key advances in knowledge and discoveries in human genetics have been made, which are likely to underpin future research and broader impacts.
   - Through its influence in forging partnerships and supporting strategies and policies that have helped to shape the direction and openness of human genetics.

1 Including funding to the Wellcome Trust Sanger Institute.

6. Over the past 20 years much of the research leading to perhaps the most significant breakthroughs in the area of human genetics have resulted from 'big', technology-based, collaborative efforts. Now that the human genome has been deciphered, the paradigms within human genetics are shifting; the challenge is to harness and act upon this new wealth of knowledge and understanding to bring about real impact on the health and wellbeing of populations – and support the field in ways best suited to deliver these impacts.

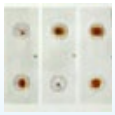
7. With the help of our experts, looking towards the future face of human genetics, we have identified several challenges on the road ahead and a need for further research and development. An overview of these challenges is presented below:

- The need for further underpinning research ('wet lab') into basic biological mechanisms and genetics, including research into monogenic disease and the genetics of infectious disease.
- A need for improved phenotypic definition and understanding among populations, including a requirement to undertake more human genetic studies on non-medical traits. Our experts felt that it would be valuable to include quantitative phenotyping elements in new and existing cohort studies such as the 2012 Birth Cohort Study. In addition, support for epigenetics research is thought to be key to enable further insight into phenotype and genotype studies.
- Our experts emphasised the need for funders to support high-quality epidemiology; well-powered cohorts and genome-wide association studies (GWAS) are required to generate robust associations between genetic factors and health outcomes and to secure adequate statistical power on rare variants. There may also be opportunities to develop genetic components to existing cohorts and longitudinal studies. Funding for longitudinal-based research needs to be sustained over the long term, however, and stability of funding is essential.
- It may be timely to consider how funders and researchers might engage with private genomics-based companies – including those not necessarily focusing on primary research (e.g. 23andMe, https://www.23andme.com/). Many private companies offering services directly to the public have a great wealth of genomic data at their disposal. If consent and ethical considerations can be accommodated, such companies could offer much to research.
- A need to ensure and facilitate the involvement of clinicians in human-genetics-related research is thought to be an important strategic goal to help assure its benefits. This could also include pathologists; our expert group felt that the establishment of a central pathology laboratory in the UK would offer many benefits to human genetics research.
- The need to guarantee global data sharing and open access to genetic and genomic data remains an important principle. To enable this, it is critical that key funders work together to develop sound governance frameworks to ensure that public and researcher confidence is maintained.
- The need to harness the opportunities to develop strategic partnerships with new and emerging science 'markets' across the world, both to develop fruitful research collaborations and to support the transfer of simple DNA-based technologies, for example, to assist in the management of communicable and non-communicable disease.
- In all of this, it remains crucial that we, as funders, work with researchers to ensure that the ethical and social implications of human genetics research are considered throughout study design and implementation.

Landmarks in human genetics

1866
Mendel publishes his work on inheritance

1900
Human blood groups described



1915
Hunt Morgan describes linkage and recombination

1951
First autosomal linkage

1952
Development of protein sequencing

1956
Genetic mutation underlying sickle-cell anaemia is identified

1959
Down's syndrome characterised

1860 1900 1910 1920 1950 1960

1859
Charles Darwin publishes *On the Origin of Species*

1869
DNA is discovered

1911
ABO blood group inheritance discovered

1924
Blood group inheritance characterised

1953
The structure of DNA is described

1956
Human chromosome number is determined

1990
Avon Longitudinal Study of Parents and Children (ALSPAC) inception

1992
Sanger Centre established

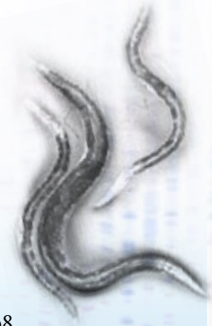
Microsatellite map of human genome

1995
Wellcome Trust agrees to fund one-sixth of human genome sequencing under the HGP

European Bioinformatics Institute opens (at the Wellcome Trust Genome Campus)

First genome of a free-living organism sequenced (*Haemophilus influenzae*)

Patrick Brown demonstrates DNA chips



1998
Nematode worm sequence published

Public Human Genome Project effort accelerated with additional Wellcome funding

2000
Working dr
genome an
by HGP and

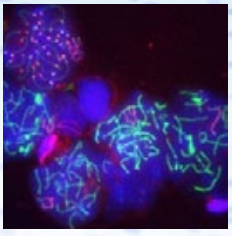
1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 20

1990
Human Genome Project begins with initial five-year plan

Discovery of SRY gene for sex determination

1993
MicroRNAs (miRNA) first described

1994
Discovery of *BRCA1* mutation (*BRCA2* mutation discovered 1995)

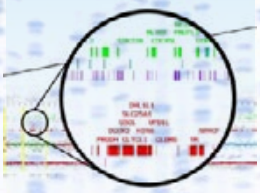


1996
First sequence of large complex organism, *Saccharomyces cerevisiae* (brewer's yeast)

Bermuda meeting of human genome scientists agree open access data-sharing principles

NIH begins funding for human genome sequencing efforts under the HGP

1999
First complete human chromosome sequence (no. 22)



Portraits
1866: Gregor Johann Mendel.
1869: Johann Friedrich Miescher.
1952: Fred Sanger.

KEY
Scientific advance
Funding development
Policy development

1966
First edition of *Mendelian Inheritance in Man*

1968
First gene mapped to an autosome

1970

1972
First recombinant DNA molecules produced

Use of somatic cell hybridisation mapping

1973
First Human Gene Mapping workshop

1974
Identification of the first human gene deletion as a cause of disease in α-thalassaemia

1975
Southern blotting developed

1977
'Dideoxy' DNA sequencing; first DNA genome is decoded

1978
The first DNA prenatal diagnosis of genetic disease (sickle-cell anaemia)

1980

1981
Human mitochondrial DNA sequence completed

1982
GenBank database formed

1983
Huntington's disease: first disease gene mapped

Polymerase chain reaction developed

1984
DNA fingerprinting developed

1985
First meeting on human genome sequencing at UCSC

Genomic imprinting described

1985
Physical map of nematode developed

1986
First positional cloning of disease gene (chronic granulomatous disease)

Automated DNA sequencing developed

1986
Dulbecco proposes the Human Genome Project (HGP)

Santa Fe workshop held to assess the feasibility of the HGP mapping

1988
Office of Human Genome Research created in NIH (later becomes NHGRI)

NIH and Department of Energy coordinating memorandum

1988
US National Center for Biotechnology Information established

Reports by the Office of Technology Assessment (OTA) and National Research Council (NRC) recommend establishing the HGP

1989
Cystic fibrosis gene identified

Human Genome Organisation (HUGO) formed

1990

aft of human
nounced
l Celera

2002
Draft sequence of the mouse genome published

Genome sequence of *Plasmodium falciparum* completed

International HapMap Project launched

2004
Chicken and rat genomes published

2006
Copy number variation of human genome described

The NCBI database of genotypes and phenotypes (dbGaP) is established

Genetic Association Information Network is launched in the USA

2008
1000 Genomes Project launched

International Cancer Genome Consortium launched

Implementation of the NIH GWAS policy

2010
'UK10K'- new project to identify the links between both rare and common variation related to human disease

00 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

2001
Draft sequences of human genome published

The SNP Consortium Ltd. maps genome variation

2003
'Gold standard' human genome sequence is completed; Wellcome Trust contributed one-third of the sequence

Structural Genomics Consortium launched

ENCODE project launched by NHGRI

2005
Publication of first draft of human haplotype map-genotyping of 1.1 million single nucleotide polymorphisms (SNPs) across four populations

Chimp genome published

Wellcome Trust Case Control Consortium established (WTCCC)

2007
Publication of largest genome-wide association study (GWAS) of common diseases (WTCCC)

ENCODE publication: new insights into genome function

Human Induced pluripotent stem cells produced

Launch of DeCODEme and 23andMe

2009
23andMe launches collaborative Parkinson's disease research initiative

Identification of two complete cancer genomes – malignant melanoma and small cell lung cancer

1. Introduction and background

8. As part of moves to strengthen the Wellcome Trust's evaluation activity, in 2008 the Assessment and Evaluation team developed an approach to review the impact of its funding at a subject, portfolio level. Conducted during the second half of 2009, this report describes the first portfolio review, focusing on the development of human genetics over the past 20 years and attempting to identify the role of the Wellcome Trust within this landscape.

9. We know that the Wellcome Trust has provided substantial funding for genetics – and, specifically, human genetics – research over its history. It has also been an active player in the formulation of international research policy and strategy in a field that has evolved radically in recent years. Support for genetics remains a cornerstone of the Wellcome Trust's funding strategy today.

10. Our portfolio review aims to be both reflective and prospective, with three specific aims:

    - to identify the key landmarks in and influences on the human genetics research landscape over the past two decades (1990–2009)
    - to consider the key features of the Wellcome Trust's impact on this human genetics research landscape
    - to speculate on the future direction of human genetics and consider where there may be opportunities for Wellcome Trust strategy and funding.

11. To deliver on these aims, we undertook three specific streams of work (see **Annex A** for detail):

    - Landscape analysis – funding and bibliometric landscape analysis
    - Narrative case studies
    - An Expert Group – convened to provide an independent view of key landmarks in the development of human genetics and the role of the Wellcome Trust and to speculate on the future.

12. By using this combination of complementary methods, we hope that this review will be helpful to the Wellcome Trust and other funders of human genetics, both by highlighting areas where new and continued research is required and in guiding the selection of mechanisms and, potentially, policies to support research.

2. Human genetics research: the global research landscape

13. The pace of change in human genetics research over recent decades has been phenomenal; major scientific discoveries have proceeded in tandem with breakthroughs in the development of enabling technologies (see **Timeline**). Together, these advances are transforming our understanding of how genes underpin biological processes in health and disease and have the potential to generate major health benefits in the coming decades.

14. In terms of financial investment in human genetics research over the past two decades, the USA has led the rest of the world by a significant margin; a survey of public and charitable funding for genomics research published in 2008 estimated that between 2003 and 2006, the USA provided just over one-third of the funding identified worldwide.[2] Within the USA, the National Institutes of Health (NIH) is the primary source of public funding for genetics research. The National Human Genome Research Institute of the NIH (NHGRI) is dedicated to human genetics and genomics research and had a budget of just over $500m per annum in 2009. Several other NIH institutes, such as the National Cancer Institute, also make substantial investments in genetics research, and there have also been a range of major trans-NIH initiatives – including, for example, an initiative on epigenomics as part of the recent Roadmap programmes.

15. In the same survey of global genomics research funding, the UK ranked second to the USA in terms of genomics research funding, accounting for 12% of funding over the same period. UK scientists have been at the fore of the field for many years.

16. Within the UK, the Medical Research Council (MRC) has played a key part in supporting outstanding scientists and developing major centres of excellence. Many of the pioneers of genetics research in the UK have been based at MRC units when they made key discoveries – including James Watson and Frances Crick, Frederick Sanger, Alec Jeffreys, Sydney Brenner, and John Sulston (see **Annex D**). Genetic research remains a key strategic priority for the MRC.

17. The MRC's support for human genetics research is supplemented by considerable support from UK medical research charities. The Wellcome Trust is the largest charitable funder, with a total expenditure of £720m in 2008/09 with major funding contributions to human genetics research in the UK over the past 20–30 years. Other UK medical research charities have also provided significant funding to genetic studies in specific disease areas. In particular, Cancer Research UK, formed in 2002 following a merger between the Imperial Cancer Research Fund and the Cancer Research Campaign, currently spends over £300m each year on research into cancer, and its funding has contributed to some significant breakthroughs in cancer genetics.

18. National funding agencies in other European countries, including Germany and France, provide significant funding for genetics and genomics research; major funding for large-scale collaborative initiatives in genomics has been provided by the European Commission via the Framework programmes.

19. Beyond Europe, Japan and other countries (including Canada and Australia) have developed significant genomic-based funding initiatives over the past decade. It is in China, however, where the growth in genetic and genomic-based research is perhaps most marked. Since becoming a partner in the Human Genome Project during the 1990s, China has ramped up investment in genome technology dramatically, and the Beijing Genomics Institute in particular has become one of the world's largest genomics research facilities.

20. In addition to the rapid pace of change, the second characteristic of human genetics research over the past few decades is its 'big', international nature; many of the key developments in human genetics have arisen through international collaboration and consortia. Such collaboration and partnership between funders and across research communities have been critical in overcoming the immense technological hurdles and in coordinating national programmes.

21. The Human Genome Project, beginning in 1990, paved the way for a new era of large-scale biology and formed a template for international partnerships to address subsequent challenges. The decision by the MRC and the Wellcome Trust to establish the Sanger Institute in 1992 enabled the UK to have a leading role in the Human Genome Project. Through continued Wellcome Trust support, the Sanger Institute has continued to have a leading role in international consortia.

22. These developments in the characteristics of the field of human genetics are reflected clearly in the patterns of publication outputs – which remain a key indication of research activity and progression – and are described in the next section.

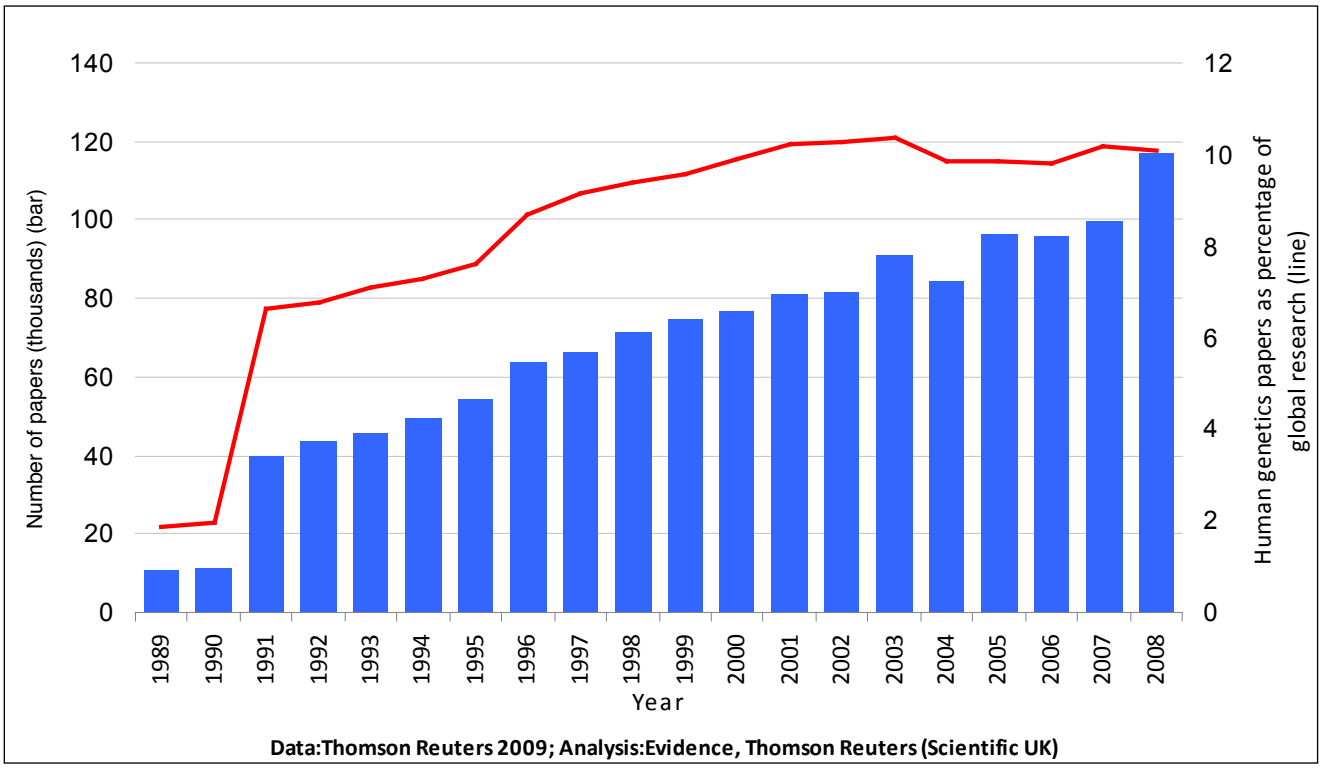
2 Pohlhaus, Cook-Deegan. Genomics research: world survey of public funding. BMC Genomics 2008;9:472

2. Human genetics research: the global research landscape

2.1 Human genetics publication output: 1989–2008

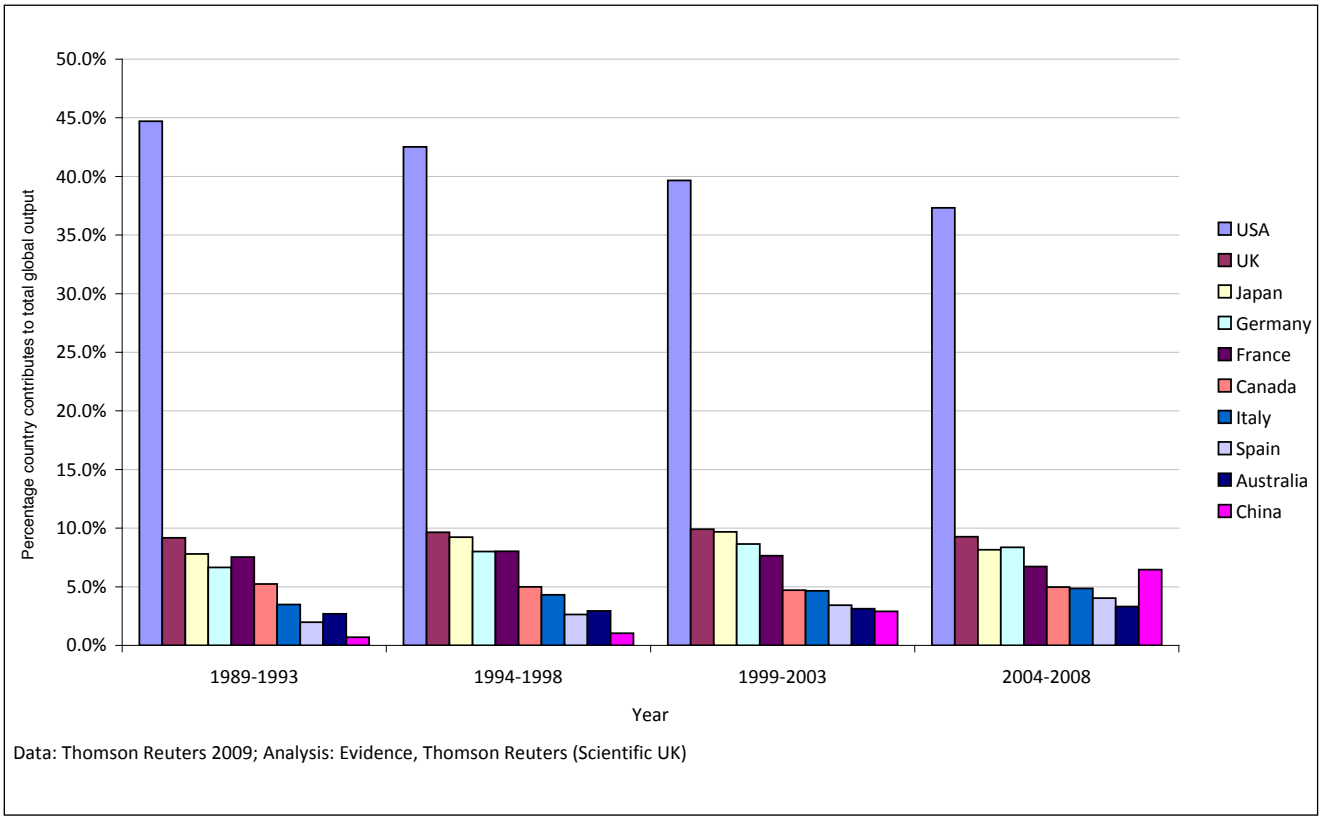
23. Based on publication output, the face of human genetics has changed somewhat over the past two decades. As a research field there has been steady growth, with major increases particularly at the beginning of the 1990s as work on human genome sequencing began in earnest; in 1990 human genetic research represented about 1.9% of all global research publications, but **by 2008 this proportion had increased to 10.1% of global research output** (see **Figure 1**).

Figure 1 Number of human genetics papers and as percentage of all research papers, 1989–2008



24. Over the past two decades human-genetics-focused research has been produced by a core set of countries, the composition of which has not changed over this period. These countries contribute to over 90% of the world's human-genetics-focused publication output (see **Figure 2**). Throughout this period, the UK has retained its position, currently linked to around **10%** of global output (see **Figure 2**), second only to the USA.

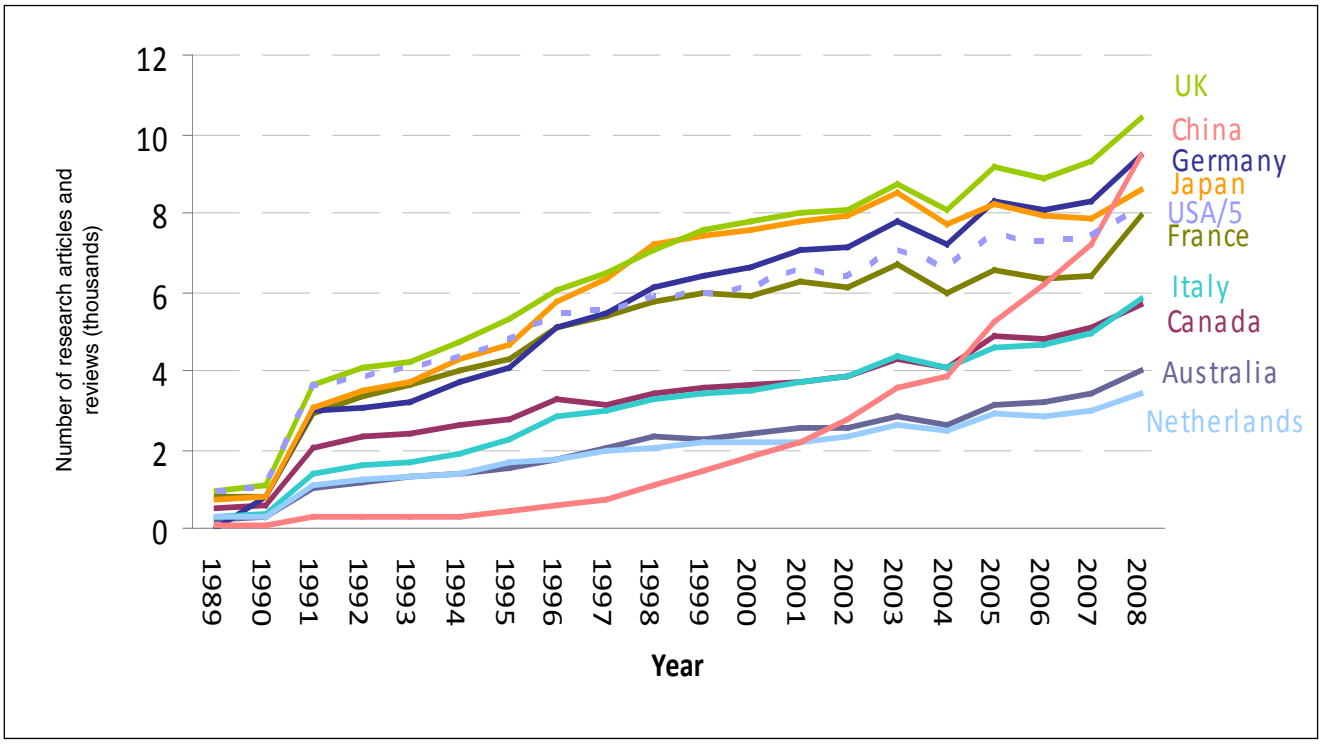
Figure 2 Human genetics papers linked with 'top ten most productive' countries, 1989–2008 (five-year periods)



25. However, although the absolute volume of human-genetics-focused papers associated with US institutions has risen throughout the period, the proportion of papers to which the USA is linked has fallen over time as other countries increase their production of human-genetics-focused research (see **Figure 2**). Countries such as Germany and China are linked to increasing proportions of papers over time and particularly in the most recent decade (see **Figure 3** and **Table 1**, **Annex C**). China has shown the most marked growth, with a growth of 171% during the past decade (see **Table 1**, **Annex C**). In terms of world publication outputs within the field of human genetics, China is currently (2008) ranked 6th and has grown from being associated with 0.7% of the world total in 1989 to 6.5% in 2008, being linked to one hundred times more publications at the end of the period than at the start. Elsewhere, countries showing fast growth in terms of human genetic publication production include South Korea, India and Brazil (see **Figure 3** and **Table 1**, **Annex C**).

2. Human genetics research: the global research landscape

Figure 3 Time trend in human genetics output – top ten countries (by volume), 1989–2008

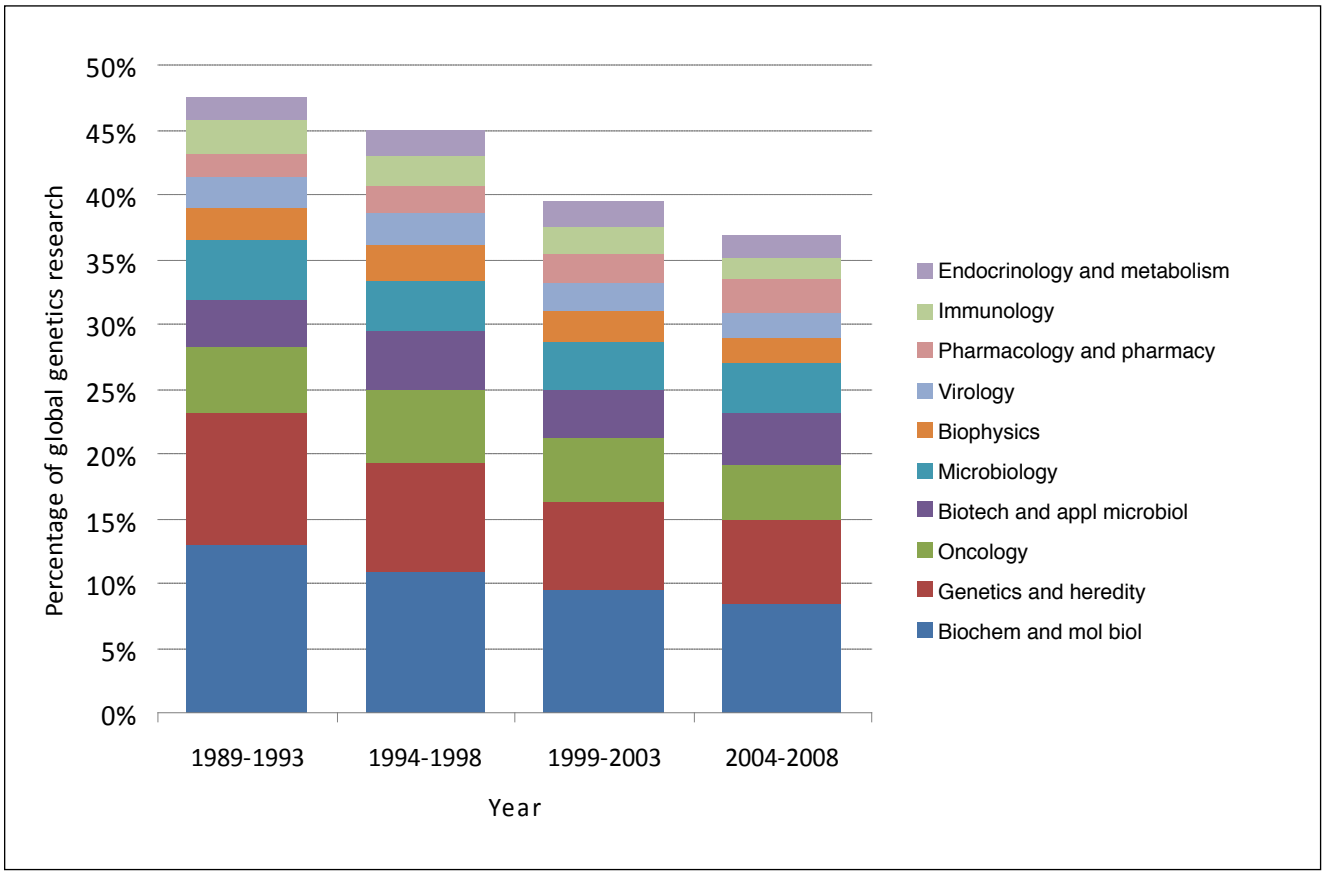


Data: Thomson Reuters 2009; Analysis: Evidence, Thomson Reuters (Scientific UK)

26. Analysis of the subject focus of human-genetics-related publication output over time shows that while much of the output retains a basic, fundamental research focus – being classified as 'biochemistry', 'molecular biology' and 'general genetics' – there has been considerable diversification over the past two decades. As the science and technology in genetics-related research has gathered pace, so the possibilities to offer insights into other disciplines and research areas have expanded. As a result, particularly in the most recent five-year period, there has been a massive growth in the volume of human-genetics-related papers appearing in journals focusing on neuroscience, clinical neurology and infectious diseases, showing growths of 649%, 510% and 477%, respectively, between 2004 and 2008.

27. The past two decades have also seen increased growth in genetics-related research in virology (147%), microbiology (181%) and infectious diseases (477%) – especially in the 1989–1998 time period (see **Figure 4** and **Table 2, Annex C**). However, although there has been an absolute increase in the volume of human genetics papers appearing in infectious-disease-related journals, the growth of human genetics papers in infectious-disease-related journals has declined markedly over the past decade – from 266% (between 1989–1993 and 1994–1998) to 9% (between 1989–1993 and 2004–2008). This slow progression in infectious-disease-focused human genetics was highlighted by the Wellcome Trust Expert Group: the relative lack of pathogen sequencing and need for more phenotyping of sample collections in populations were thought to be contributing factors to the relatively slow development of human genetics in this area over the past decade.

Figure 4 'Human genetic' papers published in five-year periods by subject category (1989–2008)



Data: Thomson Reuters 2009; Analysis:Evidence,Thomson Reuters (Scientific UK)

28. To take us beyond the volume of output, citation analysis has been conducted to help identify the origins of high-quality, highly cited research.[3] Citation data for 20 years (1989–2008) was used to determine the origins and affiliation of the most highly cited human genetics papers worldwide.

29. Over the whole period, all but eight of the 30 institutions linked to the most highly cited papers are US institutions, with Harvard and the University of Texas in first and second position throughout the 20-year period. However, UK-based institutions are increasingly featured, particularly in the most recent five-year period (2004–2008).

30. In 2004–2008, five of the top 30 institutions are UK institutions: the Universities of Cambridge and Oxford appear in the top ten (rank 7 and 8, respectively); University College London, the Wellcome Trust Sanger Institute and Imperial College are at positions 20, 23 and 24, respectively (see **Table 3, Annex C**). In addition, when the number of 'highly cited' papers produced by each institution is calculated as a percentage of their total 'human genetic' output, the Wellcome Trust Sanger Institute not only outperforms all UK institutions – 13.4% of its total output is 'highly cited' – but also surpasses the top two international producers, Harvard University and the University of Texas (see **Table 4, Annex C**).

3 In this analysis, 'highly cited papers' are those papers with an average rebased impact of at least four (i.e. they have received at least four times as many citations, by the end of 2008, as the average paper published in that year in the same subject area).

3. Looking back: the Wellcome Trust and human genetics

31. Key developments in biomedical science in the mid- to late-20th century (see **Timeline**) offered much to the understanding of human genetics. By the end of the 1980s, there was much excitement across the research community about the promise that breakthroughs in human genetics could offer. At the beginning of the 1990s, the Wellcome Trust – as an independent, internationally respected and flexible funder with financial clout – was in an ideal position to help propel the field forward and shape the direction of the research agenda, acting as a broker and catalyst to the formation of researcher groupings who have delivered many critical breakthroughs in the field.

32. This portfolio review highlighted three broad areas where the Wellcome Trust has made a significant impact on the field of human genetics: through its support for building **research capacity and infrastructure** to support human genetics and genomics; through the delivery of key **advances in knowledge and discoveries** in the field; and through its **influence** in forging partnerships and supporting issues which have shaped the direction of human genetics – both within and outside the field. These are considered in turn.

3.1 Building research capacity and infrastructure

33. Since the early 1990s, the Wellcome Trust has worked to build research capacity and expertise, providing a significant contribution to the human genetics research base, particularly in the UK. The Expert Group described one of the Wellcome Trust's major achievements as providing sustained and substantial funding for human genetics in the UK, helping to support the training of a highly skilled research and technical community able to advance the science. This is reflected in the UK's prominence in the human genetics publication output over this period – a strength which appears to be growing over time (see **Section 2**).

34. Between the Wellcome Trust financial years 1989/90 and 2007/08, the Wellcome Trust committed around £740m to human genetics; £502m through its grant funding and an additional £238m through the Wellcome Trust Sanger Institute (WTSI).

35. Through its funding divisions, the Wellcome Trust has awarded 1172 grants[4] – mainly in responsive mode – to human-genetics-focused research, accounting for £502m (just over 9% of the Trust's funding commitment over this time). Of this, approximately two-fifths (27% by value; 428 grants; £132m) of human genetics grant funding has been careers-based, supporting 398 individual researchers doing human-genetics-based projects via personal support schemes (see **Annex B**, **Table 1**): studentships (£9m), early career fellowships (£18m), and Intermediate (£26m) and Senior/Principal Research Fellowships (£79m).

36. The larger proportion of funds (74% by value; 744 grants; £370m) has been allocated to research and project support – equipment, university awards, strategic awards, buildings (Joint Infrastructure Funding) – and project and programme grants (see **Annex B**, **Table 1**). Although funding for career support research grants has remained relatively stable over the reporting period, funding for research, infrastructure and equipment-based projects has increased significantly – almost three-fold in the past decade (2000s; see **Annex B**, **Figure 3**).

37. Over this 20-year period, as described, most (97% by value) Wellcome Trust funding for human genetics research has been allocated to UK institutions; only 3% of funding (67 grants; £13m) has been committed to non-UK-based human genetics research (see **Annex B**, **Table 1** and **Table 2**).

4 This analysis identified just under 10 600 Wellcome Trust grants with a major genetics focus funded across all its funding divisions over the period 1989/90 to 2007/08; following a manual exercise, 1172 of these were deemed to be human genetics focused and hence included in this analysis (see **Annex A** for detail).

3.1.1 Wellcome Trust Sanger Institute

38. In 1992, with initial contributions from the UK MRC,[5] the Wellcome Trust Sanger Institute (see **WTSI case study**) was established as an advanced facility for mapping, sequencing and decoding the human genome and the genomes of other organisms. Since its inception, the Wellcome Trust has committed over £740m to the Institute as a whole with approximately 32% (£238m) of this being spent on human genetics (see **Annex B, Figure 2** and **Table 3**).

39. The WTSI has facilitated a sustained and coordinated programme of genome and pathogen sequencing work, which has played a key part in propelling human genetics forward. Sustained support for the development and maintenance of major resources at the WTSI, including the provision of DNA and sequencing technology tools, have facilitated the development of human genetics and were crucial in ensuring that the WTSI was able to have a major role in sequencing the human genome.

40. Through the WTSI, the Wellcome Trust drove the UK's contribution to the Human Genome Project (HGP), sequencing one-third of the human genome, the largest single contribution from any single centre. In addition, through the WTSI's leadership, the Wellcome Trust is credited with playing a major part in ensuring that genomic sequence data were made freely and publicly available, without restriction.

41. The WTSI has subsequently become a world leader for human genetics – in terms of technology development, data storage, scientific breakthroughs and its pivotal role in major human genetics collaborations and consortia. The decision to co-locate the European Bioinformatics Institute[6] (EBI) at the WTSI in 1996 has also been important in building capacity to support the data infrastructure required to underpin human genetics research. The EBI, part of the European Molecular Biology Laboratory (EMBL), is located on the Wellcome Trust Genome Campus in Hinxton. The EBI grew out of the EMBL's work in providing public biological databases to the research community.

42. This co-location is proving valuable to the field; in 2008, the WTSI and the EMBL-EBI were rated as the top two most influential research institutions in the UK, according to results published in *Science Watch*,[7] which ranked scientific publications from UK institutions from 2003 to 2007 according to their impact among researchers.

43. The WTSI, together with the EMBL-EBI, developed the Ensembl web browser (see **Annex B, Table 10**), which allows access to human and other genome sequences. In addition to Ensembl, the EBI hosts some of the worlds most important resources for biological research, including EMBL-Bank, UniProt, the Macromolecular Structure Database, ArrayExpress, IntAct and Reactome.

3.1.2 Wellcome Trust Centre for Human Genetics

44. The Wellcome Trust Centre for Human Genetics (WTCHG), which was established in Oxford in 1993, is to date one of the largest recipients of Wellcome Trust funding for human genetics, receiving approximately £54m since its establishment. Researchers at the WTCHG are linked to many of the key breakthroughs (see **Annex B**, **Table 9** and **Table 11**) in the understanding of genetics-related disease and have been integral to the inception and development of several major human genetics-focused collaborations and consortia such as the Wellcome Trust Case Control Consortium 1 and 2 (WTCCC and WTCCC2) and the 1000 Genomes Project.

45. Recently the WTCHG has established CONVERGE (China Oxford and VCU Experimental Research on Genetic Epidemiology), which is one of the largest studies to date into the genetics of depression.

5 The UK MRC contributed approx £24m to the WTSI for the period up to the end of 2002. Source: WTSI.

6 http://www.ebi.ac.uk/

7 http://www.sciencewatch.com/

3. Looking back: the Wellcome Trust and human genetics

3.1.3 Collaborations, consortia and partnerships

"It wasn't just access to technology and bioinformatics, it was changing and forcing people out of their silos into human genetics collaborations, as the silos (except the diehards) realised they couldn't do it on their own, they had to do it in partnership with funding agencies and other silos…The Trust not only provided the bricks and fellows and considerable money and equipment, but they drove that agenda, the staff of the Trust and the Governors drove that."
Wellcome Trust Expert Group on Human Genetics, October 2009

46. The Wellcome Trust has also played a leading part in the conception, establishment and ongoing support of a range of **major international collaborations and consortia** – many involving WTSI and the WTCHG. Acting as a broker and catalyst, Wellcome's independence is thought to have been key in allowing it to transcend both university and government politics.

47. Through its collaborations, consortia and partnerships, the Wellcome Trust has been instrumental in building a critical mass of expertise in the field of human genetics by bringing together collaborating parties with different expertise and/or resources contributing in diverse ways to an important scientific question, most notably the Human Genome Project (HGP), through its collaboration with the US Department of Energy (DoE) and the NIH.

48. Through the WTSI, the Wellcome Trust was able to have a leading role in the 13-year-long Human Genome Project (HGP) by providing support to the WTSI, which took responsibility for sequencing one-third of the human genome – the largest single contribution from any one centre.

Other examples of major research collaborations and consortia in which the Wellcome Trust has played a key part are listed below and are detailed, along with their key accomplishments to date, in **Table 9** (see **Annex B**).

- SNP Consortium (TSC)
- International HapMap Project
- Structural Genomics Consortium (SGC)
- establishing the joint Juvenile Diabetes Research Foundation-Wellcome Trust Diabetes and Inflammation Laboratory (JDRF/DIL)
- Wellcome Trust Case Control Consortium (WTCCC)
- Mouse Genome Sequencing Consortium (MGSC)
- The 1000 Genomes Project
- Cancer Genome Project and the International Cancer Genome Consortium

3.1.4 Research resources and data

"Not only did the Wellcome Trust fund serious sample collection, and try to make sure it was good, it was shared and openly available to anyone that wanted the samples [and] so the Wellcome Trust invested very heavily in that…To use these samples they began the case control consortium idea and then a couple of years later the Wellcome Trust really cemented that."
Wellcome Trust Expert Group on Human Genetics, October 2009

49. The vast quantities of genomic data generated from major projects such as the WTCCC, HGP, HapMap and the 1000 Genomes Project have driven the need for improvements in sequencing technologies and sophisticated methods of data storage, and analysis tools to enable researchers to share and interpret the data being generated by sequencing projects.

50. Since its involvement in the HGP, Wellcome has played a key part in the creation, development and maintenance of major genomic analysis tools, research resources and databases, including (see **Annex B**, **Table 10** for details and key accomplishments to date):

- Ensembl Genome Browser
- HapMap
- The 1000 Genomes Project
- Encyclopaedia of DNA elements (ENCODE)
- DECIPHER
- CONVERGE
- UK10K

51. In addition, the Wellcome Trust has supported the development of several other population, cohort and family studies, which – although not originally set up with a genetics design – are now evolving to incorporate significant genetic elements. These include the Avon Longitudinal Study for Parents and Children (ALSPAC) and the UK Biobank (see **Annex B, Table 10**). Such studies offer promise for providing linkages between genotype and both medical and non-medical phenotype, a requirement highlighted by the Wellcome Trust Expert Group (see **Section 4**).

52. In addition, through the Wellcome Library for the History and Understanding of Medicine, Wellcome provides an international research resource for the history of medicine and medical and clinical genetics.

3.2 Advancing knowledge and making discoveries

"The important thing is that we discovered a whole new class of genes that really were just not known and the existence of the Wellcome Trust's investment in the sequence were critical."
Wellcome Trust Expert Group on Human Genetics, October 2009

"The small number of genes is very surprising; I think that is one of the most fascinating [findings of the HGP]. Everyone gets depressed over that but it is incredibly exciting, that and microRNAs."
Wellcome Trust Expert Group on Human Genetics, October 2009

53. Over the past two decades, the Wellcome Trust has supported some of the most important discoveries in the field, through its grant funding, through research consortia and partnerships (see **Annex B**, **Table 11** and **Timeline**), and perhaps most notably through the Human Genome Project.

54. In addition to deciphering the sequence itself, the HGP revealed a few surprises that are thought to be changing paradigms across human genetics, including the discovery of a much lower number of protein-encoding genes than previously assumed (a total of between 20 000 and 25 000 genes), the discovery of microRNAs, and the discovery of copy number variation.

55. The HGP and the delivery of other genome sequences to the research community sit alongside a range of other discoveries in human genetics. Virtually all of these are breakthroughs at the most basic, fundamental end of the biomedical spectrum – and at the start of any research 'pipeline' – but will hopefully provide the foundations for future research and subsequent applications. The specific landmarks in human genetics to which the Wellcome Trust is linked are listed below (see **Timeline**):

- Sequencing of the first complex organism with cells similar to those found in humans – *Saccharomyces cerevisiae* (baker's yeast).
- Sequencing of the first animal genome: *C. elegans* (nematode worm).
- Decoding the first complete sequence of a human chromosome (no 22).
- Deciphering the human genome – first draft in 2000, 'gold standard' in 2003. The WTSI was responsible for sequencing one-third of the code.
- Linked to the HGP, the discovery of the relatively low number of protein-encoding genes in the human genome.
- Discovery of microRNAs (miRNAs) and copy number variation (CNV). This has stimulated a shift back to basic genetic research, in particular gene regulation and gene function. It has also stimulated research into the emerging field of epigenetics.
- Deciphering the genomes of a range of human pathogens responsible for significant global human morbidity and mortality, including *Plasmodium faliciparum*, *Schistosomiasis mansoni*, *Leishmania* species (*L. major, L. braziliensis* and *L. infantum*) and *Clostridium difficile.*

3. Looking back: the Wellcome Trust and human genetics

- Through the Wellcome Centre for Human Genetics at Oxford:
  - identification of several SNPs, in three unlinked members of the *CADM* gene family, demonstrating association with tuberculosis
  - major contributor to the WTCCC. Identification of complex disease loci in human genome (via WTCCC)
  - identification of the first 20 loci for height
  - identification of the second obesity locus
  - identification of the second tier of six previously unknown type 2 diabetes loci
  - identification of the role of *Pcsk5* in cardiac development, body patterning and VACTERL syndrome
  - discoveries providing the first set of evidence that demonstrates that a substantial fraction of human recombination hotspots share a common mechanism
  - elucidation of the role of the *KIAA0319* dyslexia susceptibility gene.
- Through the ALSPAC, a range of associations, including playing a significant part in the identification of the obesity-associated *FTO* gene.
- Through the WTCCC:
  - identification of the first genetic link (*PTPN2* gene) between Crohn's disease and type 2 diabetes
  - identification of four chromosome regions containing genes that can predispose to type 1 diabetes[8]
  - identification of six new genetic variants that increase the likelihood of developing coronary artery disease
  - discovery of the first gene linked to obesity; the *FTO* gene linked to childhood and adult obesity (see ***FTO* case study**).
- Through the Cancer Genome Project, providing the first comprehensive analysis of two complete **cancer genomes – small-cell carcinoma of the lung and malignant melanoma** – identifying almost all of the mutations associated with the two cancers.

3.3 Advancing knowledge and making discoveries: within the field of human genetics

"A very important role of the Trust...the Trust is a charity and it is not or doesn't have to be devoted to wealth creation. This is a unique position to be in, in a world where governments have become businesses and are furiously competing with each other."
Wellcome Trust Expert Group on Human Genetics, October 2009

56. At the beginning of the 1990s, the Wellcome Trust's independence, reputation as a funder of excellent science and financial stability enabled it to gain access to other significant international funders to help facilitate the step change in genomics that was imminent with breakthroughs in knowledge and sequencing technology.

57. At this time the Wellcome Trust began discussions with the US DoE and the NIH, who had embarked on the journey to sequence the human genome. Developing a partnership with the US DoE and the NIH, the Wellcome Trust – and scientists at the WTSI – were able to:

i. play a key part in **accelerating the pace at which the human genome was sequenced**. In response to the formation of the private venture Celera, in 1998 the Wellcome Trust took the decision to increase its funding for the sequencing work at the WTSI helping the public consortia decipher the human genome first, to protect ready availability of the sequence data to researchers world-wide – it was not just pride and irritation and competitiveness, it was critically important in facilitating subsequent growth of science.

ii. ensure that the **human genome sequence was placed in the public domain** without restrictions, which was pivotal to the subsequent growth of the 'open access' movement and the more equitable access/sharing of research data across science and research. Wellcome Trust staff – and Michael Morgan is highlighted in particular – worked with the NIH to convene the Bermuda meeting in 1996, where genomic data sharing principles were first agreed and later publicly endorsed by US President Bill Clinton and British Prime Minister at the time, Tony Blair.

8 Todd JA et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 2007;39(7):857–64

"The Bermuda meeting was actually very important... we agreed here [Wellcome Trust], and in Sanger, that we needed the meeting because people were running off all over the place doing private bits of sequencing. We were clearly just going to waste an awful lot of money if we didn't get it together so we called the Bermuda meeting and managed to get some sort of agreement, it was absolutely amazing. We made the agreement about the human genome."
Wellcome Trust Expert Group on Human Genetics, October 2009

"The public HGP effort accelerated with additional Wellcome Trust funding...I believe that [Wellcome coming on board] was an absolutely crucial moment in the development of the way in which the genomes were made public...This would basically influence the battle in the US Congress, as to whether or not NIH was going to continue to get appropriation for sequencing which in turn would, if that had not happened, then almost certainly the public databases would have collapsed."
Wellcome Trust Expert Group on Human Genetics, October 2009

58. Over the past two decades the working relationship between the Wellcome Trust and the US NIH has been particularly productive in the area of human genetics; in addition to the HGP, Wellcome has established several jointly funded genetic initiatives with the NIH, including:

- Mouse Genome Sequencing Consortium
- The International HapMap Project
- The 1000 Genomes Project
- Pathogen Genome Sequencing – the Wellcome Trust co-funded the sequencing of the *Plasmodium falciparum* genome with the National Institute of Allergy and Infectious Diseases (NIAID). NIH have also co-funded the sequencing of the leishmania parasite and other pathogen genomics projects conducted at the WTSI
- Following the International Structural Genomics Meetings in 2000 at Hinxton and in 2001 in Virginia, the International Structural Genomics Consortium (ISGO) was formed to promote international co-operation in structural genomics research
- ENCODE

The WTSI has become a focal point for international human genetics research, providing a 'campus' for a critical mass of leading scientists who are major players in international human genetics research: Sir John Sulston was the first Director of the WTSI and was a key player in driving the HGP in the UK and Allan Bradley, now former Director of the Sanger Institute, was appointed in 2000 when John Sulston stepped down. He brought the Institute into a new era: studying the biology of genomes. Professor Mike Stratton, one of the world's leading cancer geneticists and joint head of the Cancer Genome Project, is the current director of the WTSI. Professor Leena Peltonen-Palotie (1952–2010) – considered one of the world's leading geneticists – took on the mantle of Head of Human Genetics at the WTSI until her death in March 2010. Also at the WTSI are Dr Richard Durbin, a world leader in genomics and bioinformatics who led the 1000 Genomes Project; Matt Hurles, who has been key in the continued evolution of CNV discovery work; and Inês Barroso, a world expert in the genetics of metabolic disorders. Researchers at the WTSI are among the most highly respected researchers in the field globally.

"The Wellcome Trust was open to forming partnerships ...and that was key."
Wellcome Trust Expert Group on Human Genetics, October 2009

59. In addition, the Wellcome Trust's openness to forging partnerships has been a key ingredient in its strategy to support research in human genetics. These partnerships have provided a mechanism for individuals, organisations or groups of scientists, committed to working collaboratively on 'big science' projects that no one partner could achieve alone, delivering a range of accomplishments.

3.4 Advancing knowledge and making discoveries: beyond the field of human genetics – 'ripple' effects

60. In addition to the direct influences that the Wellcome Trust has made on the development of human genetics, there are a range of other significant 'ripple effects' beyond the field that can be, in part, attributed to the Wellcome Trust and other funders involved in supporting human genetics over the past two decades.

61. First, the profile of the **open data release** research agenda was raised significantly following the race to sequence the human genome and adoption of the 'Bermuda principles' to place sequence assemblies of 1–2 kb or greater in the public domain, within 24 hours of their generation, without restriction. The Wellcome Trust and the other funders of the HGP helped to further the

3. Looking back: the Wellcome Trust and human genetics

data sharing agenda across the world.

"This question of data sharing and early release is a ripple that goes well beyond genetics. It hasn't quite got there but it is getting there, it is getting into all sorts of other fields."
Wellcome Trust Expert Group on Human Genetics, October 2009

62. Following the adoption of the Bermuda principles in relation to genomic sequence data, in 2003, in response to requests from the large-genome sequencing scientific community, the Wellcome Trust also sponsored a meeting in Fort Lauderdale in Florida, to consider pre-publication data release and other large-scale genome sequence data.

63. As described previously, the Wellcome Trust is largely credited with having taken the lead on open access policy in the UK. The Wellcome Trust was pivotal in the establishment of UK PubMed Central[9] (UKPMC) in 2006 and is currently involved in discussions to set up a European counterpart system. It has helped to set a new paradigm for the sharing of research information and data. This extends to all parts of the research process, from making research data available to making the analysis and conclusions available in the public domain through open access publishing.

64. A second 'ripple', which stems largely from developments in human genetics, is the **growth in the field of bioinformatics**. Bioinformatics has become essential to genome studies as unprecedented amounts of genomic and proteomic data have necessitated the development of new tools and techniques to permit effective analysis and interpretation. In the post-genomic era, huge volumes of data continue to be generated and the demand for new informatics resources continues.

"The development of bioinformatics as a sort of functional global tool and that didn't just happen…the Trust had a reasonable impact there."
Wellcome Trust Expert Group on Human Genetics, October 2009

65. A third ripple effect is the **development and growth of bioethics research**. The Expert Group ascribed a large part of the rapid expansion of the academic field of biomedical ethics since the early 1990s to the development of science around human genetics.

66. Advances in the field of human genetics, in particular at the height of the HGP, raised important ethical, legal and social implications about genomic research and its medical applications.

"One could actually argue that the whole of modern bioethics, that it was increased by several orders of magnitude in bulk because of human genetics."
Wellcome Trust Expert Group on Human Genetics, October 2009

67. In 1997, the Wellcome Trust established its Biomedical Ethics Research Programme, to build research capacity in this area. At about the same time, it became a key funder of the Nuffield Council on Bioethics.

68. A fourth 'ripple' is the **technical development of the field of forensic science**. Human genetics and associated technology has had a major impact on the field of forensic science. DNA fingerprinting or DNA profiling (typing), developed by Sir Alec Jeffreys in 1984 (see **Timeline**), has revolutionised the field of forensic science, becoming a key tool in forensic medicine. In addition, the development of new DNA and sequencing technologies, including the polymerase chain reaction (PCR) (see **Timeline**), have increased the validity and reliability of DNA evidence in criminal and forensic investigations. PCR technology took off in the UK in 1995, enabling another huge leap in forensics – the development of the National DNA Database.

69. DNA evidence collected at crime scenes has enhanced our ability to identify both the perpetrators and the victims of crimes, thus increasing the appeal and subsequent efforts by criminal justice agencies and ministries of justice to develop national DNA databases. However, there are a range of challenges associated with the development of DNA databases, not least the need for strict regulation on their use, access to information and how to store and hold data. The UK's National DNA Database is currently the largest forensic DNA database in the world.

70. Forensic scientists are also using DNA evidence to identify human remains, determine paternity, and study human populations, diversity and medical disease. However, the emergence of forensic genetics also brings with it new ethical and legal challenges, particularly around the consent of individuals whose DNA is gathered.

9 http://ukpmc.ac.uk/

71. A fifth 'ripple' extends to **diagnostic and predictive medicine in healthcare.** Today, genetic testing is commonly used in diagnostic medicine and public health care, for rare and common disorders. Many pregnant women undergo routine genetic tests such as a prenatal screen or prenatal test and a growing number of individuals are now opting to have their DNA profiled. Genetic testing for a range of conditions and susceptibilities is the single largest application of genomic knowledge to date.

72. Again, the increasing prevalence of, and demand for, genetic testing in healthcare brings with it the need to consider the ethical, social, economic and practical implications. Genetic testing raises considerable challenges, to which regulatory frameworks must adapt in a timely fashion. It is also becoming increasingly important that health care systems set in place the necessary frameworks to ensure the effective uptake of clinically useful technologies as they develop – such as specialist genetic counselling services – to ensure that public confidence is fostered and that the public are aware of and engage with the implications of these advances.

"Pre-natal diagnosis has probably been much more practically effective but less widely practised. The sort of paradigm that was built up in clinical genetics of developing information and trying to put it back to people in a relatively value free way and letting people make their own decisions and so forth has become everyday parlance in the NHS...that was absolutely not that way when this started 25 years ago where doctors generally told patients what to do and didn't spend a lot of time thinking about it."
Wellcome Trust Expert Group on Human Genetics, October 2009

73. A sixth ripple effect is the **ability to consider race, ethnicity and origin in genetic terms**. Elucidation of human genetic variation across and within populations is presenting the opportunity to consider whether genetic factors might be correlated with racial or ethnic groups defined by their geographic location or social or cultural characteristics. Much work is underway to clarify the relationship between race, ethnicity and disease. The International Human HapMap Project[10] opened a window into the study of human genetic variation and since the completion of the HGP there has been a heightened interest in genomic analysis across populations.

"One...potential ripple is thinking about the people debating race and ethnic groups...we now can define someone's origin by just doing a chip and I think that has brought an enormous amount of clarity. There have been hundreds of papers written about how exquisitely you can define someone's genetic origin – obviously not citizenship, but their genetic origin. It is an enormously powerful and very productive way of thinking about different populations, instead of more emotive ways of thinking about it. When you look at the African genomes, for human evolution, it is a very graphic illustration of what is going on. It is amazing to see African genomes and the diversity."
Wellcome Trust Expert Group on Human Genetics, October 2009

74. The 1000 Genomes Project aims to include whole-genome sequence data for several hundred people in different locations of Africa within the next two years. MalariaGen[11] is providing new insights into how to conduct genetic studies of common diseases in African populations, which are far more genetically diverse than European or Asian populations. Researchers at the WTSI have also demonstrated a way forward for GWAS in Africa.[12] The researchers carried out a GWA study of thousands of Gambian children, in search of genetic variants associated with resistance to fatal forms of malaria. The Human Heredity and Health in Africa Project (H3 Africa), a partnership established by the US NIH and the Wellcome Trust, will support population-based genetic studies in Africa, including research into common, non-communicable disorders such as heart disease and cancer, as well as infectious diseases such as malaria.

75. In addition, there is now conclusive evidence of the contribution of genetic factors in the aetiology of specific diseases that can be more prevalent in certain populations and ethnic groups, including sickle cell anaemia in people of West African descent, thalassemia in people of Mediterranean origin, Tay Sachs disease among ethnic groups from Eastern Europe and a higher incidence of cystic fibrosis in people from Western Europe. However, although there are undoubtedly exciting new opportunities to better understand genetic variation across populations, there are dangers in making generalisations based on genetics alone and the concepts of genetic determinism and discrimination remain contentious.

10 The International HapMap Consortium. The International HapMap project. Nature 2003;426:789–796

11 http://www.malariagen.net/

12 Jallow M et al. Genome-wide and fine-resolution association analysis of malaria in West Africa. Nat Genet 2003;41:657–665

Case study 1: Human genetics at the Wellcome Trust Sanger Institute

Background

- The Wellcome Trust Sanger Institute was established in 1992 to play a substantial part in the sequencing and elucidation of the human genome and to further our knowledge of genomes.
- Since 1992, the Sanger Institute has spent approximately 32 per cent (£238 million) of the total amount that the Wellcome Trust has committed to human-genetics-related research (£740m).

Discoveries

- Since 1995, the Sanger Institute has generated the genome sequences of a range of significant human pathogens, including *Mycobacterium tuberculosis* (1998), *Plasmodium falciparum* (2002), the 'superbug' MRSA (methicillin-resistant *Staphylococcus aureus*) (2004), *Entamoeba histolytica* (2005), *Clostridium difficile* (2006), *Schistosomiasis mansoni* (2009) and three *Leishmania* species: *L. major* (2005), *L. braziliensis* (2007) and *L. infantum* (2007).
- In 2003, the International Human Genome Sequencing Consortium announced the completion of the Human Genome Project. The Sanger Institute drove the UK's contribution to the Project, sequencing one-third of the human genome, the largest single contribution from any centre.
- The Cancer Genome Project, developed in 1999 and seen as a paradigm for the use of sequence information, is using the human genome as a reference to systematically search for the genes implicated in human cancer. The Project achieved its first success in 2002 with the discovery of mutations in the *BRAF* gene that lead to malignant melanoma; in 2009, it published the first complete genome sequence of two cancers: malignant melanoma and small-cell carcinoma of the lung.
- The ENCODE initiative, a research consortium launched in 2003, aims to identify and catalogue all functional elements in our DNA sequence by analysing in detail one per cent of the human genome. Results of the pilot programme in 2007 revealed that almost every base in the test section of the genome is transcribed into RNA.
- The Wellcome Trust Sanger Institute was a major lead in the inception and development of the Wellcome Trust Case Control Consortium (WTCCC). Between 2005 and 2007, the WTCCC pioneered the use of genome-wide association studies to examine genetic variation in seven common, complex diseases including bipolar disorder, coronary artery disease, Crohn's disease, hypertension, rheumatoid arthritis, type 1 diabetes and type 2 diabetes. This study has identified more than 90 disease-causing variants.
- The 1000 Genomes Project aims to analyse 1000 genomes from different global populations. It will generate a highly detailed map of human genetic variation, from single-letter changes in the DNA to larger areas that have been duplicated or deleted.
- In 2006, scientists from the Sanger Institute were part of a multinational team who discovered that copy number variation (CNV) – the loss or duplication of chunks of DNA from the genome – was much more common than previously thought, affecting 12 per cent of the genome. CNV has been implicated in several human genetic diseases and affects susceptibility to HIV and malaria infection.

Research leaders

- The Sanger Institute is home to more than 30 faculty, whose expertise ranges from genomics to biomedicine.
- In 2003 the Institute established four-year basic PhD programmes and in 2008, three-year clinical PhD programmes and a programme for postdoctoral fellows.
- The Wellcome Trust Advanced Courses programme, hosted at the Sanger Institute, provides practical training in the latest biomedical research techniques and bioinformatics tools for senior PhD students, postdocs and clinicians in dedicated teaching facilities. The programme runs overseas courses in South America, Africa and South-east Asia in collaboration with local institutions and the Wellcome Trust Major Overseas Programmes.

Research environment

- Research at the Sanger Institute is underpinned by world-class core facilities, including one of the most powerful computer facilities in Europe.
- Since its inception, the Institute has increased output of finished DNA sequence by a factor of over 1000 with new technology for next-generation sequencing.
- The Institute provides open access and free release of data and resources for biology. These include highly curated databases such as the Ensembl genome browser, COSMIC (a database of somatic mutations in human cancers) and the DECIPHER database (which re-uses the Ensembl infrastructure to connect human clinical phenotypes with genotype information).
- The Institute has an active Public Engagement team, who hosted over 5500 visitors during 2006–2009, including schoolchildren aged 14+. All students and postdoctoral fellows at Sanger are required to take formal training in communication and public engagement.

Influence

- In 2008, the Thomson Reuters index ScienceWatch announced that the Sanger Institute and the European Bioinformatics Institute – which shares the Genome Campus in Cambridge with Sanger – were the top two most influential research institutions in the UK. The index ranked scientific publications from UK institutions from 2003 to 2007 according to their impact among researchers. The listing also named the Sanger Institute's Dr Richard Durbin as the most influential UK researcher.
- As well as the human genome, the Institute has sequenced the genomes of numerous human pathogens, including those that cause tuberculosis, malaria, leprosy and diphtheria. Its genome projects continue to identify potential targets for new drugs or vaccines.

Key publications

Aulchenko YS et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet 2009;41(1):47–55

Conrad DF et al. Origins and functional impact of copy number variation in the human genome. Nature 2010;464(7289):704–12

Greenman C et al. Patterns of somatic mutation in human cancer genomes. Nature 2007;446(7132):153–8

Jallow M et al. Genome-wide and fine-resolution association analysis of malaria in West Africa. Nat Genet 2009;41(6):657–5

Pleasance ED et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 2010;463(7278):191–6

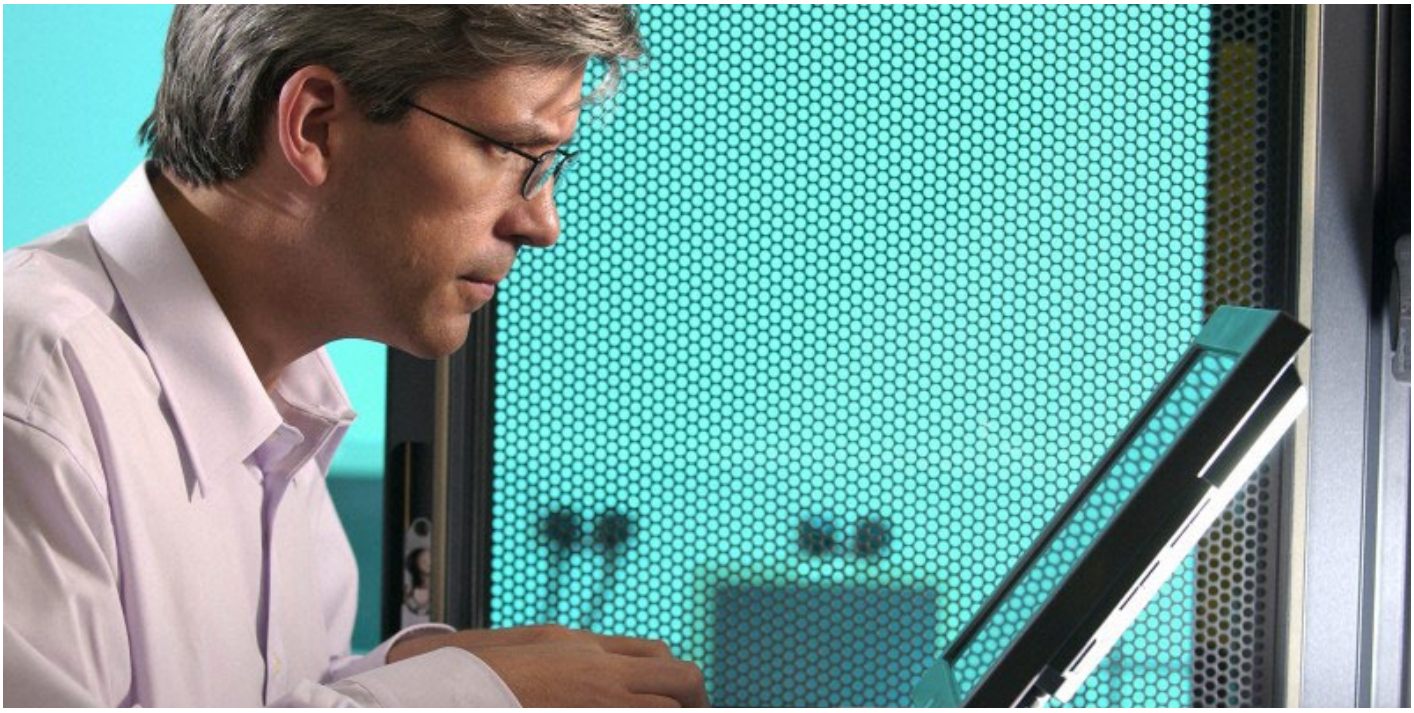
Stranger BE et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 2007;315(5813):848–53

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004;431(7011):931–45

Key projects

- Human Genome Project
- Wellcome Trust Case Control Consortium
- 1000 Genomes Project
- Cancer Genome Project
- Malaria Genomic Epidemiology Network (MalariaGen)

Case study 2: Professor Lon Cardon investigating the genetics of complex human disease



Summary

As a statistical geneticist, Professor Lon Cardon has spent his career trying to decipher the relationship between a person's genes and the complex diseases that affect them. His career combines time working in academia with periods spent in the pharmaceutical industry. He has been part of several major national and international collaborative studies looking at the genetic basis of common diseases.

Background

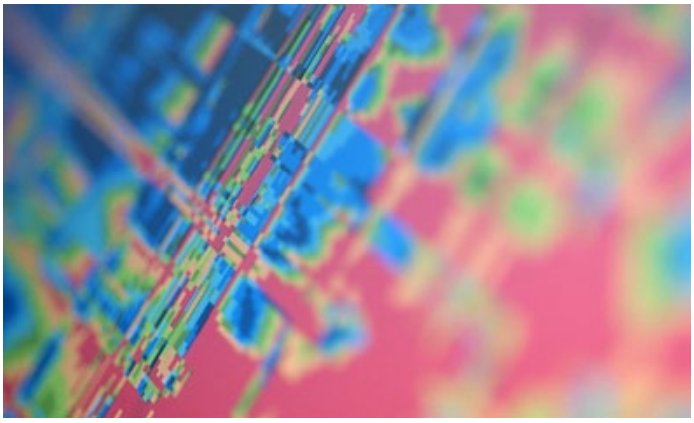
Professor Cardon completed a PhD in quantitative genetics and a postdoc in mathematics before joining a small biotechnology company in California. It was this experience that helped him to recognise the importance of bringing together researchers from different disciplines to unpick the genetics of human disease.

In 1998, with support from the Wellcome Trust, Professor Cardon moved to the UK and joined a large interdisciplinary academic research group at the Wellcome Trust Centre for Human Genetics, University of Oxford, under the Directorship of Professor Tony Monaco. In 2001, he was made a Wellcome Trust Principal Research Fellow, and he continued working there until 2008.

As head of Bioinformatics and Statistical Genetics at the Centre, Professor Cardon played a lead role in the development of several large international collaborations. These included two major projects attempting to understand genetic differences and similarities in humans. The SNP Consortium identified and mapped over 1.5 million single-nucleotide polymorphisms (SNPs), which occur when a single nucleotide (building block of DNA) is replaced with another. These changes may be associated with a predisposition to or protection from particular diseases and may affect how a person reacts to bacteria, viruses, drugs and other substances. The International HapMap Project looked at combinations of SNPs that are inherited together (known as haplotypes). This brought together researchers and funding bodies from Canada, China, Japan, Nigeria, the UK and the USA. The HapMap is expected to be a key resource for helping researchers to identify genes affecting health, disease, and responses to drugs and environmental factors.

During his time at the Oxford Centre, Professor Cardon also co-initiated and worked on the Wellcome Trust Case Control Consortium (WTCCC), a collaboration between 50 research groups across the UK designed to explore the use of genome-wide association studies in understanding the genetic basis of many common diseases. Established in 2005, the WTCCC represented a new phase in biomedical research involving large and complex datasets and encouraging research groups to collaborate and work towards new discoveries rather than competing with each other. This new model for large genetic association studies, with interdisciplinary research at its heart, has set a standard for subsequent international scientific collaborations.

Professor Cardon is currently Head of Genetics at GlaxoSmithKline.



Advance

The WTCCC, drawing on data from the SNP Consortium and the International HapMap Project, has started to reveal key insights into the genetic basis for common diseases. The WTCCC's first paper, published in *Nature* in June 2007, highlighted new genetic regions associated with seven common diseases including rheumatoid arthritis, coronary heart disease, type 1 diabetes and Crohn's disease. It also detailed the first finding of a gene linking the latter two conditions, *PTPN2*.

It is anticipated that the strength of genome-wide association studies, as used by the WTCCC, lies in identifying the range of genes that are implicated in particular conditions, therefore highlighting an individual's susceptibility. Although this does not mean that such studies will enable us to predict the occurrence of illness, they will be hugely valuable in helping to unravel the mechanisms involved in disease development.

In his current role, Professor Cardon continues to collaborate with colleagues in academia. He sits on the Board of Directors of the Serious Adverse Events Consortium, an international collaboration founded in August 2007 to investigate the genetic basis of serious drug-induced side-effects, bringing together the Trust, medicine regulators, academic researchers and representatives from ten international pharmaceutical companies.

References

Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447(7145):661–78.

Wellcome Trust Case Control Consortium. Association scan of 14,500 nsSNPs in four common diseases identifies variants involved in autoimmunity. Nat Genet 2007;39(11):1329–37.

Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. Nat Genet 2006;38(6):659–62.

International HapMap Consortium. A haplotype map of the human genome. Nature 2005;437(7063):1299–20.

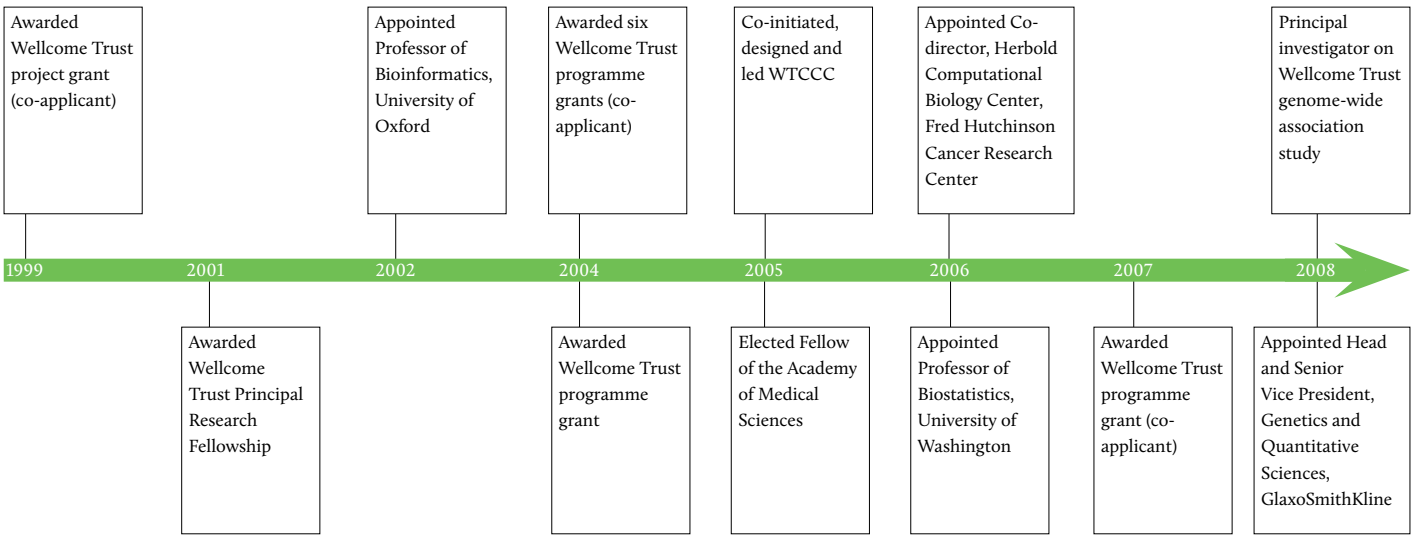
Palmer LJ, Cardon LR. Shaking the tree: mapping complex disease genes using linkage disequilibrium. Lancet 2005;366(9492):1223–34.

Marchini J et al. Genome-wide strategies for detecting multiple loci that influence complex traits. Nat Genet 2005;37(4):413–7.

Abecasis GR et al. MERLIN – rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 2002;30(1):97–101.

Cardon LR, Bell JI. Association study designs for complex traits. Nat Rev Genet 2001;2:91–9.

Timeline of Professor Lon Cardon



Case study 3: The Avon Longitudinal Study of Parents and Children (ALSPAC)

Summary

The Avon Longitudinal Study of Parents and Children (ALSPAC), based at the University of Bristol, is the most detailed and comprehensive prospective birth cohort study ever undertaken. It was also the first to have a genetic component, including appropriate consent for DNA analysis, built into the design from the outset. Research arising from the study continues to provide key insights into how genetic and environmental factors influence development, health and disease.

Background

The participants of ALSPAC are 14 000 children from the Avon area of England who were due to be born between April 1991 and December 1992 (the children's parents being recruited during the mothers' pregnancies). ALSPAC's value is in the large amounts of detailed information and biological samples – including maternal pregnancy blood, cord blood and subsequent blood samples – collected systematically.

From pregnancy to the age of six, data were collected mainly by questionnaire, although a 10 per cent subset – the 'Children in Focus' – had six-monthly hands-on examinations. Then, every one or two years from the age of seven, researchers have recorded large numbers of facts about each subject child and their parents. These include a physical examination and the collection of urine and blood samples at different times. There are also detailed questionnaires for the participants about factors such as diet, lifestyle, parent contact and socioeconomic status.

Many scientists collaborate with the ALSPAC team on particular research areas such as genetics, cognition and education, and clinical studies. An independent steering committee oversees the strategies for data collection, archiving and subsequent use. The information is used primarily for the study of the genetic and environmental factors influencing development, health and disease. These include the factors affecting childhood obesity, allergies, dyslexia, visual deficiencies, autism, asthma, antisocial behaviours and cognition.

Advance

ALSPAC faced many challenges from the outset. Many regarded such a large and comprehensive study design as unfeasible. At the time, some thought it was unethical to bleed babies purely for research – particularly for genetic analysis, the individual results of which would not be disclosed to the participants. There was concern about recruiting sufficient eligible women, and finally, the major funders made it clear that the large sums needed would have to be pieced together from project grants.

ALSPAC has been able to meet all of these challenges with the help of the University of Bristol. Grant support was gathered from 20 different funders, including the Wellcome Trust. A dedicated Law and Ethics Committee was created in 1989 and its research, advice and advocacy allowed appropriate consent procedures and strategies to be developed. ALSPAC pioneered the use, in child population studies, of anaesthetic cream when taking blood from a vein. A Genetic Advisory Committee was established in 1995, with members drawn nationally. Funding from the Medical Research Council helped to establish the ALSPAC DNA bank, and in 2001, core funding helped to consolidate the DNA bank and cell lines in a purpose-designed laboratory, which opened in 2003. The laboratory was also able to take on DNA banking and cell-line generation for the British 1958 birth cohort, which among other studies provided a control series for the Wellcome Trust Case Control Consortium.

With developments in large-scale genotyping, the number of published papers incorporating ALSPAC genotypes has risen over the years, reaching 47 by 2009. Studies range from large collaborative genome-wide association studies incorporating other sample sets to focused candidate gene studies in relation to specific outcomes, exposures or both.

One major study in which ALSPAC played a significant role was the discovery of the influence of variation in the *FTO* gene on body mass index (BMI). The huge range of data in ALSPAC allowed rapid follow-up studies examining how the *FTO* effect on BMI might be mediated. For example, *FTO* genotyping provided an opportunity to explore causal links between obesity and bone outcomes, which suggested that fat mass is on the causal pathway for bone mass in children.

The ALSPAC design also allows for an assessment of population risk for reported disease-associated variants. For example, analysis has shown that the two commonest *Filaggrin* null mutations (carried by 6 per cent of the cohort)

accounted for 15 per cent of the population's attributable risk for eczema and for atopic asthma.

Another strength of ALPSAC is the availability of maternal DNA, which allows researchers to explore the effect of the maternal genotype on fetal outcomes independently of the fetal genotype. In one study, two variant genes linked to blood glucose regulation and diabetes, glucose kinase and *TCF7L2*, were shown to influence normal-range birth weight with the effect driven by the maternal genotype. Transgenerational effects have also been demonstrated down the male line, with a correlation found between a father who started smoking in his mid-childhood and greater BMI in his sons.

The transgenerational theme is one element of a large ongoing (and Trust-funded) genome-wide association study looking at the association of pregnancy phenotypes, such as weight gain and blood pressure change, with future cardiovascular and metabolic phenotypes in women and their offspring. Genome-wide genotypes are also currently being generated for ALSPAC children. The study is part of several international consortia including GIANT (anthropometric traits), MAGIC (continuous glycaemic traits), EGG (early growth genetics), GABRIEL (asthma and related phenotypes) and EAGLE (early genetics and lifecourse epidemiology).

There is increasing evidence that enduring life-course effects are associated with changes in the epigenetic features of the individual's genome, such as DNA methylation. ALSPAC is well placed to explore this, as it holds DNA samples collected at different points in the participants' lives, which can be used to investigate such associations. DNA methylation profiling is likely to form a key part of ALSPAC's continued exploration of how genetic and environmental factors influence development, health and disease.

References

Mumford SE. Children of the 90s: ethical guidance for a longitudinal study. Arch Dis Child Fetal Neonatal Ed 1999;81(2):F146–51.

Mumford SE. Children of the 90s II: challenges for the ethics and law committee. Arch Dis Child Fetal Neonatal Ed 1999;81(3):F228–31.

Jones RW et al. A new human genetic resource: a DNA bank established as part of the Avon longitudinal study of pregnancy and childhood (ALSPAC). Eur J Hum Genet 2000;8(9):653–60.

Golding J et al. ALSPAC – the Avon Longitudinal Study of Parents and Children. I. Study methodology. Paediatr Perinat Epidemiol 2001;15(1):74–87.

Pembrey ME et al. Sex-specific, male-line transgenerational responses in humans. Eur J Hum Genet 2006;14(2):159–66.

Weedon MN et al. A common haplotype of the glucose kinase gene alters fasting glucose and birth weight: association in 6 studies and population genetics analyses. Am J Hum Genet 2006;79(6):991–1001.

Frayling TM et al. A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. Science 2007;316(5826):889–94.

Freathy RM et al. Type 2 diabetes *TCF7L2* risk genotypes alter birth weight: a study of 24,053 individuals. Am J Hum Genet 2007;80(6):1150–61.

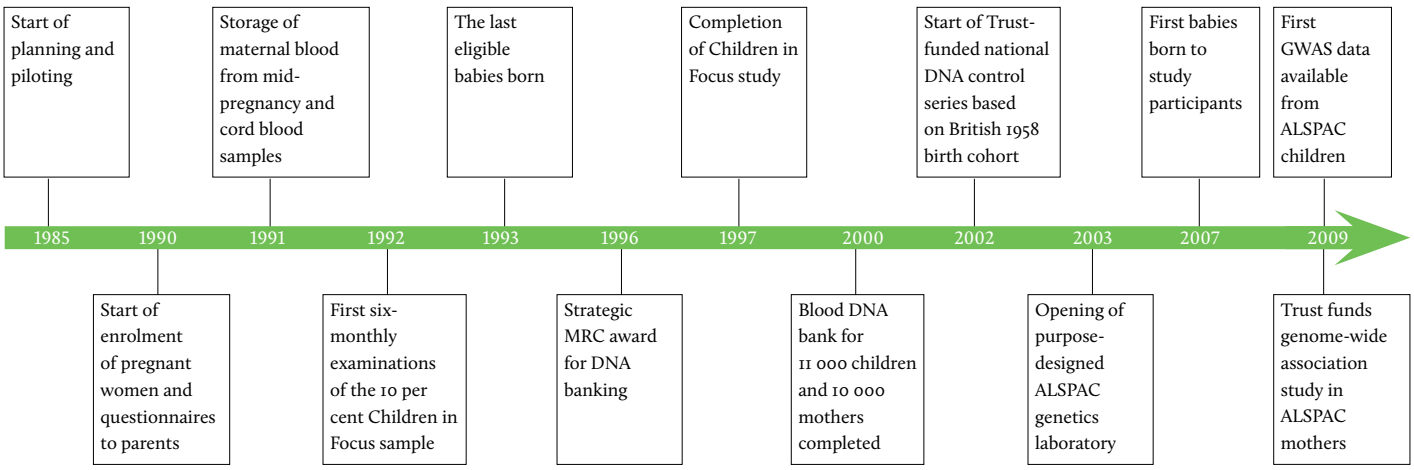
Henderson J et al. The burden of disease associated with filaggrin mutations: a population-based, longitudinal birth cohort study. J Allergy Clin Immunol 2008;121(4):872–7.e9.

Timpson NJ et al. The fat mass- and obesity-associated locus and dietary intake in children. Am J Clin Nutr 2008;88(4):971–8.

Johnson L et al. Dietary energy density affects fat mass in early adolescence and is not modified by *FTO* variants. PLoS ONE 2009;4(3):e4594.

Timpson NJ et al. How does body fat influence bone mass in childhood? A Mendelian randomization approach. J Bone Miner Res 2009;24(3):522–33.

Timeline



Case study 4:
Professor John Todd
a multidisciplinary approach to understanding type 1 diabetes

Summary

Professor John Todd is Director of the Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory at the University of Cambridge. He uses a multidisciplinary approach to understanding type 1 diabetes that ranges from the genetics of susceptibility to the cellular and molecular events underlying the disease. His findings extend to other autoimmune diseases and are establishing the principles of complex disease genetics.

Background

A major priority in diabetes research is to understand the early events that lead to the destruction of the beta cells in the pancreas that produce insulin. A milestone in the field came in 1987, when Professor Todd discovered that a particular variant of an HLA class II gene of the major histocompatibility complex (MHC) is a key determinant of susceptibility and resistance to type 1 diabetes.

In 1990 Professor Todd was awarded a Wellcome Trust Senior Research Fellowship in Basic Biomedical Science. With Trust funding he has gone on to either lead or be directly involved in the identification of almost all of the genes that are known to influence the development of type 1 diabetes. These findings have led to an increased understanding of the principles of complex genetic disorders in general.

Many of these discoveries have been facilitated by the research resources and collaborations that he has developed throughout his career. In particular, as founding director of the Diabetes and Inflammation Laboratory, he has built a world-class research centre at the University of Cambridge. He has also been central to the establishment of two major international consortia: the Type 1 Diabetes Genetics Consortium and the Wellcome Trust Case Control Consortium.

He led the creation of the Juvenile Diabetes Research Foundation/Wellcome Trust case sample set that includes over 8000 samples and cell lines from diabetic people. In 2009, this resource was central to the success of the Diabetes and Inflammation Laboratory's genome-wide association meta-analysis (as part of the Type 1 Diabetes Genetics Consortium) that identified 42 genetic loci involved in susceptibility to type 1 diabetes, 18 of which had not been identified previously.

With initial funding from the Trust and the Juvenile Diabetes Research Foundation, Professor Todd has also founded and assembled the Cambridge BioResource, a registry of over 9000 volunteers who are willing to donate blood samples repeatedly for studies of the effects of genes and their variants. This resource, now funded mainly by the National Institute for Health Research Cambridge Biomedical Research Centre, has been central to studies of the connections between genotypic and phenotypic variation in complex genetic disease. The success of this resource has led to efforts in the USA to develop registries of genotyped normal individuals willing to be recalled to donate blood samples for use in experiments. It is hoped that this kind of bioresource will be extended across the UK.

Professor Todd is extending his findings in collaborations with biotechnology companies – including Life Technologies, Roche and Illumina – to develop new approaches to the study of medical genetics and gene and protein expression. Part of this will take place in the newly funded Eastern Region Sequence and Informatics Hub, of which Professor Todd is the Director. Supported by a grant from the Medical Research Council as well by the University of Cambridge and the Cambridge Biomedical Research Centre, the Hub will provide next-generation sequencing facilities for the University and researchers across the region.

Advance

Professor Todd and colleagues are investigating how type 1 diabetes develops by identifying and analysing genetic variants that affect disease susceptibility. He first mapped the major genes involved – the MHC HLA class II genes, which are normally involved in T-cell development, immune tolerance and responses to infection. He then went on to find one of the first non-MHC genes to be implicated in the disease, *CTLA-4*, which is also involved in immune tolerance.

In 2009, Professor Todd and colleagues used high-throughput sequencing to show that four rare variants of the *IFIH1* gene lower the risk of type 1 diabetes. The *IFIH1* gene encodes a cytoplasmic protein, often called MDA5, that detects viral RNA, triggering an interferon response to the infection. Importantly, the variants that protect against diabetes are likely to have reduced *IFIH1* activity. This is consistent with the previously known link between

enteroviral infection and type 1 diabetes. It suggests that one of the events in disease development could be the body's response to infection with an enterovirus via the normal function of *IFIH1*, which includes raised levels of the type 1 interferons that could accelerate beta-cell destruction by the cytotoxic T cells of the immune system.

Professor Todd is currently investigating correlations of genotype and cellular and molecular phenotype to identify the events causing type 1 diabetes. For example, his laboratory has identified levels of the high-affinity interleukin-2 receptor on certain T cells in blood plasma as a potential factor influencing the disease.

The progress in understanding type 1 diabetes also has implications for other autoimmune diseases. Some of the key susceptibility genes for type 1 diabetes are shared by other autoimmune diseases such as coeliac disease, multiple sclerosis, Graves' disease and rheumatoid arthritis. Professor Todd and his colleagues have recently looked at the *IL2RA* gene region to see whether there are common or distinct alleles of the gene that contribute to type 1 diabetes and multiple sclerosis susceptibility. While one allele is common to both diseases, each also has a disease-specific allele of *IL2RA*.

The next big challenge for medical geneticists is to understand how genetic variants influence phenotype and contribute to disease states. Autoimmune diseases such as Graves' disease, type 1 diabetes, coeliac disease and rheumatoid arthritis are known to often cluster in families. Professor Todd's work on the extent of genetic overlap between these diseases is helping to establish the principles of complex genetic diseases.

References

Ueda H et al. Association of the T cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 2003;423(6939):506–11.

Smyth DJ et al. A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus at the interferon-induced helicase (IFIH1) region. Nat Genet 2006;38(6):617–9.

Todd JA et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 2007;39(7):857–64.

Nejentsev S et al. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-A and HLA-B. Nature 2007;450(7171):887–92.

Smyth DJ et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med 2008;359(26):2767–77.

Nejentsev S et al. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 2009;324(5925):387–9.

Barrett JC et al. Genome-wide association study and meta-analysis indicates that over 40 loci affect risk of type 1 diabetes. Nat Genet 2009;41(6):703–7.

Dendrou CA et al. Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable bioresource. Nat Genet 2009;41(9):1011–15.

Wallace C et al. The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 encodes susceptibility to type 1 diabetes. Nat Genet 2010;42(1):68–71.

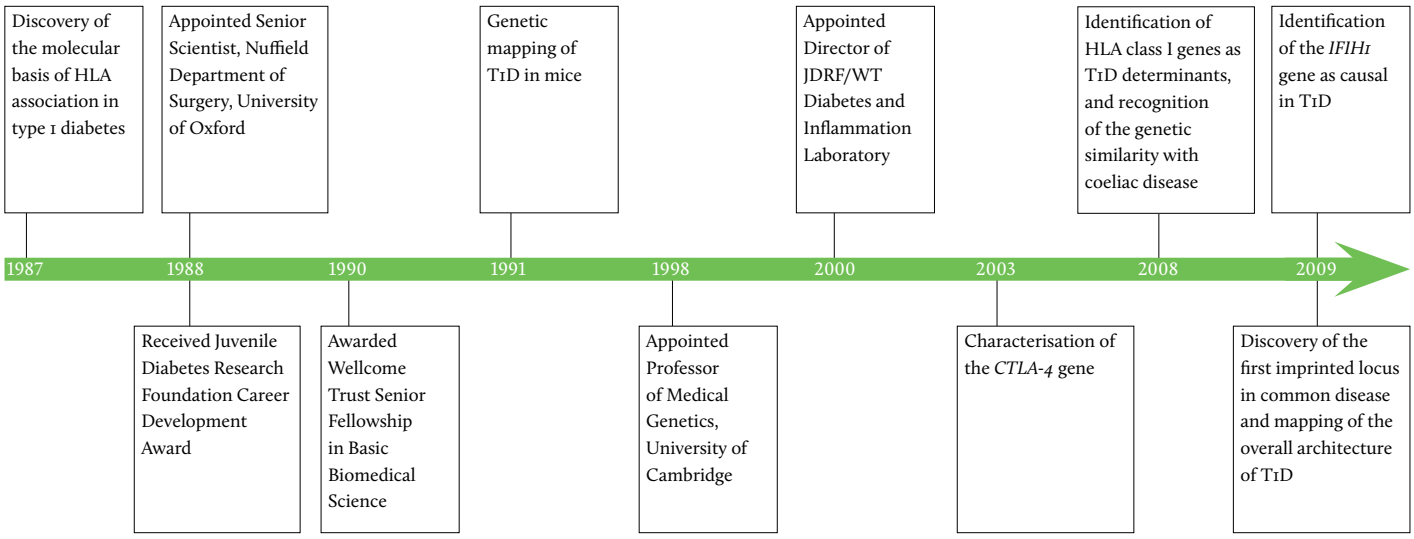
Discoveries

- Sequence variation of HLA class II genes within the MHC predisposes to type 1 diabetes (T1D)
- Identified four new chromosome regions that increase the risk of developing T1D
- Implicated the IL2RA region polymorphism in T1D

Partnerships

- Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammatory Laboratory established
- Cambridge Institute for Medical Research
- T1D Genetics Consortium and WTCCC

Timeline of Professor John Todd



Case study 5: Food, fat and *FTO* – genetic clues to obesity

Summary

The *FTO* gene was one of the first outputs from the Wellcome Trust Case Control Consortium, a pioneering collaboration between the UK's leading human genetics researchers. First identified as a risk factor for type 2 diabetes, a common variant in the *FTO* gene was found to exert its effects by increasing fat mass. Its discovery has opened up an entirely novel area of research, which is providing new insight into weight control and suggesting possible therapeutic strategies for obesity. More generally, it illustrates how large genomic studies can identify genetic influences that would never have been uncovered by conventional approaches.

Background

For some rare, highly heritable diseases, there is a straightforward link between gene and disease. But these diseases are the exceptions. For most common conditions, the link between gene and disease is much more tenuous. Many different genes (and environmental factors) contribute to a disease.

Over the years, researchers have searched for these genetic risk factors. Despite many claimed associations, only a handful have withstood rigorous scrutiny. The main problem was one of 'false positives' – statistical associations between gene and disease identified in one group of people but not subsequently confirmed in another.

By 2006, the field was ripe for a major leap forward. A successful association study has two main requirements: a large, well-characterised set of patients, and tools to scan the entire genome at high resolution speedily and economically. The Wellcome Trust Case Control Consortium, established by the Wellcome Trust in 2005, combined both elements. A collaboration between 50 leading groups, the Consortium was able to pool samples from many centres across the UK. Most critically, information from the Human Genome Project – along with associated advances in technology – provided a means to characterise ('genotype') many hundreds of thousands of genetic markers across the entire human genome. A huge logistical operation, this analysis was made possible by the high-throughput genotyping facilities at sites such as the Wellcome Trust Sanger Institute.

Importantly, associations were rigorously tested in other patient populations, to ensure that the genes identified genuinely were contributing to disease.

The first results from the Consortium were published in 2007, with a landmark paper describing associations for seven common diseases in 14 000 patients. Subsequent papers described associations in particular diseases in more detail.

FTO was among the first wave of gene discoveries, thanks principally to the work of Andrew Hattersley and Tim Frayling at Peninsula Medical School in Exeter and Mark McCarthy at the Wellcome Trust Centre for Human Genetics in Oxford. Although it turned up as a risk factor for type 2 (adult-onset) diabetes, its effects were mediated solely through its impact on fat mass – excess fat mass being a known risk factor for diabetes. *FTO* was the first gene found to affect body mass index and fat mass in the general population.

The researchers identified a position in the DNA sequence of the *FTO* gene that occurred in two versions (alleles), one of which is the risk allele for increased fat mass. An individual can carry none, one or two copies of the risk allele. Around one in six Caucasians carries two copies of the risk allele; they have a 1.7-fold higher risk of obesity and on average are 2.5 kg heavier than people with no risk alleles. Larger fat deposits account for essentially all of the extra weight.

FTO had never previously been considered a candidate gene for obesity and nothing in its sequence would suggest a role in weight control. It illustrates a strength of genome-wide studies, which begin with a blank canvas and make no initial assumptions about which genes might be involved in a condition.

Advance

In a remarkably short time, the number of verified disease associations has gone up from a handful to many hundreds – with the numbers still rising. The Consortium has acted as a model for many other large-scale international collaborations tackling other diseases, metabolic traits known to affect disease and other physiological variables – and even traits such as reading ability. Researchers from the Consortium are now assessing the impact of other forms of genetic variation, such as copy number variation (loss or gain of chunks of DNA). Indeed, working with Matt Hurles at the Wellcome Trust Sanger Institute and making use of Consortium data, Wellcome Trust Senior Research Fellow Sadaf Farooqi and colleagues have identified several rare deletions causing severe early-onset obesity.

The discovery of *FTO* has opened up an entirely new area of research. In just two years, it has spawned more than 100 papers. Its links to weight have been confirmed in numerous global populations (including groups that have very different diets and lifestyles from western Europeans, such as Japanese and Chinese people).

Already a great deal has been discovered about *FTO*. The risk variant is globally distributed. It is slightly less common in the Far East. It had no impact in a rural African population, possibly because the more complex genetic structure of

Africans means the effect is hard to find. Alternatively, it is possible that the effect of the *FTO* variants only becomes apparent when food is plentiful (as such, its effect on weight may be a recent phenomenon, linked to the ready availability of food in high-income countries over the past few decades).

In terms of function, evidence is accumulating that *FTO* affects food intake (although some studies in mice suggest that it might also affect metabolism). Work on the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort, for example, has suggested that the risk variant is associated with greater food intake. The fact that the *FTO* gene is highly active in areas of the brain involved in energy balance and appetite control also implies a role in food consumption. Indeed, a team led by Giles Yeo, a Wellcome Trust Research Career Development Fellow in Steve O'Rahilly's lab in Cambridge, has found that manipulating *FTO* protein levels in these brain areas in the rat can increase or decrease the animals' food intake.

In terms of biochemical activity, Chris Schofield, Chris Ponting and Frances Ashcroft in Oxford, Tomas Lindahl at Cancer Research UK's London Research Institute and Professor O'Rahilly have shown that the *FTO* protein is an enzyme that modifies nucleic acids. How this is linked to its metabolic role is far from clear. Moreover, *FTO* almost certainly has a range of roles. Mutation of *FTO* has been found to cause a lethal genetic disease with multiple developmental abnormalities, suggesting the gene has an important role in early human development.

Much has been discovered about *FTO* in a remarkably short time, from molecular studies of its mode of action to epidemiological findings on the distribution of different forms of the gene. As more is uncovered about its role, work on *FTO* could even reveal new targets for therapeutics designed to control body weight.

References

Frayling TM et al. A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. Science 2007;316(5826):889–94.

Bochukova EG et al. Large, rare chromosomal deletions associated with severe early-onset obesity. Nature 2010;463(7281):666–70.

Church C et al. A mouse model for the metabolic effects of the human fat mass and obesity associated *FTO* gene. PLoS Genet 2009;5(8):e1000599.

Timpson NJ et al. The fat mass- and obesity-associated locus and dietary intake in children. Am J Clin Nutr 2008;88(4):971–8.

Tung YC et al. Hypothalamic-specific manipulation of Fto, the ortholog of the human obesity gene *FTO*, affects food intake in rats. PLoS One 2010;5(1):e8771.

Gerken T et al. The obesity-associated *FTO* gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. Science 2007;318(5855):1469–72.

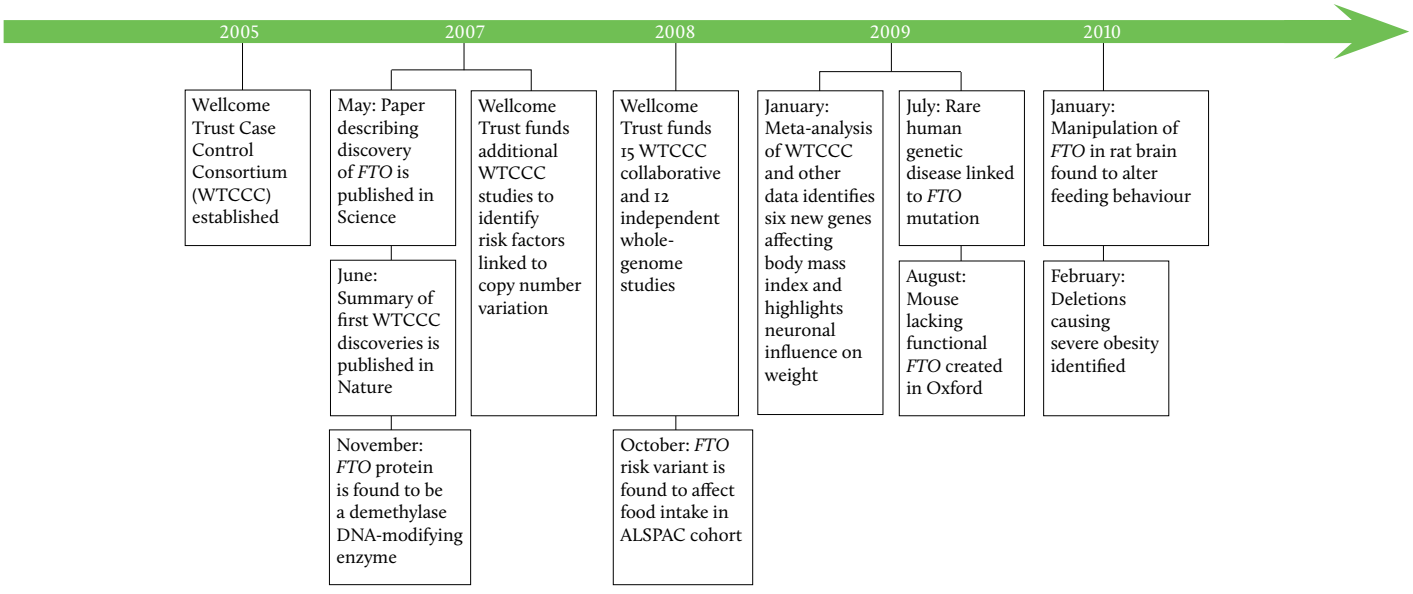
Boissel S et al. Loss-of-function mutation in the dioxygenase-encoding *FTO* gene causes severe growth retardation and multiple malformations. Am J Hum Genet 2009;85(1):106–11.

Funding

Many of these studies received support from other funders as well as the Wellcome Trust.

www.wtccc.org.uk

Timeline



4. Looking forward: speculations on the future of human genetics

"[The future is…] partnership on specific scientific issues and the attempt to get the ground rules and the behavioural rules to scientific collaboration onto a sharing and helpful basis with [emerging] places and those that we just haven't had those relationships with – that seems very important, very, very important.
Wellcome Trust Expert Group on Human Genetics, October 2009

76. As described, many of the Wellcome Trust's major contributions to the field are thought to have arisen from its timely intervention and fostering of international partnerships and collaborations. It has not always been about how much money is invested; the Wellcome Trust's strengths in this field over the past 20 years have been its independence, influence, foresight and agility, enabling it to take key strategic decisions, which are thought to have helped to shape the development of human genetics.

77. In the area of human genetics, since the early 1990s much of the research leading to perhaps the most significant breakthroughs has resulted from 'big', technology-based, collaborative efforts. Now that the human genome has been deciphered, the paradigms within human genetics are shifting; the challenge is to harness and act upon this new understanding to bring about real impact on the health and well-being of populations – and support the field in ways best suited to deliver these impacts. This may involve working with new partners and supporting new fields.

78. This portfolio review highlighted several areas where there is a need to break new ground and build on the opportunities presented by human genetics to enhance our understanding of the basis of disease aetiology.

79. The following list is by no means exhaustive and, as with any review or consultation, a different set of researchers and/or experts might have reached a slightly different set of conclusions and/or priorities. Nevertheless, we believe that by addressing some of the research needs highlighted below and building on existing infrastructures, critical mass and collaborations – which may represent relatively modest marginal costs and economies of scale and effort – the Wellcome Trust and/or other funders interested in supporting the field will help to drive the field forward. The following issues were raised at the Expert Group meeting on human genetics hosted by the Wellcome Trust in October 2009 and are detailed in the following sections: (**4.1–4.11**):

4.1 A need for underpinning research into basic biological mechanisms and genetics

"The complexity of control, of expression, where and when and so on is great…which once again absolutely hammers home the point about funding plenty of basic work on specific mechanisms, finding out what is going on…I think it is absolutely fine to go on collecting data on a large scale, I think it is very important, but that should not be the only thing going on. We need the most intelligent best work being done in the traditional way."
Wellcome Trust Expert Group on Human Genetics, October 2009

80. The Expert Group identified several areas where the Wellcome Trust could have a role in underpinning research into basic biological mechanisms and genetics. The main areas identified concerned:

- continued importance and need for research into monogenic disease
- increased capacity for research into the genetics of infectious disease and pathogen sequencing is needed
- the relative neglect of pharmacogenetics
- potential overemphasis on translational research
- the development of systems biology.

81. As described, and in the past decade in particular, there has been a marked shift in the nature of human genetics, from the time spent on **'wet-lab'** experiments to the time spent on **'dry-lab'** analysis. Since the culmination of the HGP, there has been an explosion in the amount of basic genomic data being generated, as more and more organisms are being sequenced. However, much is still unknown about gene regulation and function, and while genomic science has much to offer basic biology, there is synergy in bringing the two together.

82. To make sense of the vast amount of data being generated, coordination between basic biologists and bioinformaticians is essential; computer-based genome hypotheses still require experimental validation by biologists. Collaboration between wet- and dry-lab scientists will accelerate experimental analyses and move the field forward.

83. A vast amount has already been learnt about both normal and abnormal gene action from studying monogenic diseases, due in large part to knowledge gained from human genome sequencing work. Scientists have been using the human genome sequence as a reference point in the attempt to map **monogenic (single-gene) disease** to its corresponding genes. Much of our current understanding of genetics and disease has come from our understanding of monogenic diseases such as sickle cell disease and thalassaemias, which have revealed the multi-layered complexity of even the simplest monogenic disease. To this end, monogenic disease research is worthy of our attention and it seems vital that work on these is not neglected in the future; a better understanding of some of them at least has had, and will have, enormous importance for a better understanding of some of the commoner diseases with relatively small genetic components. There have been some advances in prenatal diagnosis and therapy of monogenic diseases, although the Expert Group were in agreement that there is an opportunity for funders to play a more active part in supporting research into developing cures for monogenic disorders, particularly because of the extraordinary light they can throw both on normal gene regulation and on diseases where there is a much smaller genetic component.

"In terms of understanding monogenic disease, it has gone reasonably well. I think with the therapy of monogenic disease, people expected too much too soon. We are just getting to the stage where this is starting to look feasible for one or two diseases. I think this is going to be an area that is well worth supporting."
Wellcome Trust Expert Group on Human Genetics, October 2009

84. As was also reflected in the bibliometric analysis, there was a view among the Expert Group that the impact of **human genetics on infectious disease** biology is less than might have been anticipated to date and the growth rate in human genetics output related to infectious diseases has certainly declined in the most recent decade. There has, for example, been relatively little success in studies looking for common genetic variants that affect susceptibility to infectious diseases such as TB and malaria. In addition, while there have been several significant sequencing breakthroughs of a range of pathogens responsible for major global morbidity and mortality, in general there was a view that more capacity is required for **pathogen sequencing** alongside a better understanding of pathogenesis.

85. In terms of cures and treatments for genetics-based conditions, the Expert Group discussed a relative "lack of any practical pharmacogenetics" to date. Although much is known about the genetic variability in human metabolism and action of drugs, warfarin for example, what is not yet known is whether such genetic knowledge, and hence tailoring of drug dosage, is more effective (cost or otherwise) than careful follow up and control of these drugs at standard dosages. A limited number of trials to address this question are underway but more are required if the value and impact of genetic information on prescribing and treatment regimes is to be maximised.

86. The perceived lack of progress in pharmacogenetics was thought to be the result of a lack of direct funding for pharmacogenetics, the absence of effective collaborations between pharmacologist and geneticists. In addition, the development of the field was thought to be hampered by the lack of phenotypic definitions among populations, which was raised throughout the discussion.

87. Another key issue highlighted by the Expert Group that has arisen due to the practical implications of the human genome project being somewhat over-hyped after its success is the current **overemphasis on "translational" research** (which focuses on a particular disease). Given the enormous biological complexity that is being unravelled by genomic studies, the Expert Group recommended to continue to support research into areas such as systems biology – and any aspect of the further analysis of gene regulation and interaction, as key medical breakthroughs often occur from our enhanced understanding of the basic fundamental molecular and cellular mechanisms of an organism.

"I think it is part of this sort of mass hypnotism with translational research. You have to have a disease name attached before it becomes respectable and my instinct is, not that one shouldn't do this, that would be mad, there has just been an over-concentration on studying disease populations as opposed to studying mechanisms of the genetics of normal biochemistry and development in humans."
Wellcome Trust Expert Group on Human Genetics, October 2009

"Unfortunately, not just the UK, the world, has got so hooked up on this translational research business, over the last while. I think many agencies have really been misled into misdirecting funds in the naïve hope that things that are actually difficult will become easy if you call them translational research."
Wellcome Trust Expert Group on Human Genetics, October 2009

4. Looking forward: speculations on the future of human genetics

4.2 A need for improved phenotypic definition and understanding among populations

"My worry about some of the work is the level of phenotypic definition that is going into some of these studies. For example some of the big scale malaria studies, how do you define severe malaria? Can we define heterogeneity before we even give it to the geneticists?"
Wellcome Trust Expert Group, October 2009

88. The Expert Group emphasised that research into phenotypic definitions in both healthy and non-healthy populations needs to be strengthened for the field of human genetics to advance. The **lack of definitive phenotypic definition/s** in the field within populations is a limiting factor on the ability of many existing cohort studies to support genetics-based analysis. This presents a major challenge when recruiting populations for studies, and particularly in studies on the genetics of infectious disease; if the phenotype is not precisely defined, it is difficult to detect true gene–disease associations.

89. While perhaps an obvious priority in trying to impact upon human health, much of the human genetics population-based work thus far has been focused on disease populations. There remains much that is unknown about the genetic mechanisms underlying basic cell and developmental processes in healthy individuals. There have, for example, been few large human genetic studies on non-medical traits, such as height or facial features, to enable human model systems to be developed and enhance understanding of how genetic variation in the population contributes to phenotypic variation.

90. **Improved phenotypic data** would have value to all of the current GWAS and cohort studies. The Wellcome Trust Expert Group suggested two approaches to address this: first, to undertake more intensive phenotyping on targeted subsections of existing cohorts (e.g. UK Biobank, ALSPAC), and, second, to build **quantitative phenotyping** and behavioural studies into new studies such as the 2012 Birth Cohort Study.[13]

"In terms of a ray of hope, there could be some quantitative phenotyping done on the tail end of the UK Biobank [and] I hope there is going to be some serious quantitative phenotyping and behaviour studies in the 2012 birth cohort."
Wellcome Trust Expert Group on Human Genetics, October 2009

91. In addition, the Expert Group highlighted the potential of **epigenetics** to support understanding of the interaction between genotype, phenotype and other external factors. Epigenetics – the study of inherited changes in phenotype gene expression caused by mechanisms other than changes in the underlying DNA sequence – has emerged largely out of the acquired knowledge of the human genome. There is an opportunity for funders – including Wellcome – to consider how it might support epigenetics to complement existing research endeavours and offer further insight into phenotype and genotype studies.

"The genotyping is going to be compared with epigenetic revelation profiles, there is huge, particular methylation and histone changes and so on, and now it has become genome wide. [Epigenetics] I personally see it as, having had a little bit of a taste of what is going to come out, going to be a very big part of the interpretation [of the human genome]. I think it will be the next big challenge in the next year or so."
Wellcome Trust Expert Group on Human Genetics, October 2009

4.3 A need to secure clinical involvement in human genetics research

"If I was pushing one thing it would be to think very hard about how one can get better collaboration with critical clinicians on the input of some of these [large-scale studies]."
Wellcome Trust Expert Group on Human Genetics, October 2009

92. Recognising the importance of clinical involvement to help capitalise on the rapid and significant advances that are occurring in human genetics research, the Expert Group discussed the need for greater interaction between geneticists, biologists and clinicians.

13 ESRC providing the lead for the 2012 Birth Cohort Study.

93. Part of the reason for the relative absence of phenotypic data across populations is thought to be the **lack of clinical involvement** in many genetics-based association studies. Clinicians are well placed to help deliver improved phenotypic definitions, and the Expert Group highlighted the importance of closer collaboration between clinicians, geneticists and basic biologists to help take the field forward – and particularly in the context of large-scale cohort studies.

94. In the UK specifically, there is a wealth of potentially useful information held by the NHS, but it is difficult to access and secure the time of many NHS clinicians. In addition, within the Expert Group, there was some support for the concept of establishing a central pathology laboratory in the UK; a facility that could carry out high-quality, secure genetic testing for all pathological disciplines. This could be a facility for use by all clinicians and scientists and would facilitate cross-disciplinary networks and ultimately increase the availability and raise the value of quantitative phenotyping.

"The recommendation of the House of Lords' Genomic Medicine report about having a central pathology laboratory that draws together somewhat disparate disciplines in a hospital and really upping the ante on quantitative phenotype centrally in a hospital, would be of major benefit."
Wellcome Trust Expert Group on Human Genetics, October 2009

95. There is also an opportunity for funders like the Wellcome Trust to consider strategies to secure and support the involvement of clinicians in human-genetics-based work. There is potential for facilities such as the Wellcome Trust Clinical Research Facilities to provide the infrastructure and environment to facilitate **collaboration between clinicians, geneticists and biologists.**

"A major impact for [researchers] has been the Wellcome Trust Clinical Research Facilities. I think that was an absolutely brilliant strategy, the CRFs have been fundamentally important at the very lowest level, just a depot and a place to bring people to take a blood sample or a blood pressure."
Wellcome Trust Expert Group on Human Genetics, October 2009

4.4 The need for high-quality epidemiology and well-powered cohorts and GWAS

"The numbers, population numbers which you need in order to start getting statistical power on variants are actually disturbingly large. And the take that I have is that we are going to have to pull together these phenotype cohorts. We need larger numbers. I actually think there is a big gap."
Wellcome Trust Expert Group on Human Genetics, October 2009

96. Linked to the need for better phenotypic information, the Expert Group highlighted the need for greater breadth and statistical power in GWAS and cohort studies that consider genotype–phenotype interactions. To date, most GWAS have been medical phenotype-driven (e.g. via the WTCCC). The Wellcome Trust Expert Group envisaged that in the foreseeable future, association studies will begin with the genotype, enabling a **'bottom-up approach' to GWAS studies** and that such studies will require the use of longitudinal well-powered cohorts, to advance research into the significance of genetic variation. For more disease-focused cohort analysis, the sample population size required to secure statistical power on rare variants is also very large. Unfortunately, although the potential benefits are high, the costs associated with securing and managing large sample sizes and associated data can be very high and the time frames involved in cohort and longitudinal studies very long.

97. In addition, it was thought that the Wellcome Trust, with its current networks and involvement in major consortia, is well-placed to **support the development of new components to existing cohorts and longitudinal studies** – including ensuring that the appropriate statistical tests will have adequate power. As described, there is an opportunity to work with existing major collaborations, consortia and cohort studies (such as ALSPAC, UK Biobank and the 2012 birth cohort), to maximise the value of the data collected and potentially impact upon understanding of the incidence of morbidity.

4. Looking forward: speculations on the future of human genetics

98. The Expert Group argued that considerable **investment in epidemiology** is essential to increase the value of population-based studies, which requires data analysis, storage and correlation to interpret how genetic variation is related to disease. Given that such studies, by design, require long-term and stable funding commitments to ensure they are adequately resourced and positioned to deliver on their aims, this 'funding for the long term' is something that the Wellcome Trust is well placed to provide.

99. The Expert Group noted that supporting epidemiology had not been a major focus area of the Wellcome Trust to date. However, given that high-quality epidemiology-based research is essential to serve and support strategically important areas of human genetic research, such as GWAS and UK Biobank, the Wellcome Trust might consider how it could support this area in relation to GWAS in the future.

"If you look at the [epidemiology] portfolio prior to half a dozen years ago, the Trust has not been strong in epidemiology…Everyone makes choices but the choices that the Trust has made have fundamentally been to hardly fund epidemiology."
Wellcome Trust Expert Group on Human Genetics, October 2009

4.5 The need to support data sharing and open access to data

"The ability to handle data, enormous amounts of data, has become absolutely critical. It really does need to be greatly strengthened, doesn't it? It is not just handling it in the lab, it is the fact that people doing biology, in pretty well any sphere anywhere in the world [need to be able] to access and make sense of quite sophisticated information."
Wellcome Trust Expert Group on Human Genetics, October 2009

"The establishment of database data resources are important in the transient world of genetics."
Wellcome Trust Expert Group on Human Genetics, October 2009

100. Despite the importance of sharing research findings across the world, the need to adopt rigorous governance policies and procedures is paramount to safeguard public and researcher confidence in human genetics research. The Wellcome Trust has historically played a leading part in developing strict and comprehensive guidelines to support its open access and data sharing principles. As the science and technology develop, it is important to continue to **retain sound governance** around this to ensure public and researcher confidence is maintained.

4.6 The opportunities and threats of personalised genomic medicine

"Already they [genomic companies] are beginning to build up cohorts and people are genuinely motivated, who have some disease in the family, to put themselves forward and pay for their genotype, and people are collecting cohorts for studies. The shape of the landscape is going to be set by different forces over the next few years, definitely some of them commercial, different sorts of establishments, but I think the research and the science agenda should be heard in that debate. Because it will affect the way that we can do science, but it is also about the use of science."
Wellcome Trust Expert Group on Human Genetics, October 2009

"500 000 people signing up to a company might provide a serious database on which research could be done and the Trust should be aware of that and willing to exploit that."
Wellcome Trust Expert Group on Human Genetics, October 2009

101. The Expert Group raised and discussed potential approaches to meet the challenges of personalised genomic medicine and proposed directions for future work, including:

- engaging with private genomic companies
- reinforcing public confidence in human genetic research.

102. Public interest in genomics and genetic susceptibility to certain conditions is now an integral part of the genetic landscape. Private genomic companies – such as 23andMe – servicing the public demand for access to personalised genetic information and risk profiles are beginning to build up very large cohorts, which are potentially a serious research resource. The Wellcome Trust Expert Group felt there was an enormous opportunity for funders and researchers to **engage with private companies** to explore the options for sharing genomic data.

103. Controlling the dissemination of genetic information and its uses by companies, often offering services without the back-up of adequate counselling, is a very important question but one primarily for those that control our health services. However, it is important for organisations such as the Wellcome Trust to advocate the provision of reliable sources of information – refuting any bogus claims from personal genomics companies – to help **reinforce public confidence in human genetics research**. The Expert Group emphasised the importance and value of public consultation and participation in the governance of genetic research, as beneficial medical applications of human genetic research are only realised if they have the full support of the public. Engaging with the public will prepare society for the emerging medical applications of human genetics research and will foster public confidence in medical genomics.

"The Trust needs to keep a very good watching brief on, along with other dispassionate sources of information. The only way we can see of dealing with this, in general, is to have trusted sources of information and try to direct people towards those."
Wellcome Trust Expert Group on Human Genetics, October 2009

4.7 Evolving partnerships with low-income countries

104. One final issue identified by the Expert Group concerned the Wellcome Trust's global health programmes. In light of the success of these programmes in recent years, the Expert Group felt that the Wellcome Trust should look for any opportunity to evolve partnerships with low-income countries, for the transfer of simple DNA technology for the control of both communicable and non-communicable disease.

105. The bibliometric analysis pointed to the changing map of international human genetics research. The Wellcome Trust has an opportunity to explore potential strategic partnerships – and given its ability and track-record in securing a seat at the influential tables, the timing might be right to do this.

Annex A: Methodology

1. As part of moves to strengthen the Wellcome Trust's evaluation activity, in 2008 the Assessment and Evaluation team developed an approach to review the impact of its funding at a subject, portfolio level. Finding the optimum way to review the impact of scientific research – and to use such information in strategic decision making – remains a significant challenge. In this portfolio review we do not attempt to provide an assessment of the impact of awards at an individual grant and/or funding scheme level but instead we take a macro, holistic view of the development of a field over time and consider the Wellcome Trust's role within this.

2. As described, the specific aims of the portfolio review were three-fold:
   - to identify the key landmarks and influences on the human genetics research landscape over the past two decades (1990–2009)
   - to consider what have been the key features of the Wellcome Trust's impact on this human genetics research landscape
   - to speculate on the future of human genetics and consider where there may be opportunities for Wellcome Trust strategy and funding.

3. One of the key challenges encountered in embarking on this project related to the definition of the portfolio of human genetics. It was decided that to attempt to assess the impact of the Wellcome Trust on 'genetics' as a whole would be impractical given the range and diversity of the field. Given the activity of the Wellcome Trust in much of the work surrounding the sequencing of the human genome, the field of human genetics might be a more useful focus for an initial portfolio review. Human genetics encompasses a wide variety of overlapping disciplines, including: population genomics and genetics of complex traits, cytogenetics, functional genomics, bioinformatics, statistical genetics, clinical genetics, genomic epidemiology and structural biology. As animal models set the stage for subsequent investigation of humans, where appropriate, experimental animal genetics research is included in the definition of the field of human genetics.

4. To provide some boundaries and restrict the scope of the analysis, for the purposes of this review, human genetics is defined broadly as

   "any aspect of genetics which has impacted on human biology or human medicine."

5. On this basis, over the 20-year timeframe (1990–2009), support for human genetics has accounted for around 10% of total Wellcome Trust funding – allocated broadly consistently across each year – accounting for £740m.[14]

6. In addition, in doing this review of a portfolio of funding – adopting a more macro approach to a review, looking at trends across the field and bringing subject experts into the heart of a review – we hope to test whether this approach is valuable to the Wellcome Trust and strategy more generally.

7. Finding the optimum way to review the impact of scientific research – and to use such information in strategic decision making – remains a major challenge. When attempting to assess the impact of a particular funding stream or funder over a substantial amount of time, we made the decision not to review on a micro, individual grants basis but instead to take a macro, holistic view of the development of the field over time and consider the role that the Wellcome Trust and its funding has played within the field. We know that the Wellcome Trust has committed a substantial amount of its funding for genetics – and specifically human-genetics-focused – research over a long period. Support for genetics remains a cornerstone of its research funding strategy.

8. Thus, the review involved three complementary streams of work:

   - landscape analysis
   - narrative case studies
   - an Expert Group.

14 Including funding to the Wellcome Trust Sanger Institute.

Landscape analysis

9. The landscape analysis had three components:

- Wellcome Trust funding analysis
- international scientific, policy and funding landscape analysis
- bibliometric analysis.

Wellcome Trust funding analysis

10. A search of the Wellcome Trust's AS400 Grants System was conducted to identify human-genetics-related funding provided by the Wellcome Trust from 1990 onwards. This involved three separate searches:

1. A search on specific search terms, as identified by Wellcome Trust Science Funding staff – capturing 6514 grants.
2. A more general search on 'Genetic*' or 'Genom*' – capturing a further 1793 grants (after de-duplication of grants following stage 1).
3. A search by Molecules Genes and Cells Committee codes – capturing a further 2281 (after de-duplication from stages 1 and 2).

11. A total of 10 588 genetics-related grants were identified. The full list of grants was then manually filtered by Science Funding staff to identify grants with a human genetics focus – resulting in a revised total of 1868 grants. After accounting for grant extensions and WTSI-based grants,[15] a total of 1172 individual grants with a human genetics focus have been funded by the Wellcome Trust over this time period. In terms of the total financial value of funding committed by the Wellcome Trust, supplements and extensions to grants were included. Due to limitations in the Wellcome Trust's grant subject classification systems, no attempt was made to further classify grant subjects within human genetics.[16]

International scientific, policy and funding landscape analysis

12. To provide context to the development of the field of human genetics research over the time period under investigation, a timeline of key events was produced by Wellcome Trust staff. The timeline contains key historical scientific advances and/or 'firsts' that have taken place or had a major impact on human-genetics-related research, focusing particularly on 1990 to the present day. The events contained on the timeline are coded into whether they describe a 'scientific advance', 'policy development' or 'funding development'.

13. The timeline was used as a prompt during the consultation with the Wellcome Trust Expert Group and their feedback on the content was invited; the revised version of the timeline is included in **Annex D.**

Bibliometric analysis

14. To identify key trends in the type, nature and location of human-genetic-related research over this period, an analysis of publication outputs was conducted. This bibliometric analysis was conducted by Evidence Ltd, part of Thomson Reuters (Scientific UK) and draws on the databases underlying the Web of Science, which include comprehensive coverage of more than 10 000 journals. All research papers included within the databases are allocated by Thomson Reuters to one or more of 253 different subject categories according to which journal the paper is published in. However, it is not possible to easily identify papers as concerning 'human genetics' from these journal subject categories alone, so a subject filter was created.[17]

15 Only a small minority of WTSI grants feature on the WT grant system. Sanger Institute data are therefore provided directly by WTSI unless otherwise stated.

16 Scoping work to refine and update the Trust's approach and system to 'subject classify' its grants is underway.

17 The filter was created by Wellcome Trust scientific staff in collaboration with Evidence Ltd.

Landscape analysis

15. Based on a set of agreed keywords (see **Table A**), a search of the Web of Science databases was conducted to identify papers that contained one or more of the keywords in either their title or their abstract. Papers that fell only into journal subject categories considered 'out of scope' for 'human genetics' were removed from the dataset; these included plant sciences (which was ranked seventh among the categories in the original search), geosciences, and astronomy and astrophysics. As such, while not a perfect fit for 'human genetics', the final dataset is thought to be a good proximate for the field and contained 1 356 417 human-genetics-related papers published over the period 1989–2008.[18]

16. To provide insight into the quality of papers emerging over the period, citation analysis was also conducted. To account for variation in citation practices across fields and the impact of publication date on the number of citations accumulated, Evidence Ltd rebase (or normalise) all raw citation data to the world average in the relevant subject field in the year of publication. In the context of this report, 'highly cited papers' refers to those papers with an average rebased impact of at least four (i.e. they have received at least four times as many citations as the average paper published in that year, in the same subject area).

Table A Search terms used for bibliometric analysis of human genetics

GENETIC*	EVOLUTION	HEREDITY
GENOM*	GENE	HETEROZYGOTE
ALLELE	GENES	HOMOZYGOTE
ANEUPLOIDY	GENOTYPE	MUTATION
CHROMOSOME	HAPLOIDY	POLYPLOIDY
DIPLOIDY	HAPLOTYPE	SNP
DNA		

17. With scientific research becoming an increasingly collaborative and multi-location activity, research papers are often linked to multiple authors based at more than one research institution. This report uses 'integer counting', meaning that any single paper counts as 'one' output for each author, institution and country contributing to its publication. For example, a paper with two authors from Harvard University, one author from the University of Texas and one author from the University of Oxford is counted as one output for each of the four authors concerned, one output for each of the three named institutions, and one output each for the USA and the UK. The tables of top countries, institutions and authors, therefore, will feature an element of double-counting.

18. Evidence Ltd are able to supply data on UK organisations with a high degree of accuracy as they ensure that organisational name variants are reconciled into one 'name' to counter the limitations of the raw Thomson Reuters address data on the papers themselves. For non-UK data, the tables in this report rely on the raw data and so paper numbers should be considered as indicative rather than absolute, although address reconciliation was effected for 'significant' research organisations. This methodology will have most impact where major organisations have several institutions or research centres (e.g. the UK MRC will be split into its constituent parts, while papers associated with the separate institutes of the CNRS in France, CSIS in Spain or the Chinese Academy of Sciences are all indexed under the name of the parent organisation).

18 Due to availability of citation data, the period of bibliometric analysis covers the period 1989–2008.

Narrative case studies

19. A series of case studies were compiled to tell the story of and highlight the key accomplishments associated with specific Wellcome Trust investments. These narratives were selected to reflect the range of funding types from major infrastructure, to fellowship support, to collaborations.

Wellcome Trust Expert Group on Human Genetics

20. To complement the landscape analyses and case study work, a key component of this portfolio review was the engagement of human genetics experts to provide the review of the key influences on the field and assessment of the Wellcome Trust's role within this. Instead of consulting experts on an individual basis, we tested whether we could usefully employ a cohort of independent subject experts in a group setting.

21. Witness Seminars[19]: In 1990 the Wellcome Trust created a History of Twentieth Century Medicine Group, as part of the Academic Unit of the Wellcome Institute for the history of medicine, to bring together clinicians, scientists, historians and others interested in contemporary medical history. Among several other initiatives, the format of Witness Seminars – to promote interaction between these different groups – was adopted. The Witness Seminar is a particularly specialised form of oral history, where several people (approximately 40) associated with a particular set of circumstances or events are invited to come together to discuss, debate, and agree or disagree about their memories. The Witness Seminar initiative is led by Professor Tilli Tansey, at the Wellcome Trust Centre for the History of Medicine, UCL. Further information about Witness Seminars can be found in volumes of *Wellcome Witnesses to Twentieth Century Medicine*.[20]

22. Drawing on the Witness Seminar approach to history and influence mapping, originally developed by the Institute of Contemporary British History in 1986, a small number of experts (n = 7, plus a Chair) were invited to debate and discuss the status of human genetic research 1990–2009 over the course of an afternoon. Existing and increasing pressures on experts generated by the peer review system embedded into science funding have made it traditionally (and increasingly) difficult to involve experts in post-award review, where arguably they could play a major part in helping to consider future funding allocation at a more strategic level. Experts were selected on the basis of their expertise in the field and, to ensure the maximum relevance to the review's aims, at least half of the experts had a fairly good knowledge of the Wellcome Trust over the period in question.

19 See 'Introduction' to recent volumes of *Wellcome Witnesses to Twentieth Century Medicine* published by the Wellcome Trust and the Wellcome Trust Centre for the History of Medicine at University College London.

20 http://www.ucl.ac.uk/histmed/publications/wellcome_witnesses_c20th_med

Wellcome Trust Expert Group on Human Genetics

23. The Expert Group received a summary of the landscaping and bibliometric analysis in advance of their meeting hosted at the Wellcome Trust in October 2009. Under the Chairmanship of Professor Martin Bobrow, the discussion was framed around the following broad question areas (see **Table B**). The meeting was recorded and the tapes transcribed. We found, alongside significant landscape analysis, that allowing our experts to be both retrospective and speculative in the same space allowed us to draw out learning from the past with future implications and also to consider how these might link to current and potential future funding strategies.

Table B Outline agenda – Expert Group on human genetics

Past/retrospective
Key impacts and influences on the field
What have been the key developments in the field? What are the origins of the key developments? What have been the key impacts within the field? What have been the key impacts beyond the field?
Role of the Wellcome Trust and key influences on the field
What has been the role of the Wellcome Trust throughout this period?
Present
Field progression
Has the field got as far as and/or gone in the direction you anticipated? What are the current limiting factors? Who is driving the agenda?
Future/prospective
Speculation and futures
What are the next big challenges in the field? What might the Wellcome Trust's role be? What strengths could the Wellcome Trust bring to the field?

Members of the Wellcome Trust Expert Group on Human Genetics

Chair
Professor Martin Bobrow

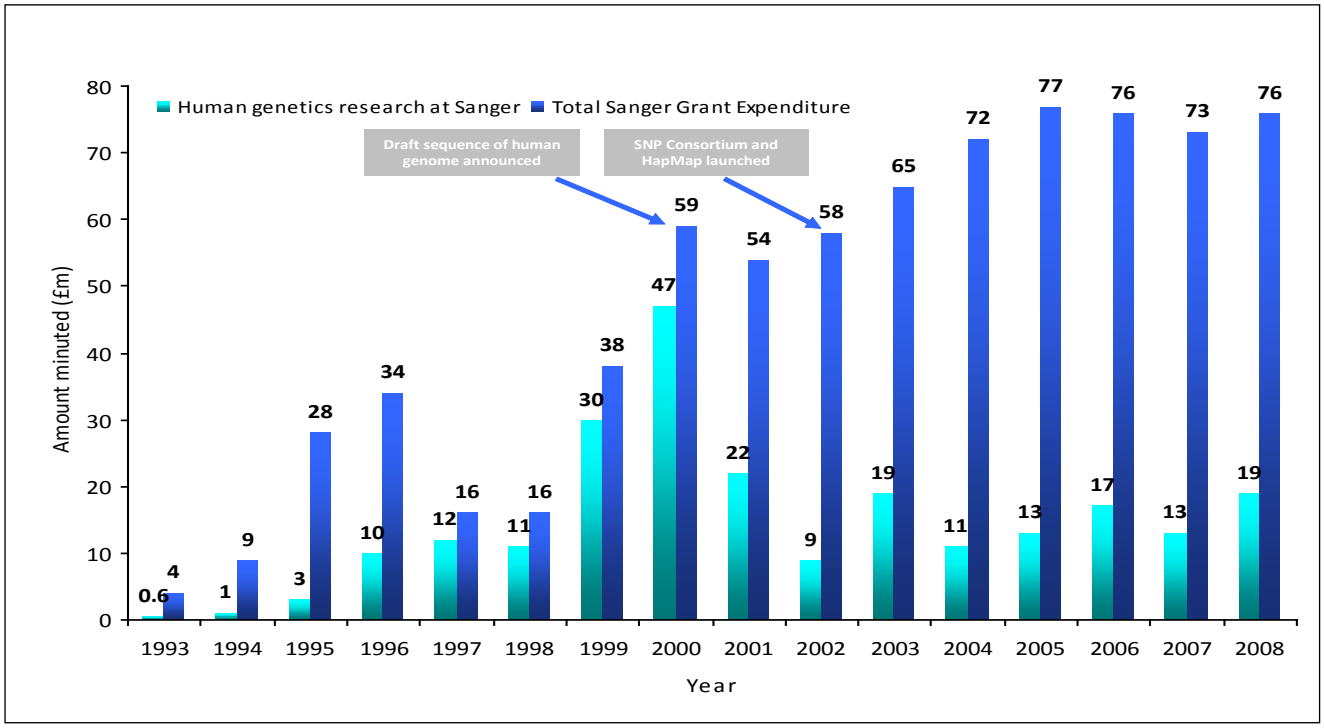
Members
Professor Dame Kay Davies
Dr Richard Durbin
Professor Marcus Pembrey
Professor Sue Povey
Sir John Sulston
Professor John Todd
Professor Sir David Weatherall

Annex B: Wellcome Trust funding for human genetics

24. Between the Wellcome Trust financial years 1989/1990 and 2007/2008, Wellcome Trust has awarded 1172 grants, mainly in responsive mode, to human-genetics-focused projects across all its funding divisions – accounting for £502m, representing just over 9% of the Trust's funding commitment over this time (excluding WTSI, see **Table 1**). Of this, approximately two-fifths (37% by number; 27% by value; 428 grants; £132m) of human genetics grant funding has been careers-based, supporting 398 individual researchers doing human-genetics-based projects via personal support schemes (**Table 1**). These personal support schemes include studentships (£9m), early career fellowships (£18m), intermediate (£26m) and senior/principal research fellowships (£79m).

25. The larger proportion of funds (64% by number, 74% by value; 744 grants; £370m) has been allocated to research and project support (equipment, university awards, strategic awards, buildings (Joint Infrastructure Funding, or JIF) and project and programme grants (**Table 1**). While funding for career support research grants in human genetics has remained relatively stable over the reporting period, funding for research, infrastructure and equipment-based projects has increased significantly – almost three-fold in the past decade (2000s; see **Figure 3**).

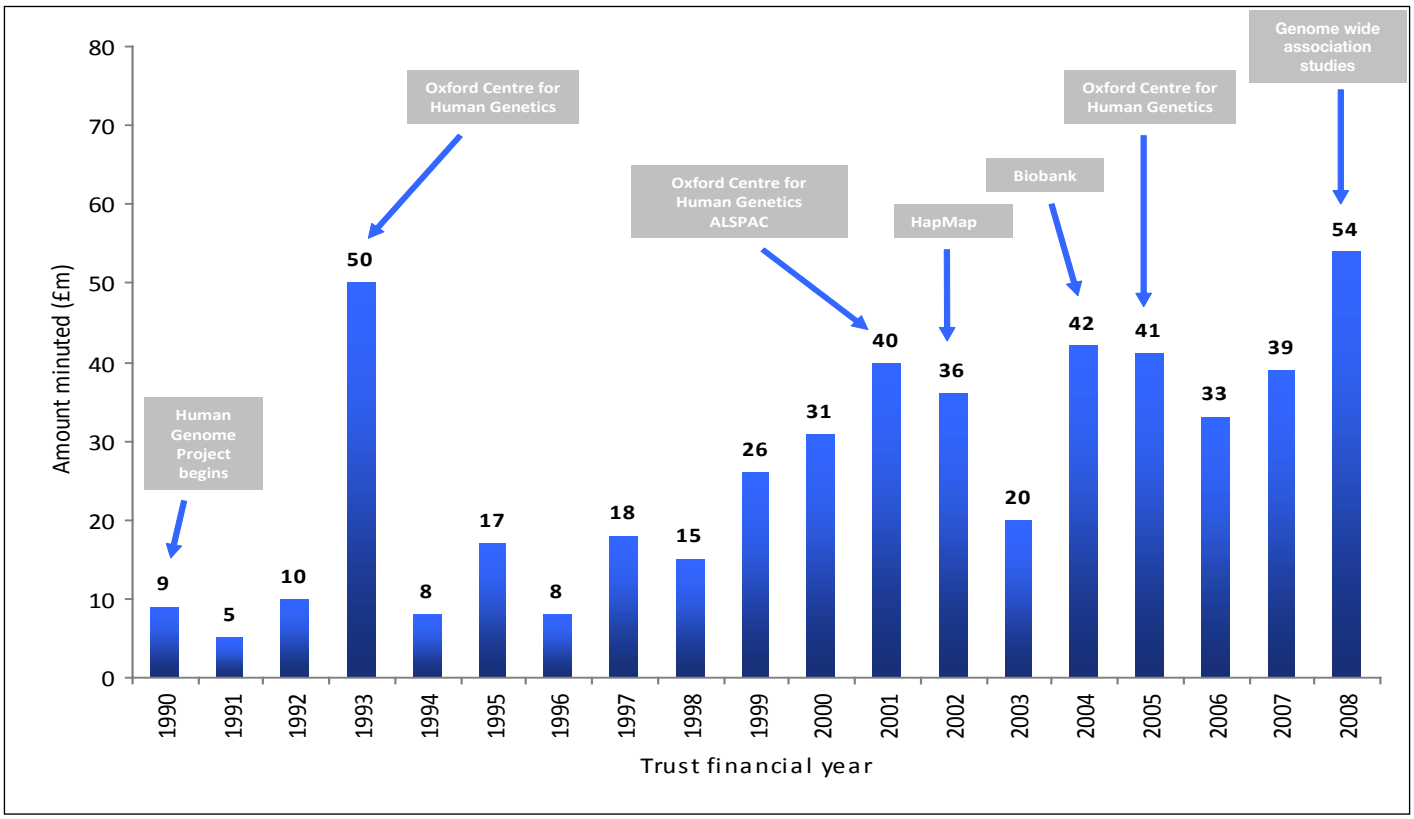
Figure 1 Wellcome Trust funding for human genetics research – 1990–2008 (£m)



NB All data is commitment by the Wellcome Trust; amounts are as awarded and have not been corrected for inflation. **Base:** 1172 Wellcome Trust grants associated with human genetics (renewals are counted as new grants in the Wellcome Trust grants system), excluding all funding to Wellcome Trust Sanger Institute.
Source: Wellcome Trust.

Annex B: Wellcome Trust funding for human genetics

Figure 2 Spend on human genetics by year at the Wellcome Trust Sanger Institute – 1993–2008 (£m)



Base: Wellcome Trust Sanger Institute spend (disbursement), excluding buildings and initial grant of £43m in 1992. **Source:** Wellcome Trust Sanger Institute.

Table 1 Wellcome Trust funding for human genetics by grant type 1990–2008[a]

Grant type	Number of grants	Per cent of total human genetics grants [b]	Amount (£m)	Per cent of total human genetics funding
Personal funding:				
Studentship	128	11%	9	2%
Early career fellowship	144	12%	18	4%
Intermediate fellowship	101	9%	26	5%
Senior/principal research fellowship	55	5%	79	16%
Total personal funding	**428**	**37%**	**132**	**27%**
Research funding:				
Project	420	36%	93	19%
Programme	52	4%	83	17%
JIF/SRIF	8	1%	42	8%
Strategic award	6	1%	25	5%
Equipment	55	5%	11	2%
University award	9	1%	3	1%
Other	194	17%	113	22%
Total research funding	**744**	**64%**	**370**	**74%**
TOTAL	**1172**		**502**	

[a] Excludes funding to the Wellcome Trust Sanger Institute – grant type data not available in comparable format; [b] percentages are rounded and may not equal 100%.
Base: 1172 WT human genetics grants. **Source:** Wellcome Trust AS400.

Annex B: Wellcome Trust funding for human genetics

Table 2 Wellcome Trust funding for non-UK-based human genetics research by grant type 1990–2008[a]

Grant type	Number of grants	Per cent of total human genetic grants	Amount (£m)	Per cent of total human genetics funding
Personal funding:				
Studentship	1	1%	0.08	1%
Early career fellowship	2	3%	0.3	2%
Intermediate fellowship	6	9%	1	7%
Senior/principal research fellowship	5	7%	3	23%
Total personal funding	**14**	**21%**	**4**	**32%**
Research funding:				
Project	11	16%	2	14%
Programme	1	1%	1	11%
Equipment	12	18%	2	13%
University award	1	1%	0.2	1%
Other	28	42%	4	29%
Total research funding	**53**	**79%**	**9**	**68%**
TOTAL	**67**		**13**	

[a] Excludes funding to the Wellcome Trust Sanger Institute – grant type data not available in comparable format; [b] percentages are rounded and may not equal 100%.
Base: 67 WT grants associated with human genetics overseas. **Source:** Wellcome Trust AS400.

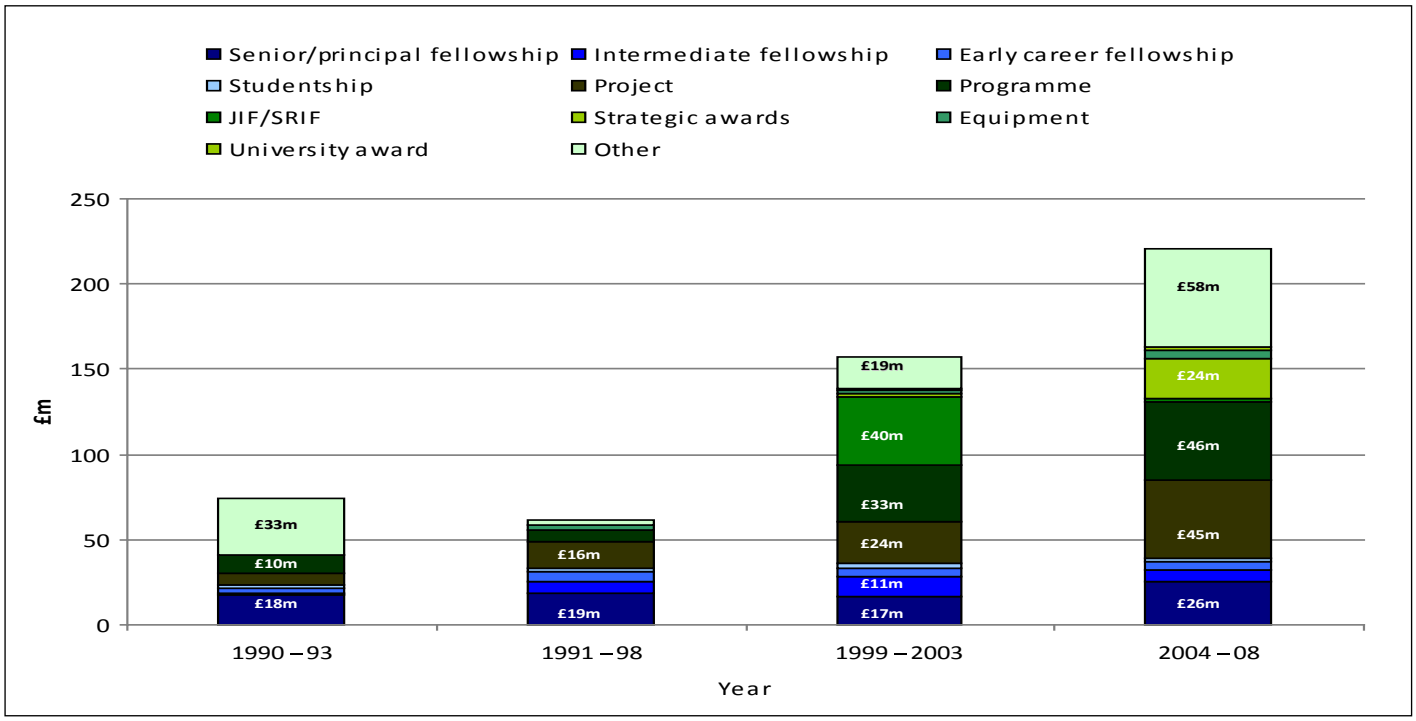
Table 3 Wellcome Trust (WT) and WTSI funding for human genetics – 1990–2008 [a, b, c]

Year	WT grant funding (commitment £m)[d]	WT human genetics [e] (Figure 4)		WTSI (disbursement £m)[b] (Figure 5)	WTSI human genetics (Figure 5)	
		(£m)	% of WT grant funding		(£m)[b]	% of WTSI total funding
1990	53	9	17%			
1991	60	5	8%			
1992	86	10	12%			
1993	437	50	11%	4	0.6	15%
1994	193	8	4%	9	1	11%
1995	198	17	9%	28	3	11%
1996	168	8	5%	34	10	29%
1997	222	18	8%	16	12	75%
1998	212	15	7%	16	11	69%
1999	354	26	7%	38	30	79%
2000	480	31	6%	59	47	80%
2001	388	40	10%	54	22	41%
2002	419	36	9%	58	9	16%
2003	395	20	5%	65	19	29%
2004	258	42	16%	72	11	15%
2005	344	41	12%	77	13	17%
2006	325	33	10%	76	17	22%
2007	359	39	11%	73	13	18%
2008	525	54	10%	76	19	25%
Total	**5476**	**502**	**9%**	**754**	**238**	**32%**

[a] Data from Wellcome Trust grants system is 'commitment'; data from WTSI is disbursement; [b] all WTSI data provided by Sanger Institute; [c] data rounded to nearest £m; [d] excludes all funding to the WTSI and WT running costs; [e] excludes data to WTSI. **Base:** 1172 WT human genetics grants. **Source:** Wellcome Trust AS400.

Annex B: Wellcome Trust funding for human genetics

Figure 3 Wellcome Trust funding for human genetics research by grant type – 1990–2008 (£m)



NB All data are 'commitment' by the Wellcome Trust; amounts are as awarded and not corrected for inflation; renewals are counted as new grants.
Base: 1172 Wellcome Trust human genetics grants, excluding all funding to Wellcome Trust Sanger Institute. **Source:** Wellcome Trust AS400.

Table 4 **Wellcome Trust Strategic Awards 2003–2008**

Amount awarded (£)	Year	Principal applicant	Institution of award	Grant title
1,250,812	2003	Prof Sydney Brenner	Babraham Institute	Massively parallel sequencing of populations of genomes.
5,882,186	2007	Prof Linda Partridge	University College London	Genomic and biochemical analysis of ageing and age-related disease.
4,968,596	2007	Prof Angus Lamond	University of Dundee	Centre for Gene Regulation and Expression.
2,020,414	2008	Prof Peter Donnelly	University of Oxford	Strategic Award for the Wellcome Trust Centre for Human Genetics.
5,757,592	2008	Dr Ewan Birney	Embl at WTSI	Trace Archive and 1KG DCC.
4,873,724	2008	Prof Janet Thornton	Embl at WTSI	Chemogenomics.

Source: Wellcome Trust AS400.

Table 5 **Wellcome Trust funding for human genetics research by funding division – 1990–2008**

Date	Science funding		Medicine, Society and History		Technology Transfer	
	No. grants	Amount awarded (£m)	No. grants	Amount awarded (£m)	No. grants	Amount awarded (£m)
1990–1993	210	79	6	0.1	0	
1994–1998	262	61	17	0.6	0	
1999–2003	324	146	82	5	2	2
2004–2008	210	202	57	5	2	1
Total	**1006**	**488**	**162**	**10**	**4**	**3**

Source: Wellcome Trust AS400.

Annex B: Wellcome Trust funding for human genetics

Table 6 MH&E and TTD grants by grant type

Grant type	1990–1993		1994–1998		1999–2003		2004–2008	
	No. grants	Amount awarded (£m)	No. grants	Amount awarded (£m)	No. grants	Amount awarded (£m)	No. grants	Amount awarded (£m)
Biomedical Ethics								
Project grants					19	2	4	0.6
Fellowships					7	0.7	4	0.3
Studentships					10	0.6	6	0.3
Other			1	0.002	17	0.1	11	0.4
Total	**0**	**0**	**1**	**0.002**	**53**	**4**	**24**	**1**
History of Medicine								
Project grants			1	0.003			1	0.09
Fellowships	2	0.09			1	0.1		
Studentship							1	0.06
Other	3	0.006	1	0.06	6	0.003	4	0.001
Total	**5**	**0.09**	**2**	**0.07**	**7**	**0.1**	**6**	**0.2**
Public Engagement								
Symposia	1	0.002	2	0.004	3	0.008		
Project grants			10	0.5	7	0.4		
Fellowships					1	0.1		
People Award					2	0.06	15	0.4
Society Activity Awards							6	0.8
Society Research Awards							1	0.1
Total	**1**	**0.002**	**12**	**0.6**	**13**	**0.5**	**22**	**1**
Other MH&E			2	0.001	9	0.5	4	2[a]
Total MH&E	**6**	**0.1**	**17**	**0.6**	**82**	**5**	**57**	**5**
Technology Transfer								
TTD Projects					1	1		
University Translation Award					1	0.6		
Strategic Translation Award							2	1
Total TTD					**2**	**2**	**2**	**1**

Base: 166 Wellcome Trust grants associated with human genetics in the MH&E and TTD funding divisions.

[a] Includes an MH&E Capital Award for £1.7m.

Table 7 Researchers in receipt of most Wellcome Trust (research and personal support) funding for human genetics 1990–2008 [a, b] (top ten)

Researcher	Current host institution	Amount (£m) [b]	Grant type
Prof John Todd	University of Cambridge	11	Senior research fellowship (x2); project (x5); principal research fellowship
Prof Peter Donnelly	University of Oxford	10	Strategic award; genome-wide association study funding
Prof William Cookson	Imperial College London	10	Senior research fellowship (x2); project (x3); programme; genome-wide association study funding
Prof Linda Partridge	University College London	9	Programme; strategic award
Prof Jonathan Flint	University of Oxford	8	Research training fellowship; clinician scientist fellowship; project (x3); senior research fellowship; biomedical research collaboration; programme
Prof Anthony Monaco	University of Oxford	7	Principal research fellowship (x2)
Prof Timothy Spector	King's College London	6	Programme (x3); project (x4)
Dr Ewan Birney	European Bioinformatics Institute	6	Prize studentship; strategic award
Prof Douglas Blackwood	University of Edinburgh	6	Programme
Prof George Davey Smith	University of Bristol	6	Programme; genome-wide association study funding

[a] Excludes funding to the Wellcome Trust Sanger Institute; [b] data rounded to the nearest £m. **Base:** 1172 WT human genetics grants. **Source:** Wellcome Trust AS400.

Annex B: Wellcome Trust funding for human genetics

Table 8 Institutions in receipt of most Wellcome Trust funding for human genetics 1990–2008 (top 20)

Institution	Amount (£m)[c]
Wellcome Trust Sanger Institute[a]	754
University of Oxford	157
University of Cambridge	45
University College London	38
Imperial College London	29
University of Edinburgh	27
UK Biobank	27
Cardiff University	22
European Bioinformatics Institute[b]	21
King's College London	18
University of Bristol	12
University of Newcastle	9
University of Dundee	9
Institute for Cancer Research	9
University of Glasgow	8
University of Leicester	7
University of Manchester	7
University of Leeds	6
University of Birmingham	5
University of Southampton	4

[a] Includes *all* funding to the Wellcome Trust Sanger Institute; [b] includes EBI extension (£8.6m); [c] data rounded to the nearest £m.
Base: 1172 WT human genetics grants. **Source:** Wellcome Trust AS400.

Table 9 Major human genetics research consortia and partnerships where the Wellcome Trust has played a major part (funded solely by the Wellcome Trust shaded)

Title	Established	Funding	Aim	Key accomplishments and discoveries (to end of 2009)
Human Genome Project (HGP)	1990	The HGP was a 13-year project coordinated by the US Department of Energy and the National Institutes of Health. In 1995, the Wellcome Trust pledged £12 million for seven years and in 1999, £110 million. Overall, the Wellcome Trust has contributed over £200m. Additional funding was provided by the governments of Japan, France, Germany, China and others.	To identify all of the genes in human DNA. To determine the sequences of the 3 billion chemical base pairs that make up human DNA. To store this information in databases and to develop tools for data analysis. Additional thought was also given to the ethical, legal, and social issues (ELSI) associated with the project and its findings.	• The revelation of microRNAs (miRNAs) (1993). • Completion of a working draft sequence of the human genome in 2000, published in *Nature* in 2001. • Open Access. • Completion of the "gold standard" human genome sequence in 2003, published in *Nature* in 2004. • Sequencing science is thought to have revolutionised basic biological research; the elucidation of the human genome, finally completed in 2003, is thought every bit as significant to basic biological science as the identification of the structure of DNA in 1953. • Identification of smaller number of protein-encoding genes in human genome – 20 000–25 000 genes. • Identification of copy number variations (CNVs) in human genome. • Collaboration with TSC enabled TSC to proceed at a much faster pace.
SNP Consortium (TSC)	1999	The Wellcome Trust contributed £9m to the SNP Consortium – a £30m collaboration involving the Wellcome Trust, leading academic centres (Whitehead Institute, Washington, University School of Medicine, St. Louis, and the Sanger Institute) and 13 pharmaceutical and technological companies. The international member companies are APBiotech, AstraZeneca Group PLC, Aventis, Bayer Group AG, Bristol-Myers Squibb Co., F. Hoffmann-La Roche, Glaxo Wellcome PLC, IBM, Motorola, Novartis AG, Pfizer Inc., Searle, and SmithKline Beecham PLC. Wellcome Trust funding for TSC comes from Wellcome Trust Core funding to WTSI (see **Annex B: Table 3**).	To identify and map the single-nucleotide differences in our DNA code – called single nucleotide polymorphisms, or SNPs.	• Identified and mapped 1.5m genetic variants. • Major driving force behind the creation of the public database (dbSNP)[1] that describes these genetic markers. • Produced first map of human genome sequence variation.[2] • Collaboration between the HGP and TSC demonstrated that public–private cooperation can be an efficient means for developing basic research tools essential for the application of genetic information to the understanding and treatment of diseases. • Influenced the creation of the International HapMap (see below), a project in which SNPs are used to understand how genetic variation contributes to health and disease. • Providing basis for the search for genes involved in common diseases, such as asthma, diabetes, cancer and heart disease, and have become essential to initiatives such as the Wellcome Trust Case Control Consortium.

Annex B: Wellcome Trust funding for human genetics

Title	Established	Funding	Aim	Key accomplishments and discoveries (to end of 2009)
Cancer Genome Project (CGP)	1999	The Cancer Genome Project was initiated by scientists at the Institute of Cancer Research: Professor Mike Stratton (now acting director of WTSI) and Dr Richard Wooster, in collaboration with the Wellcome Trust. The CGP is funded by the Wellcome Trust. In 1999, the Wellcome Trust awarded the CGP £10m over five years. Since 2005, the CGP have obtained in the region of £11m in external funding from a range of funding sources including ICR, Adenoid Cystic Carcinoma Research Fund, GSK, Key Kendall Leukaemia Fund, Breakthrough Breast Cancer and the European Commission. CGP also has additional 'out of envelope' funding from the Trust, which includes a senior clinical fellowship for Dr Peter Campbell, a Strategic Award for a drug screening project and a HICF award (currently under final financial negotiation), also for Dr Peter Campbell. These three awards total in the region of £17m. Several of these external projects are for ICGC pilot studies; the European Commission FP7 grant is for ER+ PR+ HER- breast cancer project. Mike Stratton and Andy Futreal plan to use the core funding for the remainder of this quinquennium and part of the next for an ICGC project on triple-negative breast cancer. **See also: International Cancer Genome Consortium (ICGC).**	To use the human genome sequence and high-throughput mutation detection techniques to identify somatically acquired sequence variants/ mutations and hence identify genes crucial in the development of human cancers.	• In 2002, discovery of mutations in the *BRAF*[3] gene linked to malignant melanoma. • Identified several other genes, including *ERBB2*[4], implicated in small-cell carcinoma of the lung. • In 2009, ten years after its inception, the Cancer Genome Project described the first comprehensive analysis of two complete cancer genomes – small-cell carcinoma of the lung and malignant melanoma – identifying almost all of the mutations associated with the two cancers.[5,6] • *BRAF* and *ERBB2* are now candidates for drug therapy. • Professor Marais – an author of the landmark *BRAF* 2002 paper – and Professor Caroline Springer at the Institute for Cancer Research (ICR) in London are now developing drug-like inhibitors of *BRAF*.[7]

Title	Established	Funding	Aim	Key accomplishments and discoveries (to end of 2009)
Juvenile Diabetes Research Foundation–Wellcome Trust Diabetes and Inflammation Laboratory (JDRF/WT DIL) located in the Cambridge Institute for Medical Research	2000	In 2000, the Wellcome Trust and the JDRF formed a partnership in order to leverage funds for research into type 1 diabetes. The Wellcome Trust[8] and the JDRF provided grants of around £10m each, over five years, to set up the JDRF/WT DIL under the leadership of Professor John Todd. In 2005 the DIL was renewed: to date, the total awarded to the DIL is approx £29m (including PRF awards), of which the Wellcome Trust contributed approx £17m.	The JDRF/WT DIL, centred in the Cambridge Institute for Medical Research, is a multi-disciplinary research programme within the University department of Medical Genetics. The aims of JDRF/WT DIL are to identify and characterise the effects of the susceptibility genes for type 1 diabetes to better understand the earliest events in human physiology that lead to autoimmune destruction of the insulin-producing beta-cells of pancreas.	• Created a world-class facility and multidisciplinary team to locate, identify and analyse the function of genes determining type 1 diabetes risk. • Professor John Todd (see **case study**), a Wellcome Principal Research Fellow, helped to establish the JDF/WT DIL. • The JDRF/WT DIL was centrally involved in the establishment and success of two major international consortia, the NIDDK- and JDRF-funded Type 1 Diabetes Genetics Consortium and the Wellcome Trust Case Control Consortium. Both collaborations drove forward the genetic analyses of type 1 diabetes. • In 2003 the JDRF teamed up with the Wellcome Trust in a multi-million pound effort to promote the UK's contribution to advancing research into stem cells. The two organisations are working together to promote and fund stem cell research. The development of stem cell lines and stem cell research is not only of primary importance to diabetes but could also lead to therapies and cures for other chronic illnesses. • Identified and analysed at least four new chromosome T1D susceptibility regions.[9] • Identified the HLA-B and HLA-A genes in the MHC as additional determinants of type 1 diabetes.[10] • Launched a new GWAS of JDRF/WT cases and controls with the Type 1 Diabetes Genetics Consortium (T1DGC).[11] • Detailed genetic mapping of the IL2RA region and correlation of IL2RA genotype with plasma concentration of the immune activation biomarker soluble IL-2RA/CD25.[12]

Annex B: Wellcome Trust funding for human genetics

Title	Established	Funding	Aim	Key accomplishments and discoveries (to end of 2009)
Mouse Genome Sequencing Consortium (MGSC)	2000	Members of the Mouse Sequencing Consortium (MSC) and their contributions to the effort are the Wellcome Trust ($7.75m), SmithKline Beecham ($6.5m), the Merck Genome Research Institute ($6.5m), Affymetrix, Inc. ($3.5m), and six of the National Institutes ($34m), including the National Cancer Institute, the National Human Genome Research Institute, the National Institute on Deafness and Other Communication Disorders, the National Institute of Diabetes and Digestive and Kidney Disease, the National Institute of Neurological Disorders and Stroke, and the National Institute of Mental Health.	To accelerate the sequencing of the mouse genome. In 2008 the WTSI committed to sequencing a further 17 mouse strains[13] used in biomedical research, with support from the UK MRC.	• In 2002, the WTSI, with funding from the Trust, had a major role in decoding the mouse genomes – contributing **25%** to the sequence work published in *Nature* by the MGSC.[14] • Produced the high-quality finished sequence of four mouse chromosomes. • Sequences released immediately through the Ensembl database,[15] jointly run by the European Bioinformatics Institute and WTSI, providing researchers with unrestricted access to the information. • The mouse genome sequence is being used to study genes involved in deafness,[16] cancer[17] and the immune system.[18]
International HapMap Project	2002	The HapMap Project[19] was launched by an international consortium of scientists and funding agencies and has been identifying and mapping sites of variation in the human genome. Funding for the $100m public–private effort to create the next generation map of the human genome was provided by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT); Genome Canada; Genome Quebec; the Chinese Academy of Sciences, the Chinese Ministry of Science and Technology, the Natural Science Foundation of China; and the US NIH. The SNP Consortium co-ordinated private funding, while the Wellcome Trust provided charitable funding through its core funding to WTSI (see **Annex B: Table 3**). The WTSI has been the main UK partner since its inception.	To compile data on patterns of association between SNPs in four different populations (African, Han Chinese, Japanese and North Americans of West and Northern Europe descent) and make them freely available in the public domain to identify genes of medical importance. The role of the UK research based at the WTSI, University of Oxford, was to analyse 24% of the genome. Professor Lon Cardon, a Wellcome Principal Research Fellow at the WTCHG, was key in the development and subsequent analysis of the HapMap projects (see **Lon Cardon case study**).	• HapMap Project data has been instrumental in the development of methods for the design and analysis of **GWAS**. • In 2005, the HapMap Project completed the first draft of the human **haplotype map**,[20] or HapMap, which involved the genotyping of 1.1 million SNPs across four populations, published in *Nature*.[21] • Published a free-to-access internet-based database of human genetic variation in October 2005.[1,2] • Identified a genetic variation that substantially increases the risk of age-related macular degeneration.[22] • Shed new light on the suggested role of the gene **dysbindin** (*DTNBP1*) in schizophrenia.[23] • Data samples served as a foundation for a major research collaboration, led by Matthew Hurles at the WTSI, on copy number variation.[24] • Phase II of the HapMap project delivered a further 2.1m SNP genotypes on the same population.[25]

Title	Established	Funding	Aim	Key accomplishments and discoveries (to end of 2009)
Structural Genomics Consortium (SGC)	2003	The SGC is funded in a distinctive way. Like TSC, the SGC is a not-for-profit organisation funded by public and private funders. In 2003, the Wellcome Trust contributed £18m to the three-year phase I budget (a total of £48m) to establish the SGC. GlaxoSmithKline contributed £3m, and another £19m came from four Canadian funding agencies: Genome Canada (through the Ontario Genomics Institute), the Canadian Institutes for Health Research, the Canada Foundation for Innovation and the Government of Ontario. In 2005, four Swedish sponsors joined the SGC: the Karolinska Institute, the Knut and Alice Wallenberg Foundation, the Swedish Foundation for Strategic Research, and the Swedish Governmental Agency for Innovation (VINNOVA). In 2007, funding for phase II was agreed (in excess of £50m over four years). Two additional pharmaceutical companies– Merck and Novartis – joined the consortium of funders for phase II.	To determine the three-dimensional structure of medically important proteins using a high-throughput approach and place them into the public domain, with no restrictions on use. The Consortium operates out of three sites – Oxford, UK; Toronto, Canada; and Stockholm, Sweden – each of which focuses on proteins of known or potential importance in different disease areas.	• The SGC[26] is a model for an open access public–private partnership. • Reached its phase I targets (386 human and parasite proteins) ahead of schedule and on budget. • Additional support was awarded in 2007 (C$105m) and two new partners joined the SGC, Novartis and Merck & Co., Inc., to support phase II. • Published 166 scientific articles from 2005 to 2009. More than 90% include outside collaborators. • Deposited 949 structures in the WorldWide Protein Data Bank (wwPDB).[27] • Contributes ~20% of global output of human structures each year and accounts for **15%** of the total structural coverage of human proteins. • Between 2005–2009 the SGC has been the dominant contributor to structural coverage of DrugBank protein targets. • SGC's open access data-sharing policy where all structures determined by the SGC are made immediately and freely available on the internet facilitated a trio of papers on the structure of a key epigenetic component in *Nature* in 2007. • SGC determined the 3D structure of a key protein component – UHRF1 – involved in enabling the 'epigenetic code' to be copied accurately from cell to cell.[28]

Annex B: Wellcome Trust funding for human genetics

Title	Established	Funding	Aim	Key accomplishments and discoveries (to end of 2009)
Wellcome Trust Case Control Consortium (WTCCC)	2005	One of the UK's largest academic collaborations to date, the WTCCC was funded solely by the Wellcome Trust and was launched with almost £9m of funding.[29] In January 2008, the Wellcome Trust awarded a further £30 million to support 27 new GWAS.	To bring together top research groups to examine genetic variation in common, complex diseases, including bipolar disorder, coronary artery disease, Crohn's disease, hypertension, rheumatoid arthritis, type 1 diabetes and type 2 diabetes.	• Pioneered the use of genome-wide association (GWA) studies to identify genetic variants in relation to common diseases, applying high-throughput genotyping techniques to screen entire genomes of large numbers of patients. • Identified more than 90 new disease-causing variants. • Detection of the first gene associated with obesity, the *FTO* gene.[30] • Identification of variants in three genes (IGF2BP2, *CDKAL1* and *CDKN2A/CDKN2B*) that predispose people to type 2 diabetes.[31] • In 2007 publishes largest genome-wide association study of common diseases in *Nature*.[32] • Identification of the first genetic link – the *PTPN2* gene – between Crohn's disease and type 2 diabetes. • Confirmation of the importance of autophagy, or 'self-eating', in the development of Crohn's disease. • Identification of four chromosome regions containing genes that can predispose to type 1 diabetes.[33] • Identification of six new genetic variants that increase the likelihood of developing coronary artery disease.[34] • Identification of the first gene, known as HMGA2, a common variant of which directly influences height.[35] • Identification of genetic risk variants implicated in coeliac disease, in the first ever GWAS study carried out for coeliac disease.[36] • Identification of six new genes that have a role in the development of type 2 diabetes, extending the total number of genes implicated in common forms of the disease to 16.[37] • Classification of more than thirty distinct susceptibility loci for Crohn's disease by genome-wide association.[38]

Title	Established	Funding	Aim	Key accomplishments and discoveries (to end of 2009)
1000 Genomes Project[39]	2008	This international research consortium receives major support from the Wellcome Trust Sanger Institute in Hinxton, England, the Beijing Genomics Institute, Shenzhen (BGI Shenzhen) in China and the National Human Genome Research Institute (NHGRI), part of the National Institutes of Health (NIH). In 2008, three sequencing companies joined the consortium: 454 Life Sciences, a Roche company, Branford, Conn.; Applied Biosystems, an Applera Corp. business, Foster City, Calif.; and Illumina Inc., San Diego. The Wellcome Trust provided US$50m support for the project; which is now led by Richard Durbin at the Sanger Institute alongside David Altshuler from the US Broad Institute. This funding came from Wellcome Trust Core funding to WTSI (see **Annex B: Table 3**).	To generate a detailed map of human genetic variation – a haplotype map – by analysing in depth the genomes of at least 1000 people. To provide insight into the genetic susceptibility to common disease. The 1000 Genomes project goes a major step beyond the HapMap Project by mapping, SNPs and also structural variants at a high resolution.	• During the pilot phase of the project the WTSI produced around 20% of the sequence data and around 10m new SNPs and genetic rearrangements have been discovered, as well as confirming a similar number of previously known variants. An initial publication is expected in 2010. • Data from the first three pilot studies have been deposited with the European Bioinformatics Institute and the National Center for Biotechnology Information (NCBI),[40] part of the US National Library of Medicine.
International Cancer Genome Consortium (ICGC)	2008	There are 12 fully funded ICGC projects at present supported by the following funding organisations from around the world, including the Wellcome Trust and Breakthrough Breast Cancer, UK, the National Health and Medical Research Council, Australia, the Ontario Institute for Cancer Research and the Ontario Ministry of Research and Innovation, Canada, the Chinese Cancer Genome Consortium, China, the European Commission FP7, the Institut National du Cancer, France, Federal Ministry of Education and Research and German Cancer Aid, Germany, Department of Biotechnology, Ministry of Science & Technology, India, University of Verona, Italian Ministry of Education, University and Research, Italy, RIKEN and the National Institute of Biomedical Innovation, Japan, Spanish Ministry of Science and innovation, Spain. Each project within the consortium has an estimated cost of US$20m. Wellcome Trust funding for the ICGC comes from the Wellcome Trust Core funding to WTSI (see **Annex B: Table 3**).	Building on the Cancer Genome Project,[41] and modelled on the HGP, the ICGC aims to generate comprehensive catalogues of genomic abnormalities (somatic mutations, abnormal expression of genes, epigenetic modifications) in tumours from 50 different cancer types and/or subtypes of global clinical and societal importance, making the data available to the research community as rapidly as possible, with minimal restrictions.	• ICGC[42] is a international confederation of members comprising research teams from Australia, Canada, China, France, India, Japan, Singapore, the UK and the USA.

Annex B: Wellcome Trust funding for human genetics

Title	Established	Funding	Aim	Key accomplishments and discoveries (to end of 2009)
CONVERGE: China Oxford and VCU Experimental Research on Genetic Epidemiology	2008	This five-year project is funded by a £1.4m grant from the Wellcome Trust.	A major study on the genetics of depression in a population of women in China. The project is a CONVERGE collaboration, between the University of Oxford, Hua Shan Hospitalat Fudan University, China, and the Virginia Commonwealth University (VCU), USA.	• Prof J Flint's group from the WTCHG initiated the major study COVERGE into the genetics of depression in China. • CONVERGE is one of the largest studies into depression ever carried out in China (http://www.well.ox.ac.uk/flint/china/).
UK10K	2010	Led by researchers at the WTSI, Wellcome has committed £10.4m to the project.	The project aims to sequence 4000 samples from the ALSPAC and TwinsUK cohorts (at a 6x coverage); directly associate this variation with known phenotypic traits in the sequenced samples; carry out exome sequencing in 6000 cases with extreme phenotypes and provide all these data to the wider community as a sequence variation resource.	• Using next-generation sequencing technology to sequence human DNA samples with medically important phenotype data to study the contribution of rare sequence variants in health and disease.
The Human Heredity and Health in Africa Project (H3 Africa)	2010	H3 Africa is a £25m partnership established by the NIH and the Wellcome Trust. Over the next five years, H3 Africa will receive at least $12m (approx £8m) from the Wellcome Trust and $5m (approx £3.4m) a year from the NIH, along with administrative and scientific support. The NIH provided $750 000 to kick-start the project.	To support population-based genetic studies in Africa, including research into common, non-communicable disorders such as heart disease and cancer, as well as infectious diseases such as malaria.	

1 http://www.ncbi.nlm.nih.gov/SNP

2 The International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001;409:928–933

3 Davies H et al. Mutations in the BRAF gene in human cancer. Nature 2004;417:949–954

4 Stephens et al. Lung cancer: intragenic ERBB2 kinase mutations in tumours. Nature 2004;431:525–526

5 Pleasance ED et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 2010;463(7278):191–6

6 Pleasance ED et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature 2010;463(7278):184–90

7 Ménard D et al. Novel potent BRAF inhibitors: towards 1nM compounds through optimisation of the central phenyl ring. J Med Chem;52:3881–91

8 As part of the DIL, Professor David Clayton and Dr Linda Wicker were both awarded Wellcome Trust PRFs, both were renewed in 2005.

9 Todd JA et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 2007;39(7):857–64

10 Nejentsev S et al. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. Nature 2007;450:887–892

11 https://www.t1dgc.org

12 Lowe C et al. Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. Nat Genet 2007;39(9):1074–1082

13 http://www.sanger.ac.uk/Projects/M_musculus/

14 Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. Nature 2002;420(6915):520–62

15 http://www.ensembl.org/Mus_musculus/Info/Index

16 Spiden S et al. The novel mouse mutation oblivion inactivates the PMCA2 pump and causes progressive hearing loss. PLoS Genetics 2008;4(10)

17 Uren AG et al. Large scale mutagenesis in p19ARF- and p53-deficient mice identifies cancer genes and their collaborative networks. Cell 2008;133(4):727–741

18 Rodriguez A et al. Requirement of bic/microRNA-155 for normal immune function. Science 2007;316:608–611

19 http://hapmap.org/index.html

20 http://hapmap.org/index.html

21 The International HapMap Consortium. A haplotype map of the human genome. Nature 2005;437:1299–1318 http://www.genome.gov/Pages/Research/DER/HapMap.pdf

22 Klein RJ et al. Complement factor H polymorphism in age-related macular degeneration. Science 2005;308(5720):385–9

23 Mutsuddi M et al. Analysis of high-resolution HapMap of DTNBP1 (Dysbindin) suggests no consistency between reported common variant associations and schizophrenia. Am J Hum Genet 2006;79(5):903–9

24 Redon et al. Global variation in copy number in the human genome. Nature 2006;444:444–454

25 The International HapMap Project. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449:851–861

26 http://www.thesgc.org/

27 http://www.wwpdb.org/

28 Avvakumov GV et al. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature 2008;455:822–825

29 Support for the sample collections was provided by the Medical Research Council, British Heart Foundation, Juvenile Diabetes Research Foundation, Diabetes UK, the Arthritis Research Campaign, the National Association for Colitis & Crohn's Disease and MDF and the Bipolar Organisation.

30 Frayling et al. A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. Science 2007, 316:889–94

31 Zeggini et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 2007;316:1336–41

32 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447(7145):661–678

33 Todd JA et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 2007;39(7):857–64

34 Samani NM et al. Genomewide association analysis of coronary artery disease. N Engl J Med 2007;357:443–53

35 Weedon MN et al. Common variant of HMGA2 is associated with adult and childhood height in the general population. Nat Genet 2007;39(7):1245–50

36 Van Heel DA et al. A genome wide association study for celiac disease identifies risk variants in the region harbouring IL2 and IL21. Nat Genet 2007;39(7):827–9

37 Zeggini E et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 2008;40:638–45

38 Barrett JC et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 2008;40:955–962

39 http://www.1000genomes.org/page.php

40 http://www.ncbi.nlm.nih.gov/

41 http://www.sanger.ac.uk/genetics/CGP/

42 http://www.icgc.org/

Annex B: Wellcome Trust funding for human genetics

Table 10 Major research resources and data with major funding and development support from the Wellcome Trust (funded solely by the Wellcome Trust shaded)

Title	Established	Funding	Aim	Key accomplishments (to end 2009)
Avon Longitudinal Study of Parents and Children (ALSPAC)	1991	In June 2001, a grant totalling £11m was awarded to the project by the Wellcome Trust (£7m), the Medical Research Council (MRC) and the University of Bristol. In late 2005, the Wellcome Trust and the MRC each awarded £4.47m to ALSPAC, supplemented by £5.3m from the University of Bristol. The Trust provided an additional £2m towards a DNA samples collection already part-funded by the MRC.	Long-term, large-scale cohort study of 14 000 children, and their parents, born in the Avon area of England between 1 April 1991 and 31 December 1992 (see **ALSPAC case study**).	• Observation that the risk of peanut allergies is increased by using skin creams containing peanut oil during pregnancy. • Use of paracetemol in late pregnancy correlates with the development of childhood asthma. • Correlation between visual development in the pre-school period and fish consumption of the mother in pregnancy and/or breast feeding (indicating a positive benefit of exposure to omega-3 fatty acids). • Association between low levels of sleep in early childhood and obesity at the age of seven. • Link between maternal anxiety in pregnancy and increased risk of behavioural problems in young children. • Significant contribution to the identification of the obesity-associated *FTO* gene, one of the first outputs from the WTCCC.
UK Biobank	2000	In 2002, the Wellcome Trust contributed £20m (total initial funding contributed from the Wellcome Trust is £62m). Wellcome Trust supports the Biobank in partnership with the MRC, the Department of Health, and the Scottish Executive and the Northwest Regional Development Agency.	To provide scientists with an insight into the genetic and environmental factors that are associated with common multi-factorial diseases. UK Biobank will collect samples and medical information for 500 000 UK citizens between the age of 40 and 69.	• Feasibility study in 2000/01. • Pilot phase in 2006 involving recruitment of 4000 participants from South Manchester, UK; led to stringent ethical and legal governance framework. • Starting in April 2007, a three-to-four-year recruitment phase is underway (with 35 centres in England, Scotland and Wales, each open for a six-month period) to recruit participants and collect data, to collect a representative sample of the UK's population. • By the end of 2007, 50 000 people had taken part. Recruitment reached 100 000 in April 2008, 200 000 in October 2008, 300 000 in May 2009 and 400 000 in November 2009.

Title	Established	Funding	Aim	Key accomplishments (to end 2009)
Ensembl genome browser[43]	2000	Ensembl is funded principally by the Wellcome Trust with additional funding from EMBL and NIH-NIAID. The Ensembl genome browser is a joint project between the European Bioinformatics Institute (EMBL-EBI) and the WTSI. The Wellcome Trust invested more than £20m to establish Ensembl.	To provide a free, automatic, high-quality annotated analysis of human and other genome sequences to the research community using a web-based browser. To provide researchers with new ways to explore complex biological interactions between species.	• The most widely used vertebrate annotation resource: allows access to human and other genome sequences, which are freely available for use by researchers worldwide. • The Ensembl genome browser supports a significant fraction of the world's biomedical research enterprise. • Ensembl is one of the top two most accessed resources provided through the Sanger Institute web site, the other being Pfam. • Ensembl Plants, Ensembl Fungi, Ensembl Bacteria, Ensembl Metazoa and Ensembl Protists provide an integrated portal to access fully sequenced genomes and related data for the rest of cellular life, including many organisms of medical and agricultural importance. • A very highly cited resource, Ensembl was ranked fourth (171 citations) and appeared with many Sanger Institute resources in the top ten most cited papers in the *Nucleic Acids Research* database issue (2008). • Decipher is built upon the Ensembl genome browser – see below.
HapMap	(see **Annex B: Table 9**)			
The 1000 Genomes Project	(see **Annex B: Table 9**)			

Annex B: Wellcome Trust funding for human genetics

Title	Established	Funding	Aim	Key accomplishments (to end 2009)
ENCODE – an encyclopaedia of DNA elements	2003	The Encode[44] initiative, a US$36m project led by the US National Human Genome Research Institute, is an international consortium involving scientists in government, industry and academia. The Wellcome Trust supports ENCODE through its Wellcome Trust core funding to WTSI (see **Annex B: Table 3**).	To identify and catalogue all functional elements in the human genome sequence, analyzing in detail 1% of the human genome.	• In 2007 the consortium published data[45] yielding unexpected findings in the genome – in particular that almost every base in the test 1% of the genome is transcribed into RNA. • In 2007 the WTSI was awarded $8.5m (£4.2m) by the US National Human Genome Research Institute (NHGRI) as part of a programme to expand the Encyclopedia of DNA Elements (ENCODE) project, which in its pilot phase yielded new insights into the organization and function of the human genome. • Dr Tim Hubbard from the WTSI and Dr Ewan Birney from the European Bioinformatics Institute (EBI)[46] are leading the initiative into its second phase of analysis. Participant enrolment is scheduled to continue until mid-2010.
DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources)[47]	2004	The Wellcome Trust supports DECIPHER through its core funding to WTSI (see **Annex B: Table 3**)	Drawing on the Ensemble infrastructure, the DECIPHER database aims to: (i) increase medical and scientific knowledge about chromosomal microdeletions and duplications; (ii) improve medical care and genetic advice for individuals or families with submicroscopic chromosomal imbalance; and (iii) facilitate research into the study of genes that affect human health.	• Set up by Nigel Carter of the WTSI and Helen Firth, a clinical genetics consultant at Addenbrookes's hospital in Cambridge. • DECIPHER is built upon the Ensembl genome browser. It is the only open-access, web-based interactive database of its type, although data from other databases are available. • Enables clinicians across the world to share information about their patients to help unravel the basis of human genomic disorders. • Plays a pivotal part in looking at how CNVs affect human health and helps reveal and define new syndromes: a new syndrome linked to loss of part of chromosome 14 has been identified.[48] • 2000 patient cases have been contributed to the DECIPHER database since its inception: its diagnostic power strengthens as new cases are added.

43 http://www.ensembl.org/

44 http://www.genome.gov/10005107

45 ENCODE Project Consortium et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007;447(7146):700–816

46 http://www.ebi.ac.uk/

47 https://decipher.sanger.ac.uk

48 Firth HV et al. DECIPHER: Database of chromosomal imbalance and phenotype in humans using Ensembl resources. Am J Hum Genet 2009;84(4):524–533

Table 11 Key human genetics discoveries with major input from the Wellcome Trust

Discovery	Key date	Funding (including WT grant type/s where appropriate)	Aim/Summary	Key accomplishments/implications/ applications (to end 2009)
Sequencing the genomes of significant human pathogens	From 1995	Funded primarily by the Wellcome Trust and via the Wellcome Trust Beowulf Genomics Panel; now directly via WTSI through the Pathogen Genomics Programme. http://www.sanger.ac.uk/Projects/Pathogens/	The Pathogen Genomics Programme at the WTSI aims to sequence the genomes of organisms relevant to human and animal health.	Wide range of pathogens sequenced, including: • *Mycobacterium tuberculosis* (1998) • *Plasmodium falciparum* (2002) • the 'superbug' MRSA (methicillin-resistant *Staphylococcus aureus*) (2004) • *Entamoeba histolytica* (amoebic dysentry) (2005) • *Clostridium difficile* (2006) • three Leishmania species: *L. major* (2005), *L. braziliensis* (2007) and *L. infantum* (2007) • *Schistosomiasis mansoni* (2009).
Sequencing of the genome of *Saccharomyces cerevisiae* (baker's yeast)	1996	Involved a collaboration of more than 100 laboratories, including WTSI.	To sequence a large (more than 12 million base pairs of DNA) genome – of an organism with cells with similarities to those of humans.	• 17% of genome sequenced at the WTSI – the largest contribution from a single centre to the sequencing project. • First elucidation of the first complete genome sequence of a eukaryote[49]; yeast genome used as a model for understanding the basic functions of human cells. • Continues to provide benchmarks and reference points during HGP and other genome sequencing.
Sequencing of the first animal genome: the nematode worm *Caenorhabditis elegans*	1998	Led by John Sulston of the WTSI and Bob Sulston Washington University, St Louis.	To sequence the first multicellular organism.	• *C. elegans* was the first multicellular organism to be sequenced.[50] • John Sulston and Bob Waterson received one of the first HGP grants to embark on sequencing the nematode *C. elegans.* • This endeavour was instrumental to the decision by the Wellcome Trust to establish the WTSI. • The data sharing ethos among the *C. elegans* community, led by John Sulston and Bob Waterson, triggered the culture of data sharing that was fostered during the HGP. • This work became an important reference tool for the HGP. • In 2002 Sir John Sulston won the Nobel prize for Physiology and Medicine for his role sequencing the nematode worm.

Annex B: Wellcome Trust funding for human genetics

Discovery	Key date	Funding (including WT grant type/s where appropriate)	Aim/Summary	Key accomplishments/implications/ applications (to end 2009)
Decoding the first complete sequence of a human chromosome – number 22	1999	This work was supported by grants from the Wellcome Trust, NIH National Human Genome Research Institute, NSF, University of Oklahoma, Human Genome Sequencing Project of Japan Science and Technology Corporation, Japan Society for the Promotion of Science, the UK MRC, the Medical Research, Council of Canada to H.E.M., Swedish Cancer Foundation, Swedish Medical Research Council, and Swedish Cancer Foundation.	To decipher the first complete sequence of a human chromosome (22).	• Chromosome 22 was sequenced at the WTSI[51] as part of the HGP. • Chromosome 22 was the first of 23 human chromosome pairs to be deciphered. • Included the largest continuous sequence determined from any organism at the time (23 million bps). • Enabled scientists, for the first time, to view the entire DNA and the organization of a chromosome at the base pair level. • Reinforced and increased confidence in the HGP to complete a 'working draft' of the DNA sequence of the human genome by 2000.
Elucidation of Human Genome Sequence (see Table 9)	Working draft 2001; complete 'gold standard' 2003	See **Table 9**.		• Most widely used vertebrate annotation resource. • Identification of smaller number of protein-encoding genes in human genome – 20 000–25 000 genes. • The revelation of microRNAs (miRNAs). • Identification of copy number variations (CNVs) in human genome.
Micro RNAs (miRNAs)	2003	Arising from the HGP – involved researchers at WTSI and EBI	Speculated that the discovery of miRNAs – small non-coding RNA molecules approximately 22 nucleotides in length – is perhaps the most exciting finding from the HGP. The discovery of vast stretches of non-coding regions in human DNA, previously considered to be 'junk DNA' serving little or no purpose, are now thought to play an important part in gene regulation and may be responsible for much of the complexity of the higher eukaryotes.	• In 2007 the consortium published data[52] yielding unexpected findings in the genome, in particular that almost every base in the test 1% of the genome is transcribed into RNA. • Dr Tim Hubbard from the WTSI and Dr Ewan Birney from the European Bioinformatics Institute (EBI)[53] are leading the initiative into its second phase of analysis. • By the end of 2009, approximately 700 miRNA's had been identified in the human genome, with a total of over 800 predicted to exist (see http://microrna.sanger.ac.uk/cgi-bin/sequences/browse.pl). Manipulating their activity may present a new way to tackle disease, especially as they have been implicated in several types of cancer.

Discovery	Key date	Funding (including WT grant type/s where appropriate)	Aim/Summary	Key accomplishments/implications/ applications (to end 2009)
Copy number variation (CNVs)	2006/07	Identified following HGP; Matt Hurles and colleagues at WTSI did much to elucidate and understand CNVs.	Identification of global variation in copy number across all human chromosomes.[54]	• CNVs are known to affect phenotype by altering how many copies of certain genes are in the genome, by disrupting the coding sequences of genes and by upsetting the regulation of genes. • Early work revealed that there is little overlap between SNPs and CNVs, indicating that both types of variation should be examined in association studies (Stranger et al, 2007).
Identification of complex disease loci in human genome (via WTCCC)	2007	Led by the WTCCC; much of the analysis involved researchers at WTSI and researchers involved in the ALSPAC.	The largest GWAS of common diseases conducted by the WTCCC in 2007, identified several novel complex disease susceptibility loci in diseases including diabetes (type 1 and type 2), bipolar disorder, rheumatoid arthritis, Crohn's disease, hypertension and coronary heart disease.	• See **WTCCC in Table 9**. • Following the success of this WTCCC, the Wellcome Trust Sanger Institute proposed and planned a second wave of GWA studies – the WTCCC2 project – which is funded by the Wellcome Trust to analyse a further 120 000 samples and 13 more conditions, including multiple sclerosis, pre-eclampsia and Parkinsons disease.
***FTO* gene**	2007		Detection of the first gene associated with obesity; the *FTO* gene.[55]	• Discovery of the *FTO* gene has opened up an entirely novel area of research – the genetics of obesity (see ***FTO*** case study). • This discovery was made by the WTCCC in collaboration with researchers leading the ALSPAC.
Identification of several SNPs, demonstrating association with tuberculosis		Led by Prof Adrian Hill at the WTCHG.		• Prof Adrian Hill's Group finalised a GWAS of TB susceptibility as part of the WTCCC. • Identified several SNPs, in three unlinked members of the *CADM* gene family, not previously implicated in tuberculosis susceptibility, demonstrated association with tuberculosis.[56]
Identification of the first 20 loci for height, the second obesity locus and type 2 diabetes loci	2007, 2008	Led by Dr C Lindgren's group at the WTCHG.		• Dr C Lindgren's group identified the first 20 loci for height, the second obesity locus and the second tier of six previously unknown type 2 diabetes loci and participated in ten other studies.[57–59]
Identification of the role of *Pcsk5*	2008	Led by the Bhattacharya group at the WTCHG.		• Identified the roles for *Pcsk5* in cardiac development, body patterning and VACTERL syndrome.[60]

Annex B: Wellcome Trust funding for human genetics

Discovery	Key date	Funding (including WT grant type/s where appropriate)	Aim/Summary	Key accomplishments/implications/ applications (to end 2009)
Identification of a degenerate 13bp sequence motif strongly associated with recombination hotspots in humans	2008	Led by Prof P Donnelly at the WTCHG in collaboration with McVean and Myers.	First evidence that a substantial fraction of human recombination hotspots share a common mechanism. Increased hotspot resolution afforded by the phase II HapMap and novel search methods were used to further the understanding of the relationship between DNA sequence and hotspot location.	• Described a degenerate 13bp sequence motif strongly associated with recombination hotspots in humans.[61] The motif appears to be responsible for ~40% of human hotspots and is also associated with other forms of genome instability, including deletions in several classical genetic syndromes, microsatellite instability, and common mtDNA deletion. • Provided further insight into the process of human recombination and evolution. • Prof P Donnelly continues to lead the WTCCC2 consortia.
Elucidation of the role of the dyslexia KIAA0319 susceptibility gene	2008	Led by Prof A Monaco's group at the WTCHG.		• Showed that the dyslexia susceptibility gene *KIAA0319* influences reading skills in the general population.[62] • Demonstrated that the gene *KIAA0319* is involved in cell–cell interactions and signalling.[63]
Deciphering of 2 cancer genomes: small-cell carcinoma of the lung and malignant melanoma	2009	Through the Cancer Genome Project at WTSI.	First comprehensive analysis of two complete cancer genomes – small-cell carcinoma of the lung and malignant melanoma.	• In 2009, ten years after its inception, the Cancer Genome Project described the first comprehensive analysis of two complete cancer genomes – small-cell carcinoma of the lung and malignant melanoma – identifying almost all of the mutations associated with the two cancers.[64,65] • *BRAF* and *ERBB2* are now candidates for drug therapy.

49 Goffeau A et al. Life with 6000 genes. Science 1996;274:563–567

50 The C. elegans Sequencing Consortium. Genome sequence of the nematode *Caenorhabditis elegans*. A platform for investigating biology. Science 1998;282:2012–2018

51 Dunham I et al. The DNA sequence of Chromosome 22. Nature 1999;402(6761):489–95

52 ENCODE Project Consortium et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature;447(7146):700–816

53 http://www.ebi.ac.uk/

54 Redon R et al. Global variation in copy number in the human genome. Nature 2006;444(7118):444–54

55 Frayling TM et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 2007;316(5826):889–94

56 Tosh K et al. Variants in the SP110 gene are associated with genetic susceptibility to tuberculosis in West Africa. PNAS 2006;103(27):10364–10368

57 Weedon MN et al. Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet 2008;40:575–583

58 Zeggini et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 2008;40:638–645

59 Loos RJ et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. Nat Genet 2008;40(6):768–775

60 Szumska D et al. VACTERL/caudal regression/Currarino syndrome-like malformations in mice with mutation in the proprotein convertase Pcsk5. Genes Dev 2008;22:1465–1477

61 Myers S et al. A common sequence motif associated with recombination hot spots and genome instability in humans. Nat Genet 2008;40(9):1124–1129

62 Paracchini S et al. Association of the KIAA0319 dyslexia susceptibility gene with reading skills in the general population. American J Psychiatry 2008;165:1576–1584

63 Velayos-Baeza A et al. The dyslexia-associated gene KIAA0319 encodes highly N- and O-glycosylated plasma membrane and secreted isoforms. Human Molecular Genetics 2008;17(6):859–71

64 Pleasance ED et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 2010;463(7278):191–6

65 Pleasance ED et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature 2010;463(7278):184–90

Annex C: Wellcome Trust bibliometric analysis for human genetics: 1989–2008

26. Bibliometric analysis was conducted by Evidence Ltd, part of Thomson Reuters (Scientific UK). The final dataset contained 1 356 417 human-genetics-linked papers published over the period 1989–2008. See **Annex A: Methodology** for a detailed description of the methods used.

Table 1 Countries (top 20) showing the fastest growth in human-genetics-focused papers

Country	Number of 'human genetics' publications per year			
	1994–98	Change[a]	2004–08	Change[b]
China	3156	198%	31,932	171%
India	3661	99%	11,320	90%
South Korea	2277	917%	11,968	88%
Brazil	2917	178%	10,895	76%
Spain	8073	171%	19,907	43%
Australia	9008	121%	16,383	29%
Canada	15,274	93%	24,537	29%
Italy	13,221	151%	24,026	28%
Netherlands	8807	110%	14,555	27%
Denmark	3381	128%	6027	27%
Belgium	4595	142%	7971	25%
Israel	3544	104%	5952	19%
Switzerland	6319	137%	10,064	18%
Germany	24,498	144%	41,325	18%
USA	130,185	93%	184,216	15%
Sweden	7168	118%	11,081	14%
UK	29,555	113 %	45,797	14%
France	24,560	115%	33,176	7%
Japan	28,239	139%	40,305	3%
Russia	6310	514%	7228	1.72%

Source: Thomson Reuters 2009; Analysis: Evidence, Thomson Reuters (Scientific UK).

[a] Per cent change 1989–93 and 1994–98 (ten years); [b] per cent change 1999–2003 and 2004–08 (ten years).

Annex C: Wellcome Trust bibliometric analysis for human genetics: 1989–2008

Table 2 Subjects showing the fastest growth in 'human genetics' publications

Subject field	Number of publications per year				
	1994–98	Change[a]	2004–08	Change[b]	Change[c]
Clinical neurology	3296	231%	7460	25%	649%
Neurosciences	4945	230%	9159	29%	511%
Infectious diseases	3384	266%	5331	9%	477%
Gastroenterology and hepatology	3245	176%	5754	18%	390%
Pharmacology and pharmacy	6072	130%	12,727	36%	383%
Res and exptl medicine	4307	140%	7257	18%	305%
Cell biology	4694	106%	8829	62%	287%
Biotech and appl microbiol	13,805	144%	19,814	28%	250%
Endocrinology and metabolism	5999	134%	8745	11%	241%
Haematology	4239	137%	5937	-1%	232%
Biology	3928	69%	7064	33%	203%
Zoology	4410	79%	7005	18%	185%
Microbiology	11,577	73%	18,874	30%	181%
Oncology	17,296	126%	21,015	8%	174%
Biophysics	8907	141%	9388	-10%	154%
Virology	7508	91%	9678	18%	147%
Immunology	7226	87%	8333	1%	116%
Biochem and mol biol	33,242	70%	41,551	9%	113%
Genetics and heredity	25,954	67%	32,006	14%	106%
General and internal medicine	4410	68%	5352	-2%	103%

Source: Thomson Reuters 2009; Analysis: Evidence, Thomson Reuters (Scientific UK).

[a] Per cent change 1989–93 and 1994–98 (ten years); [b] per cent change 1999–03 and 2004–08 (ten years); [c] per cent change 1989–93 and 2004–08 (20 years).

Table 3 Organisations producing highly cited papers[a] in 'human genetics', 1989–2008 – worldwide

World organisation (ranked by highly cited papers over 2004–2008 period)	Highly cited papers			
	1989–1993	1994–1998	1999–2003	2004–2008
Harvard University	192	311	306	559
University of Texas	90	171	154	236
MIT	89	87	106	218
Johns Hopkins University	97	176	132	207
Stanford University	62	111	152	193
University of Washington, Seattle	52	109	132	191
University of Cambridge	**29**	**60**	**79**	**184**
University of Oxford	**26**	**71**	**108**	**182**
University of California San Diego	53	89	114	163
University of Pennsylvania	36	91	78	161
Washington University, St Louis	48	104	79	156
University of California Berkeley	69	90	126	151
Yale University	60	89	84	146
University of California Los Angeles	43	79	89	144
University of California San Francisco	88	116	90	141
US National Cancer Institutes	97	115	118	140
Massachusetts General Hospital	61	84	68	140
University of Michigan	66	77	77	134
Duke University	59	88	64	128
University of North Carolina	32	67	52	127
University College London	**18**	**62**	**53**	**126**
University of Toronto	55	65	75	125
Wellcome Trust Sanger Institute	**0**	**12**	**40**	**122**
Imperial College London	**26**	**40**	**63**	**117**
Brigham and Womens Hospital	28	53	65	115
Cornell University	57	75	57	108
CNRS (Ctr Natl Recherche Sci)	27	55	49	107
University of Chicago	37	64	58	104
University of Minnesota	35	68	44	103
Karolinska Institutet	24	47	35	103

Source: Thomson Reuters 2009; Analysis: Evidence, Thomson Reuters (Scientific UK) [a] Highly cited papers refers to those papers with an average rebased impact of at least four (i.e. they have received at least four times as many citations as the average paper published in that year, in the same subject area). See **Annex A** for further detail.
Base: 30 organisations producing the most highly cited papers in 'human genetics' in the 2004–2008 period.

Annex C: Wellcome Trust bibliometric analysis for human genetics: 1989–2008

Table 4 Percentage of total papers which are highly cited[a] – for each five-year period and for selected organisations 2004–2008

Time period	Total global output of 'human genetics' papers	Output of highly cited 'human genetics' papers	Percentage of total output which is highly cited
1989–1993	151,261	3276	2.2%
1994–1998	306,105	5246	1.7%
1999–2003	405,401	5159	1.3%
2004–2008	493,650	8116	1.6%
For the 2004–2008 period:			
Harvard University	9482	559	5.9%
University of Texas	8158	236	2.9%
University of Oxford	3795	182	4.8%
University of Cambridge	3588	184	5.1%
University College London	3394	126	3.7%
Imperial College London	2830	117	4.1%
Wellcome Trust Sanger Institute	911	122	13.4%

Source: Thomson Reuters 2009; Analysis: Evidence, Thomson Reuters (Scientific UK).
Note: Organisations selected are the top two international producers, and all UK institutions featured in the top 30 producers of highly cited 'human genetics' papers in the 2004–2008 period. [a] Highly cited papers refers to those papers with an average rebased impact of at least four (i.e. they have received at least four times as many citations as the average paper published in that year, in the same subject area). See **Annex A** for further detail.

Annex D: Human genetics research timeline

27. The Timeline sets out the key scientific, policy and funding developments that have influenced the field of human genetics in recent history. In the past two decades the Wellcome Trust has been associated with many of the key developments and breakthroughs in the field, typically through its role in supporting international consortia, collaborations and partnerships.

Key landmarks in human genetics (Wellcome Trust-associated post-1990 shaded)

Date	Key	Summary	Description	People and place
1859	Scientific advance	*On The Origin of Species* is published	Charles Darwin sets out the theory of natural selection.	Charles Darwin, UK
1866	Scientific advance	Mendel publishes his work on inheritance	Mendel's work to establish the laws of inheritance remained largely unrecognised until 1900. Mendel and Darwin's theories were originally considered incompatible.	Gregor Mendel, Austria-Hungary
1869	Scientific advance	DNA is discovered	DNA is isolated from white blood cells, ultimately paving the way for its identification as the carrier of inheritance.	Friedrich Miescher, University of Tübingen, Germany
1900	Scientific advance	Human blood groups described	Austrian scientist Karl Landsteiner described three different blood types in 1900 (it emerged later that Czech serologist Jan Janský had independently classified human blood into four groups).	Karl Landsteiner, Austria
1905	Scientific advance	The term "genetics" is coined	Bateson, a proponent of Mendel, popularized the usage of the word "genetics" to describe the study of inheritance.	William Bateson, University of Cambridge, UK
1908	Scientific advance	Recessive inheritance of enzyme defects	Garrod formulated the "one gene, one enzyme" hypothesis and described the nature of recessive inheritance of most enzyme defects.	Archibald Garrod, London, UK
1911	Scientific advance	ABO blood group inheritance discovered	Ludwik Hirszfeld and E von Dungern discovered the heritability of ABO blood groups in 1910–11.	Ludwik Hirszfeld and E. von Dungern (Heidelberg, Germany)
1915	Scientific advance	Hunt Morgan describes linkage and recombination	Over several years, Thomas Hunt Morgan characterises the chromosomal basis of heredity, sex determination and recombination.	Thomas Hunt Morgan, Colombia University, USA
1924	Scientific advance	Blood group inheritance characterised	Felix Bernstein demonstrates the correct blood group inheritance pattern of multiple alleles at one locus in 1924.	Felix Bernstein, University of Göttingen
1928	Scientific advance	"Transforming principle" theory	Fred Griffith proposes that a "principle" had transformed a harmless bacterial strain into a harmful strain.	Fred Griffith, UK
1944	Scientific advance	The "transforming principle" is developed	First shown that DNA carries genetic information.	Oswald Avery, Maclyn McCarty and Colin McLeod , Rockefeller Institute, USA
1951	Scientific advance	First autosomal linkage identitifed	Jan Mohr in Copenhagen shows autosomal linkage between the Lutheran and Secretor locii; the first autosomal linkage identified.	Mohr et al (Copenhagan, Denmark)
1952	Scientific advance	Hershey-Chase Experiment	The Hershey-Chase experiments identified DNA as the genetic information of phages (Hershey shared the 1969 Nobel Prize in Physiology or Medicine for his "discoveries concerning the genetic structure of viruses").	Alfred Hershey and Martha Chase
	Scientific advance	Development of protein sequencing	Sanger worked out the first complete amino acid sequence of a protein (insulin). He won his first Nobel Prize for Chemistry for his work on this.	Fred Sanger, Cavendish Laboratory, University of Cambridge
1953	Scientific advance	The structure of DNA is described	The structure of DNA is shown to be a double helix. In 1962, the pair receive the Nobel Prize for Medicine for their work.	Francis Crick and James Watson, Cambridge UK

Annex D: Human genetics research timeline

Date	Key	Summary	Description	People and place
1956	Scientific advance	Human chromosome number is determined	Determination that the correct human chromosome number is 46.	Joe-Hin Tjio and Albert Levan – University of Lund, Sweden
	Scientific advance	Genetic mutation underlying sickle cell disease is identified	Vernon Ingram, John Hunt, and Antony Stretton determined the change in the haemoglobin molecule in sickle cell disease. Ingram showed that the amino acids of normal human and sickle cell anemia hemoglobins differed due to a single mutated gene.	Vernon Ingram, John Hunt and Antony Stretton (University of Cambridge, UK)
1957	Scientific advance	Semiconservative replication of DNA discovered	First shown that one DNA strand is a template for the other.	Matthew Meselson and Franklin Stahl, California Institute of Technology, USA
1959	Scientific advance	Down's syndrome characterised	Down's Syndrome is identified as a chromosome 21 trisomy.	Jérôme Lejeune, Centre Nationale de la Recherche Scientifique, France
1966	Scientific advance	The genetic code is deciphered	Marshall Nirenberg, Johann H Matthaei, Har Gobind Khorana and Robert Holley describe how four DNA letters code for 20 amino acids (protein subunits). Working independently, Marshall Nirenberg and Johann H Matthaei cracked the first letter of the genetic code (1961), Khorana mastered the synthesis of nucleic acids, and Holley discovered the exact chemical structure of transfer-RNA.	Har Gobind Khorana, University of Wisconsin, USA, Marshall Nirenberg, National Insitutes of Health (NIH), USA, and Robert Holley, Cornell University, USA
	Scientific advance	First edition of *Mendelian Inheritance in Man*	McKusick first published his catalogue of all known genes and genetic disorders, *Mendelian Inheritance in Man* (MIM).	Victor A McKusick, Johns Hopkins University, USA
1968	Scientific advance	First gene mapped to an autosome	Donahue and colleagues demonstrated that the Duffy blood group locus mapped to human chromosome 1.	R P Donahue (Johns Hopkins University, USA)
1969	Scientific advance	Attention is drawn to the fact that all individuals are different in their genetic constitution	Harry Harris was the first to draw attention to the fact that all individuals are genetically and biochemically different. Using simple biochemical tests on isozymes, Harris demonstrated that no two individuals except for identical twins are exactly alike in their genetic make-up. Harris, H. (1969) Enzyme and protein polymorphism in human populations. Br Med Bull 25, 1.	Harry Harris, The Galton laboratory and MRC Human Biochemical Genetics Unit, University College London, UK
1972	Scientific advance	First RNA sequence of a gene is determined	Fiers and team are the first to determine the RNA sequence of a gene (bacteriophage MS2 coat protein).	Walter Fiers et al, University of Ghent, Belgium
	Scientific advance	First recombinant DNA molecules produced	Produced the first recombinant DNA molecules.	Paul Berg, Herbert Boyer and Stanley Cohen, Stanford University, USA
	Scientific advance	Use of somatic cell hybridisation mapping	Ruddle and Bodmer (independently) use somatic cell hybridisation techniques to map genes to chromosomes.	Frank Ruddle (Yale University, US); Walter Bodmer (University of Oxford, UK)
1973	Policy	First human gene mapping workshop	First Human Gene Mapping Workshop held in New Haven, Conneticut, by McKusick and Ruddle.	Victor A McKusick, Johns Hopkins University, USA and F H Ruddle
1974	Scientific advance	Identification of the first human gene deletion as a cause of disease (alpha thalassaemia)	Identification of the first human gene deletion as a cause of disease (in the case of alpha thalassaemia).	S Ottolenghi et al
1975	Policy development	Asilomar meeting	International meeting at Asilomar, urged the adoption of guidelines regulating recombinant DNA experimentation.	
	Scientific advance	Southern Blotting developed	Southern develops a method for detection of specific DNA sequences from DNA samples.	Ed Southern, University of Edinburgh, UK

Date	Key	Summary	Description	People and place
1977	Scientific advance	'Dideoxy' DNA sequencing; first DNA genome is decoded	Fred Sanger's lab sequence the entire genome of bacteriophage F-X174 using the dideoxy method.	Fred Sanger, University of Cambridge, UK
1978	Scientific advance	First DNA prenatal diagnosis of genetic disease (sickle cell anaemia)	First DNA prenatal diagnosis of sickle cell anaemia.	Y W Kan and A M Dozy (University of California – San Francisco, USA)
1981	Scientific advance	Human mitochondrial DNA sequence completed	The complete sequence of the human mitochondrial genome is published.	Fred Sanger and colleagues (MRC Laboratory of Molecular Biology, Cambridge)
1982	Scientific advance	GenBank database formed at Los Alamos National Laboratory. The National Institutes of Health funded Walter Goad's proposal for the creation of GenBank	Scientists begin submitting DNA sequence data to the public GenBank database.	US Department of Energy (DOE)
1983	Scientific advance	Huntington's disease – first disease gene mapped	A genetic marker for Huntington's disease is found on chromosome 4.	James Gusella (Harvard Medical School, USA)
	Scientific advance	Polymerase chain reaction (PCR)	Kary Mullis develops the polymerase chain reaction – enabling the amplification of specific segments of DNA.	Kary Mullis (Cetus Corporation, USA)
1984	Scientific advance	DNA fingerprinting developed	Alec Jeffreys develops the technique of DNA fingerprinting.	Alec Jeffreys (University of Leicester, UK)
1985	Policy development	First meeting on human genome sequencing	Robert Sinsheimer holds meeting on human genome sequencing at University of California, Santa Cruz.	Robert Sinsheimer, The Santa Cruz Workshop, May 1985, Genomics 5: 954 (1989)
	Policy development	Human Genome Project (HGP) first proposed	Charles DeLisi and David A Smith commission the first Santa Fe, New Mexico conference to assess the feasibility of a Human Genome Initiative.	
	Scientific advance	Physical map of nematode developed	Working in Sydney Brenner's lab in Cambridge, Sulston, Coulson and others construct a physical genetic map of the nematode worm during the early 1980s.	John Sulston and Alan Coulson, MRC Laboratory of Molecular Biology, UK
	Scientific advance	Genomic imprinting is described	Experiments in mice demonstrate the phenomenon of imprinting.	Cattanach and Kirk (MRC Radiobiology Unit, UK)
1986	Policy development	Nobel Prize winning molecular biologist Renato Dulbecco called for the HGP	Renato Dulbecco called for the HGP in "A Turning Point in Cancer Research: Sequencing the Human Genome" Science 231: 1055 (1986) and "A Turning Point in Cancer Research: Sequencing the Human Genome in Viruses and Human Cancer" (Alan R Liss, ed, 1987).	Renato Dulbecco
	Scientific advance	First positional cloning of disease gene (CMG)	Orkin reported the first positional cloning of a human disease gene (chronic granulomatous disease) – a method for finding a gene without the knowledge of the protein it encodes is developed.	Stuart H Orkin (Harvard Medical School, USA)
	Scientific advance	Automated DNA sequencing developed	Hood and colleagues pioneered four instruments – the automated DNA sequencer and synthesizer, and the protein synthesizer and sequencer.	Leroy Hood (California Institute of Technology, USA)
	Policy development	Office of Human Genome Research created in NIH (later becomes NHGRI)	Office of Human Genome Research created to lead NIH's human genome efforts.	US National Institutes of Health, USA
	Policy development	NIH and Department of Energy (DOE) coordinating memorandum	NIH and Department of Energy (DOE) establish MOU to coordinate research activities on human genome.	US National Institutes of Health, and Department of Energy, USA

Annex D: Human genetics research timeline

Date	Key	Summary	Description	People and place
1988	Policy development	US National Center for Biotechnology Information (NCBI) established	The NCBI is established within NIH to manage and store molecular and genetic data, under the leadership of David Lipman.	US National Institutes of Health, USA
	Policy development	A report published by the National Research Council (NRC) provides justification for the government to launch the HGP	The NRC's report "Mapping and Sequencing the Human Genome" was highly influential in the establishment of the HGP in the USA.	NRC, Bruce Alberts
	Policy development	A report published by the Office of Technology Assessment (OTA) advanced the cause of the HGP, politically	The OTA's report "Mapping Our Genes: Genome Projects – How Big? How Fast?" was highly influential to the establishment of the HGP in the USA.	US Congress, Office of Technology Assessment. OTA-BA-373 (Washington, DC: U.S. Government Printing Office, April 1988)
1989	Scientific advance	Cystic fibrosis gene identified	Identified the gene coding for the cystic fibrosis transmembrane conductance regulator protein (CFTR) on chromosome 7 that, in mutated form, causes cystic fibrosis.	Francis Collins (University of Michigan, US), Lap-Chee Tsui and Jack Riordan (Hospital for Sick Children, Canada)
	Policy development	Human Genome Organisation (HUGO) formed	HUGO formed to coordinate international activities on the human genome.	International
	Policy development	DOE and NIH establish Joint ELSI Working Group	Work to explore ethical, legal and social issues (ELSI) associated with human genome is integral to US HGP from the outset.	USA
1990	Funding development	Avon Longitudinal Study of Parents and Children (ALSPAC) inception	More than 14 000 mothers enrolled during pregnancy in 1991 and 1992, and the health and development of their children has been followed in great detail ever since.	Jean Golding et al, Bristol University, UK
	Funding development	HGP begins with initial five-year Plan	DOE and NIH present joint five-year US plan to Congress for the HGP. The 15-year project formally begins.	USA
	Scientific advance	Development of pre-implantation genetic diagnosis	Handyside and collaborators' first successful attempts at testing were in October 1989 with the first births in 1990. In these first cases, PCR was used for sex determination for patients carrying X-linked diseases.	Handyside and colleagues (Hammersmith Hospital, UK)
	Scientific advance	Discovery of SRY gene for sex determination	Lovell-Badge and Goodfellow isolated the testis-determining factor gene: the master switch for mammalian sex determination (named *SRY*, for sex-determining region, Y chromosome).	Robin Lovell-Badge and Peter Goodfellow (MRC National Institute for Medical Research)
	Policy development	Human Fertilisation and Embryology Act 1990	Landmark legislation governing the use of human gametes and embryos introduced.	UK Government
1991	Scientific advance	Discovery of the molecular basis of "dynamic mutations"	The fragile X syndrome was the first trinucleotide repeat (dynamic mutation) disorder identified and served as a prototype for several diseases caused by triplet repeat expansions in the human genome.	AJ Verkerk et al
1992	Funding development	Sanger Centre established	Wellcome Trust and MRC establish the Sanger Centre in partnership to enable UK participation in the HGP (officially opened 1993).	Wellcome Trust and MRC
	Scientific advance	Microsatellite map of human genome	Creation of a microsatellite linkage map of the human genome.	Jean Weissenbach Fondation Jean Dausset (CEPH), France
1993	Scientific advance	Micro RNAs (miRNAs) first described	Micro RNAs are first described in worms. They are short RNA molecules that no do not encode proteins and serve to regulate gene expression.	Rosalind Lee, Rhonda Feinbaum and Victor Ambros (Harvard University, USA)

Date	Key	Summary	Description	People and place
	Policy development	NIH withdraw patent application for expressed sequenced tags (ESTs)		NIH
1994	Scientific advance	Discovery of *BRCA1* mutation (*BRCA2* mutation discovered 1995)	*BRCA* mutations, which confer significantly increased risk of breast and other cancers, discovered by Myriad Genetics.	Myriad Genetics, USA and collaborators; Miki et al, 1994; Wooster et al, 1995
1995	Scientific advance	First genome of a free-living organism sequenced (*Hemophilus influenzae*)	Genomic sequence of *Hemophilus influenzae* determined.	Fleischmann and colleagues (Johns Hopkins University, USA)
	Scientific advance	Patrick Brown demonstrates DNA chips	Patrick Brown develops microarrays for use in large-scale gene expression studies.	Patrick Brown (Stanford University, USA)
	Funding development	European Bioinformatics Institute building opens (at the Wellcome Trust Genome Campus)		Wellcome Trust, EMBL, MRC (UK research councils)
	Funding development	The Wellcome Trust committed to funding the sequencing of one-sixth of the human genome, through the Sanger Centre, under the HGP	The Wellcome Trust supported the proposal from the Sanger Centre to spearhead the international sequencing programme to sequence the human genome. The Wellcome Trust agreed to provide continued funding to the Sanger Centre to enable them to sequence at least one-sixth of the human genome, representing the largest financial commitment to human sequencing at a single centre in the world.	The Wellcome Trust and the WTSI
1996	Scientific advance	Yeast (*Saccharomyces cerevisiae*) genome published	The full genome sequence of budding yeast is published – the first eukaryotic organism to be sequenced.	International
	Scientific advance	Afftymetrix makes DNA chips commercially available	US company Affymetrix markets microarrays for research use.	Affymetrix, USA
	Scientific advance	Creation of Dolly the Sheep by Wilmut and Campbell	The creation of Dolly the sheep, the first mammal to be cloned from a somatic cell, is announced by researchers at the Roslin Institute in Edinburgh.	Ian Wilmut and Campbell Roslin Institute, UK
	Policy development	Bermuda meeting of human genome scientists agree open access data sharing principles	The collaborators in the HGP agree that all data produced should be made freely available within 24 hours.	Wellcome Trust
	Policy development	HUGO statement on principled conduct of genetics research	HUGO sets out a series of ethical principles for international collaborative research on the human genome.	HUGO
	Funding development	NIH began funding for human sequencing efforts under the HGP		NIH
1998	Scientific advance	Nematode worm sequence published	The nematode sequence is published – the first multicellular organism to be sequenced.	Sanger Institute (UK) and the Genome Sequencing Center (Washington, USA)
	Scientific advance	Discovery of RNA interference (RNAi)	Fire and Mello characterise RNA interference (RNAi) – a natural phenomenon that can be used to inactivate specific genes of interest (received 2006 Nobel Prize).	Andrew Fire (Stanford University, USA), and Craig Mello (University of Massachusetts Medical School, USA)
	Scientific advance	Isolation and growth of the first stem cells from human embryos	Scientists at the University of Wisconsin, led by James Thompson, isolate and grow the first stem cells from human embryos. The embryos used in these studies were created by IVF.	Thomson et al, 1998 (University of Wisconsin, Madison, USA)

Annex D: Human genetics research timeline

Date	Key	Summary	Description	People and place
	Funding development	Celera Genomics enters the human genome sequencing 'race'	Celera Genomics was established in May 1998 as a private sector effort to sequence the human genome.	Craig Venter, Celera
	Funding development	Public HGP effort accelerated with additional WT funding	In response to the formation of Celera, the Wellcome Trust substantially raise funding for human genome sequencing work at the Sanger Centre, enabling it to sequence one-third of the genome.	Wellcome Trust
1999	Scientific advance	Publication of first complete human chromosome sequence	The sequence of chromosome 22 is published.	Sanger Centre, UK
	Funding development	Cancer Genome Project (CGP) established	The Wellcome Trust and the Institute of Cancer Research (ICR) launch the CGP, which will use high-throughput mutation detection systems along with the rapidly emerging data from the HGP to find the gene abnormalities associated with all forms of human cancers.	Professor Michael Stratton, Sanger Centre, UK, Dr Richard Wooster, ICR, UK.
2000	Scientific advance	Working draft of human genome announced by HGP and Celera	The completion of a working draft sequence of the human genome is announced by Bill Clinton and Tony Blair.	HGP (international); Celera
2001	Scientific advance	Draft sequences of human genome published	The HGP publishes a 90 per cent draft of the human genome sequence in *Nature*. Celera publishes its sequence concurrently in *Science*.	HGP (international); Celera
	Scientific advance	The SNP Consortium Ltd maps genome variation	The SNP Consortium Ltd. (funded by the Wellcome Trust and 13 pharmaceutical companies) publishes a single-nucleotide polymorphism (SNP) map of human genome.	The SNP Consortium Ltd
	Policy development	Myriad Genetics granted *BRCA1* patent		
	Scientific advance	Gleevec approved by US Food and Drug Administration	Gleevec represented a new class of cancer drugs which targeted abnormal proteins fundamental to an individual cancer; drug is targeted at the bcr-abl protein in chronic myeloid leukaemia cells.	
	Policy development	UK Government and British Insurers agree moratorium on use of genetic tests		
2002	Scientific advance	Draft sequence of the mouse genome published in *Nature*	A draft sequence of the mouse genome is published in *Nature*	Mouse Genome Sequencing Consortium (UK and USA)
	Funding development	UK Biobank project gets green light	UK Biobank project is launched by MRC, Wellcome Trust and Department of Health to establish a resource collating medical samples, medical history and lifestyle information for 500 000 UK citizens.	MRC, Wellcome Trust, Department of Health
	Funding development	International HapMap Project Launched	The International HapMap Project is established to develop a haplotype map of the human genome (HapMap). First use of "click-wrap" open source licence.	International Consortium
2003	Scientific advance	'Gold standard' human genome sequence is completed	The human genome sequence is completed (published in 2004).	HGP (International)
	Funding development	Structural Genomics Consortium launched	The Wellcome Trust, four Canadian funders and GlaxoSmithKline launch the Structural Genomics Consortium for high-throughput protein structure determination.	International Consortium (UK, Canada, later Sweden)
	Funding development	ENCODE project launched by NHGRI	NHGRI launches the ENCODE (Encyclopedia Of DNA Elements) initiative to identify all functional elements in the human genome sequence – with a pilot to characterise 1 per cent of the genome.	NGHRI, US

Date	Key	Summary	Description	People and place
	Policy development	UK Department of Health publishes White Paper on Genetics	Department of Health White Paper published setting out a strategy for developing genetic services in NHS.	Department of Health, UK
2004	Scientific advance	Chicken and rat genomes published	Full sequences of the rat and chicken genomes are published in *Nature*.	International consortia
	Policy development	Human Tissue Act (2004)	The Human Tissue Act is passed by the UK Government. The legislation makes it a criminal offence to test an individual's DNA without their consent.	UK Government
2005	Scientific advance	International HapMap Project publishes catalogue of human genetic variation	The International HapMap Project published haplotype map of human genome.	International HapMap Consortium
	Scientific advance	Chimp genome published	The chimpanzee genome sequence is published in *Nature*.	International Consortium (USA, Germany, Italy)
	Funding development	Wellcome Trust Case Control Consortium (WTCCC) established	WTCCC established to explore the utility, design and analyses of genome-wide association (GWA) studies in common diseases.	Wellcome Trust
2006	Scientific advance	Copy number variation (CNV) of human genome described	In analysis of 270 people, researchers find hitherto unanticipated volume of variation – 1447 CNVs in nearly 2900 genes.	Matt Hurles et al, Wellcome Trust Sanger Institute, UK and Harvard Medical School, USA
	Funding development	The Genetic Association Information Network (GAIN) is launched in the USA	A unique public–private medical research partnership – the Genetic Association Information Network (GAIN) – is launched to unravel the genetic causes of common diseases through whole GWAS. GAIN is an analogue of the WTCCC.	The Foundation for the National Institutes of Health (FNIH), NIH, and Pfizer Global Research & Development, New London, Connecticut
	Scientific advance	NCBI database of genotypes and phenotypes (dbGaP) established	The National Center for Biotechnology Information created the dbGaP public repository for individual-level phenotype, exposure, genotype and sequence data and the associations between them.	NCBI
2007	Scientific advance	Wellcome Trust Case Control Consortium publication	Trust-funded Consortium publishes what was the largest ever study of genetics of common disease.	UK (various)
	Scientific advance	ENCODE publication – new insights into genome function	The results of the pilot phase of the ENCODE project are published – a second phase of the project is initiated.	NIH (international consortium with UK input)
	Scientific advance	Human induced pluripotent stem cells produced	Induced pluripotent stem cells were first generated from mouse cells 2006 by Shinya Yamanaka's group at Kyoto University. In 2007, iPS cells were generated from human cells by two independent teams – one led by Yamanaka and the other by James Thomson and Junying Yu at the University of Wisconsin-Madison.	Shinya Yamanaka (Kyoto University, Japan and James Thomson and Junying Yu (University of Wisconsin-Madison, USA)
	Funding development	Launch of DeCODEme and 23andMe	Two private companies enter the marketplace offering genome testing services direct to the public open for business.	International
	Policy development	Myriad *BRAC1* patent struck out in Europe		Europe
	Policy development	Herceptin is made available in NHS for certain patients	Herceptin – a monoclonal antibody therapy that targets certain forms of cancer – is made available in the NHS (one of the few pharmacogenetic therapies).	UK
2008	Funding development	1000 Genomes Project launched (UK, US and China)	The 1000 Genomes Project is an international research effort to establish by far the most detailed catalogue of human genetic variation.	International

Annex D: Human genetics research timeline

Date	Key	Summary	Description	People and place
	Funding development	International Cancer Genome Consortium launched	ICGC is established – bringing together funders from 10 countries to catalogue genetic changes in tumours from 50 different cancer types.	International
	Policy development	US Genetic Information Non-discrimination Act (GINA) passed into law	The US Genetic Information Non-discrimination Act (GINA) is finally passed after years in development. GINA prohibits discrimination based on genetic information in both health care and employment.	USA
	Policy development	UK moratorium on genetic testing extended until 2014	The existing UK moratorium on use of genetic testing in insurance is extended.	UK
	Policy development	NIH GWAS policy implemented	The goal of the GWAS policy was to facilitate broad and consistent access to NIH-supported GWAS data in order to speed the translation of basic genetic research into therapies, products, and procedures that benefit the public health.	NIH
2009	Policy development	23andMe launches collaborative Parkinson's Disease research initiative		23andMe, The Parkinson's Institute and Clinical Center and Michael J Fox Foundation, USA
	Scientific advance	Sequencing of two complete cancer genomes	Sequencing of two complete cancer genomes – small cell lung carcinoma and malignant melanoma.	Cancer Genome Project at the WTSI
2010	Funding development	UK10K	To sequence 4000 samples from the ALSPAC and TwinsUK cohorts.	Wellcome Trust
	Policy development		Federal District Court rules that Myriad *BRCA* patents are invalid.	
	Funding development	Human Heredity and Health in Africa Project, or H3 Africa	A £25 million partnership established by the NIH and the Wellcome Trust to support population-based genetic studies in Africa, including research into common, non-communicable disorders such as heart disease and cancer, as well as infectious diseases such as malaria.	NIH, Wellcome Trust

Contact

Strategic Planning and Policy Unit
Wellcome Trust
Gibbs Building
215 Euston Road
London NW1 2BE, UK
T +44(0)20 7611 7246
F +44(0)20 7611 8742
E SPPU@wellcome.ac.uk

The Wellcome Trust

We are a global charitable foundation dedicated to achieving extraordinary improvements in human and animal health.

We support the brightest minds in biomedical research and the medical humanities. Our breadth of support includes public engagement, education and the application of research to improve health.

We are independent of both political and commercial interests.

SP-5081/tba/03-2011/BS