

House of Lords Science and Technology Committee: Inquiry on Genomic Medicine

Response by the Wellcome Trust Sanger Institute

April 2008

Introduction

The Wellcome Trust Sanger Institute (WTSI) is pleased to have the opportunity to contribute to the Committee's inquiry on genomic medicine.

1. The WTSI is one of the leading academic centres in the world for genomic and genetic research. Funded by the Wellcome Trust, the WTSI generates and analyses genomic data and provides resources that support further discoveries in genetic research.

2. The prominent position of the WTSI is based on major investment in state of the art genomics technology platforms, massive computer resources and funding streams which support an environment for high throughput research that is not achievable in traditional university settings. WTSI currently has the largest operational sequencing capacity in the world and is capable of sequencing a human genome sequence per day.

3. The major WTSI platforms of DNA sequencing, genotyping and informatics together with initiatives in model organism research deliver data for international collaborative research projects in healthcare related genomics. The Institute played a key role in the International Human Genome Project, taking responsibility for producing one third of the genome sequence. WTSI investigators have subsequently taken leading positions in major international projects such as the International Hapmap Consortium, the 1000 genomes project and the recently launched International Cancer Genome Consortium. The participation of the WTSI in these international projects has helped provide the scientific community with a comprehensive picture of the human genome and has generated tools to monitor and analyze genome variations with high accuracy

4. The major research platforms of WTSI are generally employed in large collaborative projects with other researchers, frequently within the UK research community. For example the Wellcome Trust Case Control Consortium (WTCCC) was the internationally leading collaboration in 2006/7 to use whole genome association to find genes involved in common disease. The WTSI's presence in the UK helps UK medical genetics research have a leading international presence.

5. The WTSI has been a leading advocate of open access to data. The Institute supports the delivery of free annotated genome sequence to hundreds of thousands of users across the globe 24 hours a day, 365 days a year through online resources such as Ensembl (www.ensembl.org)

6. The current research programme in human genetics at WTSI focuses on large-scale studies of genetic variation in order to advance our understanding of gene function in health and disease. In this regard, the Institute has made several major discoveries ranging from the identification of cancer causing sequence changes to description of the patterns of copy number and structural variation in the human genome.

Key messages of WTSI

7. There has been a dramatic increase in the capacity to collect genetic and genomic data in recent years. Our ability to understand basic human biology has been

transformed via the high throughput data production platforms developed in collaborations between academia and industry, which have been deployed in academia on a large scale. This has resulted in the rapid advancement of genomic research and in major breakthroughs in our understanding of the biology behind human health and diseases. We are beginning to see the application of genomic technologies in health care. This will only increase. The further development of understanding and applications will play a major part in future healthcare.

8. We are seeing now a remarkable wave of discovery of new genetic variants that predispose to disease. However it is likely to take several years for our understanding of the disease risks associated with these variants to be clarified, for our understanding of how to use them in clinical disease management to mature, and for intervention strategies that might be implemented on the basis of their identification to be developed.

9. UK scientists are world leaders in genetics and genomics research. Within Europe the UK is the leading country in basic and applied genomic research. Europe and the UK are in a unique competitive position to translate this position into novel diagnostic and treatment strategies for many public health problems. If this unique position is properly exploited with strategic vision and targeted funding decisions, the UK can play a key role in the implementation of the basic innovations derived from genomics within biology based medicine and could participate in some of the commercial developments.

10. To deliver on the opportunities will require strategic vision and sustained investment. This type of research depends on large series of individuals being subjected to genetic analyses as part of research protocols, with the genetic information generated being associated with extensive phenotypic information. The research framework and infrastructure underpinning this type of research requires sustained support and expansion at both national and European levels.

11. There will need to be ongoing oversight of the requirement for regulation in this area. DNA analysis technologies are advancing very quickly. These technologies potentially will allow members of the public to access information on essentially all their sequence variation, often without direct medical supervision or advice. This will almost certainly generate questions from the public as to the significance of the results with respect to disease risk, and potentially other implications, for example insurance status. This new wave of individual-based genetic information is likely to generate a new and substantial demand on clinical services. Consideration of the regulation of genetic typing services based on these new technologies and the health service response to them may be required. However there are issues of practicality (information may be available via the internet from anywhere in the world), and individual freedom to access ones own personal information.

Research and scientific development

Questions

What is the state of the science? What new developments are there? What is the rate of change?

Who is taking the lead in the consideration and co-ordination of research and the development of new technologies?

How effective is the policy and investment framework in supporting research in this area?

How does research in the UK compare internationally? How much collaboration is there?

What are the current research priorities?

What is the role of industry? How much cross-sector collaboration takes place?

State of the science

12. Most of the germline single nucleotide variation in the human genome that is common in human populations has already been identified. The “1000 genomes” project (www.1000genomes.org) now aims to identify the remaining common and much of the moderately rare single nucleotide variation in the human genome. Through the combined efforts of major international genome centres, a comprehensive catalogue of this type of human genome variation from different global populations will be generated. Alongside this initiative are ongoing major programmes at WTSI and elsewhere to provide more complete catalogues of germline copy number and structural variation in the human genome in several populations. The WTSI Cancer Genome Project and the forthcoming International Cancer Genome Consortium will provide similar types of information for somatically acquired variation across the broad range of human cancers. These studies will therefore generate much more complete catalogues of human sequence variation, empowering the scientific community to collect information regarding the phenotypic significance of these genetic variants in health and disease.

13. Understanding of the role of human genome variation in health and disease is progressing very rapidly. We know that most common diseases in the UK have a substantial genetic component. However, for most common human diseases only a small proportion of disease susceptibility has been explained in terms of identified disease-causing variants. A wave of association studies using large numbers of cases and controls, studying both single nucleotide variation and copy number/structural variation and directed at diverse disease phenotypes over the next few years will undoubtedly explain much more of the genetic basis of many common diseases. Currently, however, we understand the phenotypic consequences (i.e. the end effect on people) of only a tiny proportion of human genome sequence variants. Large scale and ultimately complete resequencing of human genomes in large numbers of individuals will likely contribute to this discovery process; this is only a few years away.

14. Genetic testing is already part of routine clinical practice worldwide. Currently, such tests are usually conducted under medical supervision and in individual cases are directed at a small number of genes that are responsible for a specific disease which is often rare. In these circumstances the disease-causing mutation is often rare in the population and causes a greatly elevated risk of the disease. However, the recent wave of discoveries in human disease genetics has focused primarily on common diseases with the newly discovered disease-associated variants also being common and imparting only a small increased disease risk. Our current knowledge of the characteristics of the risks conferred by the latter class of variants is rudimentary and our experience of the usage of this class of disease-causing genetic variant in clinical practice is for now very limited.

15. Rapidly advancing technology, however, will allow low cost analyses of many/all variants of individual genomes in the not so distant future, and will therefore transform our concept of diagnostic and predictive tests in health care. Genetic testing for predisposition to many traits and diseases will become possible. Many of these will be for variants that are common and confer a small increased risk of disease. For some diseases/variants it will be possible to intervene at an early stage with clinical management strategies that favour a better outcome. For others, however, this may not be possible for the foreseeable future. Thus information may be provided by tests that have no direct clinical impact. Such tests are likely to be accessible directly by the public without medical supervision or advice. Diagnostic companies are already offering services for genetic risk prediction via the internet. Although such tests are generally not strongly informative currently, the level of public interest is potentially high. It is likely that the services offered and results provided to individuals will not fulfil the basic criteria of

evidence based medicine, and there is a risk of exploitation. Increasing numbers of test providers will present a huge challenge for interpretation, guidance and counselling. Although the genetic information will remain essentially the same throughout an individual's lifetime, its interpretation will change over time as the science advances. Furthermore, the relevance of the genetic information will also be different depending on the age of an individual.

Who is taking the lead in the consideration and co-ordination of research and the development of new technologies?

16. There is considerably more coordination of this area of research both in the UK, by Wellcome Trust and other funding bodies, and in the USA by NIH, than is the case generally in biomedical research. With respect to development of new technologies there is less coordination, although one example might be that the US National Human Genome Research Institute are fostering a technology drive to achieve the \$1000 genome sequence. On the other hand, it is not clear that more coordination would be beneficial in this context and the main technology drive will probably continue to originate primarily from industry. Much of the industry activity and funding is in the USA, although there have been some very notable recent UK successes (e.g. Cambridge University spinout Solexa Ltd, now owned by Illumina Inc of San Diego).

17. The genome sequence is essentially information, and collation, organisation and presentation of information is central. Support for such genome and genetics based databases is therefore important. The leading centres for this are at NCBI in Washington DC, and at Hinxton in the UK (the site of WTSI and the European Bioinformatics Institute, EBI). The European Union, EMBL and the Wellcome Trust support the Hinxton activities. Ensembl, maintained by WTSI and EBI, is a leading example that is now routinely used by researchers and clinicians in the field. In the context of human genetic disease, the DECIPHER project at WTSI is coordinating genomic data collected from samples worldwide both in research and diagnostic contexts. Similarly, the COSMIC database at WTSI organises and presents information on cancer associated somatic mutations.

How effective is the policy and investment framework in supporting research in this area?

18. The Wellcome Trust has a strong policy and investment record in this area that is well respected by scientists, fostering genomics and genetics research both at WTSI and elsewhere in UK biomedical science. These efforts have provided major data resources for genomic and genetic scientists, including the Human Genome Sequence, Hapmap, the Cancer Genome project, descriptions of copy number variation, Wellcome Trust Case Control Consortium datasets and databases such as Ensembl. The UK government has provided strong moral and philosophical support, and through the MRC has historically supported genomics and medical genetics strongly. However, for the last few years it has not been able to match the scale of Wellcome Trust support for genomics, and it does not appear to have a well articulated and coordinated strategy in this area of research. Similarly, despite a significant number of short-term individual initiatives, there has an apparent lack of large scale strategic vision and infrastructure support at the European level.

How does research in the UK compare internationally? How much collaboration is there?

19. The strong history and track record in genetics and in epidemiological research makes the UK a world leader in both basic and disease-related genomics research which is relevant for public health. The WTCCC provides one remarkable and recent example of this leadership. International collaboration in genomics research is a strong and

enduring theme and the UK has often held a leading position in major international collaborative initiatives.

20) These include:

- a. The Human Genome Project, of which one third was generated at WTSI.
- b. The International HapMap Project which characterises common variation of human genome.
- c. The Genome Structural Variation Consortium, which characterises copy-number and structural variations across the human genome.
- d. ENCODE (Encyclopaedia of DNA elements), an international initiative funded by the US National Institutes of Health characterizing functional elements in the human genome.
- e. The 1,000 Genomes Project, an international initiative aiming to sequence the genomes of 1,000 individuals from different populations.
- f. The International Cancer Genome Project which aims to systematically scan the cancer genome to identify somatically mutated cancer genes, in which the WTSI Cancer Genome Project has played a leading role.

21. The data resources produced by these international projects are provided under an open access policy to the scientific community and have dramatically expedited the progress of health related genome research around the world. This international leadership creates significant potential both for UK research and health care applications internationally.

22. Alongside these genomic studies, there are many UK case collections and cohorts that are being exploited for genetic analysis, including the 1958 birth cohort, EPIC, ALSPAC and others. Increasingly these are collaborating with each other and with other European and international collections, to give the numbers of affected individuals necessary for adequate statistical power for modern association genetics.

23. By its nature, scientific research cannot and should not be centrally dictated or coordinated. However, coordination by funders has clearly been highly effective and helpful in the context of genomic and genetics research when the strategic imperatives of major funding agencies have been configured by close consultation with the scientific community. Many large scale genomics projects have followed this pattern from the Human Genome Project itself 15 years ago to the WTCCC more recently.

24. The UK should take a more active leadership role in many European operations, especially in projects related to the development of European Research Infrastructures relevant for life sciences and health care related research. These include biobanking related activities (BBMRI), clinical and translational activities (EATRIS) and infrastructure for biological information (ELIXIR). These are all projects identified as priorities in the 2006 European Roadmap developed by the European Strategy Forum on Research Infrastructures (ESFRI). Given the very rapid growth in data generation from new technology platforms, adequate support for infrastructure to store and distribute biological information in Europe is particularly important. If not addressed, the European research community will be disadvantaged, making it more likely that European data will be integrated in the USA and that discoveries and applications will be developed there. Strategic developments in the future should increasingly involve partnerships with Asian centres (such as genome centres in Singapore, China and India). Increased strategic alliances with major European centres could also be highly beneficial, especially in the field of analyzing biobanks. Sample and sample-related data resources within industry, such as those held by GlaxoSmithKline and DECODE would also be valuable to the scientific research community if ways could be found to share them.

What are the current research priorities?

25. We identify the following immediate priorities:

- a. Identification of more of the sequence variation, particularly including a higher proportion of the rarer variants, that is present in human populations. This will, at least in part, be achieved through the activities of the international “1000 genomes” project.
- b. For a broad range of human diseases, identification of a much larger proportion of the disease-associated variants, and hence obtaining a more complete picture of the overall genetic structure of disease susceptibility. This aim will be achieved initially by further association studies, both in diseases not yet studied and larger association studies in those diseases for which studies have been carried out (together with combining already conducted association studies). Subsequently, resequencing of all or large sectors of the genome in substantial numbers of disease cases and controls may become the preferred strategy for variants of lesser effect.
- c. Detailed study of the risks of disease-associated variants in large population cohorts and large clinical study sample collections. This will provide a more comprehensive picture of the role of genetic variation and its relationship to our lifestyle and environment, and is necessary to convert our increasing knowledge of the genetic contribution to disease into public health improvements. This would require a national cohort policy as well as sufficient funding to analyze the wealth of data in population cohorts. A proper structure should be established to guarantee broad access to population cohort data to all investigators. The UK Biobank will produce an excellent data set for future generations of scientists interested in the genetics of common health problems. However, this will take substantial time to establish, and should not exclude other (including legacy) biobank initiatives within the UK. Genetic analyses of prospective cohorts are essential to the translation of genetic information to public health actions. Multiple cohorts are required to cover a range of diseases and their collection should be of the highest priority at the national level. A full evaluation of legacy cohorts and biobanks should be undertaken and those exhibiting the best potential for new information collection should be sufficiently supported. Such an evaluation could form the basis for a national cohort policy. The UK could substantially increase its visibility in this area by taking collective action at the EU level and by supporting the necessary UK infrastructure.

What is the role of industry? How much cross-sector collaboration takes place?

26. There are two ways that industry and academia can collaborate. First, academia provides leads and expertise from new research for the development of new therapies and technologies. Some of this is by direct consultancy and technology transfer, and some from public information arising from academic research. Second, industry generates information that can be very valuable for research purposes.

27. The massive data collection from tens of thousands of samples has proven to be highly successful for gene and pathway identification behind common health problems. Such basic association studies and new findings should encourage industry and academia to join forces in the early stages of research. Based on its international stature in genetic and epidemiological research, the UK could be in the leading position of research efforts aiming to understand the genetic basis of both favourable and adverse responses to various medications. Such data from numerous clinical trials could be collected in a co-ordinated manner in NHS hospitals. The ongoing development of the NHS electronic care records service (NHS CRS) and work being done within the Research Capability Programme of NHS Connecting for health (CfH) should help

facilitate this. Again, a national strategy with well structured and integrated cataloguing and tracking would be extremely useful and expedite phase 2 and phase 3 clinical trials in drug development.

28. Cross talk between scientists and pharmaceutical industries could be improved and increasing these interactions should be encouraged. For example, making Phase 2 and Phase 3 clinical trial data available to scientists (including negative results) would have the potential for expanding our understanding of the impact of genetic variants in drug response and in molecular pathogenesis of diseases more generally.

29. An important area of genome research is the genetics of infectious diseases. Combined information from pathogen genomes and the human genome will transform our understanding of pathogen host interactions. This new understanding provides new potential for diagnostics, vaccination and therapeutics of infectious diseases.

Data use and interpretation

Questions

Is genomic information published, annotated and presented in a useful way? Should there be a common, public database? If so, who should fund, and have responsibility for, such an initiative?

Who should provide the framework for optimal evaluation of data and translational opportunities? What policy and funding mechanisms are in place for recognising and utilising potential opportunities?

Is other medical information recorded in a suitable format to allow optimal interpretation of genomic data? How should genomic data be brought together with other health information?

What are the implications of the generation and storage of genome data on personal data security and privacy, and on its potential use or abuse in employment and insurance? How should these be addressed?

Is genomic information published, annotated and presented in a useful way? Should there be a common, public database? If so, who should fund, and have responsibility for, such an initiative?

30. Data resulting from genomic research is generally published, annotated and presented in a useful way, in common public databases. The Human Genome Project set a new paradigm for sharing research data within the life sciences. Sequence was placed in the public domain essentially as soon as it was produced in order that it would be freely available to all without restriction. This Open Access policy has resulted in substantial advancement of genomic research worldwide. Examples are the Ensembl database, a joint project of WTSI and EBI, and Cosmic.

31. These public archives are funded by government or charities. Public funding is necessary because these archives are an infrastructure resource for researchers. Much experience has led to a broad recognition that there is no effective mechanism for users to pay without hindering access and downstream usability of the data. These public data resources are the way that investment in genomic research is realised. Promotion and planning of data sharing policies for major genomics initiatives is a critical task for all major funding agencies, and should be systematically associated with their funding decisions. Coordination of sustained data resource funding is clearly important, and should be promoted at the national, European and International scales commensurate with the requirements from the new generation of genomics and genetics projects.

32. The new genomic medical data contains information on private individuals, and it is necessary to safeguard patient confidentiality, by working within appropriate ethical frameworks. Mechanisms for ensuring this ethical security are being established, while allowing maximal use of research data.

What policy and funding mechanisms are in place for recognising and utilising potential opportunities?

33. The UK is in an excellent position to make strategic investments for our future health care system. Such an investment could be compared to the investment made via the MRC in the Human Genome Mapping Project (HGMP) in the 1990's. This followed from an OST report, and funded research programmes and the HGMP Resource Centre. This helped set in place the UK's leadership position in Europe in human genetics. We are at a similar point now in the development of genomic medicine.

Is other medical information recorded in a suitable format to allow optimal interpretation of genomic data? How should genomic data be brought together with other health information?

34. Currently, the additional medical information that is coupled with genomic data in medical genetics studies is collected in separate research databases, specific to each investigator and project. There is potential for using the standard clinical record beyond the limited set of notifiable events that can currently be tracked, for example via the planned Research Capability Programme of NHS. However there are concerns about the consistency and quality of the information for research purposes.

35. The provision of genomic information in a format suitable for use by clinicians or health care providers is not well advanced. There will be a substantial need for resources to present information in a format suitable to support clinicians in their decision making. One example of an existing project at WTSI is the DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources) which uses genomic array technologies to identify chromosome abnormalities in children with developmental defects and provides this information linked with clinical information about the chromosomal abnormality present. Support for such national and international data transfer systems will be of critical importance for translation of genomic data for public health benefit.

36. There should be planning now for how to store genetic and genomic information as part of each patient's clinical record.

What are the implications of the generation and storage of genome data on personal data security and privacy, and on its potential use or abuse in employment and insurance?

37. In principle, genomic data and/or phenotype data can be used to identify individuals by matching and comparing datasets from two sources. This could potentially lead to a breach of personal privacy. A balance has to be drawn in both the research context and in clinical practice between, on the one hand maximising benefit to patients by not imposing highly burdensome bureaucratic structures that aim to prevent unsanctioned identification, and on the other protecting the privacy of individuals. Current practice for protection of privacy is aware of these issues, and appears to be working satisfactorily in the research context. There should, however, be a clear statement from government, supported by criminal legislation, prohibiting the identification of individuals through matching genomic or other data, without consent.

38. The UK Biobank initiative has set a gold standard for ethical principles and guidelines concerning the large population studies. Again international harmonization would be valuable. The Public Population Project in Genomics (P3G) consortium aims to promote scientific interoperability and harmonization of regulatory frameworks between major cohort studies.

Translation

What opportunities are there for diagnostics, therapeutics and prognostics - now and in the future?

Who is responsible for translation to clinical practice?

Given the pace of technological advance, how 'future-proof' is healthcare investment in this area?

How does the UK compare to other countries and what lessons can be learnt?

How meaningful are genetic tests which use genome variation data? What progress has been made in the regulation of such tests?

39. Genomics research offers considerable potential for the development of new classifications and diagnostics for many diseases.

- a. In the immediate future, comprehensive diagnostics for most monogenic diseases will be provided but their efficient use within a health care system will require further development of clinical diagnostic laboratories and specialised training of clinicians and health care providers.
- b. In the longer term, the recent development of genetic profiles of common variants will provide the basis for predictive testing for susceptibility to late onset common diseases. Here the demand for adequate education, training and counselling of health care providers, test providers and the public is substantial. This will include science driven cataloguing efforts as described above, websites, and training programs for clinicians and health care providers.
- c. It is also foreseeable that there will be a gradual emergence of pharmacogenetic tests which could produce significant benefits both for patients and the health service with better targeted medication and minimized adverse events.

40. The NHS is responsible for training clinicians and health care providers in the use of genetic tests for health care. The requirement will be substantial and will require a national strategy and funding of the infrastructure.

41. There is also a longer term potential to develop new therapeutics and prevention strategies. Here the close interaction between researchers and the pharmaceutical and biotechnology sector is important in progressing the commercial development of new healthcare products and technologies.

Biomarkers and epidemiology

In what way do genome-wide association studies contribute to the identification of biomarkers? How is the study of genetic factors and biomarkers integrated for translational purposes?

What impact will genomic data have on data emerging from projects such as UK Biobank, Generation Scotland and other biobanks?

42. There is a potential for strong synergy between the use of biomarkers and genetic information. Biomarker assays are being incorporated in genome-wide association studies.

43. The availability of genomic data and data that will emerge from biobanks will provide a very powerful tool for researchers studying various environmental and genetic factors at play in complex diseases. The UK is uniquely positioned to exploit the data emerging from such population cohorts and biobanks and in its ability to access patient data from the NHS to inform such studies. Again, close European and international collaboration is required to create the best datasets and to provide solid scientific evidence of the applicability of this new information in health care system.

Use of genomic information in a healthcare setting

What impact will genomic information have on the classification of disease? How will it affect disease aetiology and diagnostic labels?

How useful will genomic information be as part of individualised medical advice? What provisions are there for ensuring that the individual will be able to understand and manage genomic information, uncertainty and risk?

Should there be a regulatory code (mandatory or voluntary) covering the provision of this advice?

What are the implications of developments in genomic technologies for the training of medical specialists and other health professionals? Are there any gaps that need addressing? What is the assessment and planning for future needs in capacity?

44. Our knowledge of underlying genetic factors and disturbed biological pathways will contribute to an improved, more biology-based, diagnostic classification of diseases. It is likely that this will ultimately enable better treatment and preventive strategies.

45. A national strategy for informing, counselling and educating health care providers and the public needs to be designed. The underlying scientific information is just emerging, and many findings need to be verified and tested. There are special concerns in the situation where information can be provided but no effective clinical intervention exists yet.