

Submission to the House of Lords Science and Technology Committee Enquiry into Genomic Medicine

By

Professor Peter Donnelly FRS, Director of the Wellcome Trust Centre for Human Genetics, University of Oxford.

20 April 2008

Below, I will address only some of the questions on which the enquiry seeks evidence. I have also seen the submission which will be sent by the Wellcome Trust.

Research and Scientific Development

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

In the area of basic research into the genomic causes of common diseases, immense progress has been made within the last two years, and the rate of discoveries is increasing. The key reasons for this are: (i) technologies have been developed to allow determination of common genetic variants across the whole genome, both quickly and cheaply, in many individuals; (ii) large cohorts of people with disease have been brought together for comparison with similar numbers of healthy people; (iii) methods have been developed to get the most information from the resulting massive datasets.

Where 18 months ago there were only a handful of genetic variants reliably associated with common disease, there are now well over 100 (see <http://www.genome.gov/GWAstudies/> for a comprehensive list of published studies).

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

Many research initiatives are international in scope. I chaired the WTCCC (Wellcome Trust Case Control Consortium), a collaboration of 200 UK scientists responsible for the largest study to date of the genetics of common diseases, and will also chair WTCCC2, a new consortium studying 15 diseases in ~60,000 samples.

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

The UK is a major international player in this field, in each of (i) technology development and access to new technologies (ii) development of statistical methods to fully utilize the information in modern genomics data sets, and (iii) performing successful disease association studies. As noted above, the single largest study of the genetic basis of common diseases completed to date, the (WTCCC), was a UK-based collaboration. This study, with follow-ups and collaborations, has been responsible for over 50 of the currently known associations. The Wellcome Trust has recently announced £30M in follow-up funding for association studies, including WTCCC2, in 25 diseases, analyzing 120,000 DNA samples. It is evident that the UK leads the way in collaborative studies that have revolutionized large-scale genetics.

What are the current research priorities?

Current priorities, both in the UK and elsewhere, include extending the now proven success of genome-wide association studies, following up the existing findings to understand disease mechanisms and develop new interventions, and applying next-generation sequencing technologies to large-scale studies of the genomics of common diseases.

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

Broadly, private enterprise is involved in two aspects of current research: development and supply of advanced technologies (especially novel genotyping and sequencing technologies), and service provision – high-throughput sample processing and analysis. Most technology companies are US-based, although UK and European companies have been responsible for key developments (e.g. Solexa next generation DNA sequencing, developed in the UK but bought out by Illumina, a major US player in genomic technology). Numerous companies offer services to genomics research, in areas such as sample processing, genome-wide genotyping and high throughput sequencing, and the market is essentially global.

Data Use and Interpretation

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?

Completed genomic association studies will provide immensely useful information for future data-mining and meta-analysis. The (US) Database of Genotype and Phenotype (dbGAP) and the European Genotype Archive (EGA, based at the European Bioinformatics Institute, Hinxton, Cambridge) have been established to meet a growing need for the curation and storage of such genomic variation data, together with clinical and phenotypic information. Data sharing between studies is not straightforward, partly because different platforms may be

used, that result in data on different sets of markers. It is clear that significant support will be required to achieve a UK-European database that can extract the full value from resource-intensive genomic studies.

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?

While individual scientists have incentives for recognizing the commercial and translational value of their discoveries, these are not always pursued in their pressured work environment. However, a recent injection of funding via OSCHR (Office for Strategic Co-ordination of Health Research) for translational research had an extremely positive impact.

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?

There are significant challenges in efficiently bringing together medical and genomic information, partly logistically and partly because of the overarching need to maintain individual patient confidentiality. But there are huge potential benefits, for (i) basic research, (ii) translational research, and (iii) NHS R&D. So it is vital to tackle and overcome these obstacles.

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?

Like any personal data, genome data must be kept securely, not least because the success of research efforts depends on the confidence of potential study participants that the detailed genetic data obtained will not be used illegitimately to stigmatize, or to discriminate against, them. The spirit of data protection legislation seems suited to this aim, although whether it is appropriately focused may need to be examined. In insurance, the current (temporary) concordat between industry and government to govern the use of genetic test information in setting premiums seems to be working well. For some comments on the likely distribution of risks and broad utility of genome-wide profiling see section the subsequent answer to ‘How meaningful are genetic tests ...?’.

Translation

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

The term “genetic test” has a range of accepted meanings. Traditionally, it has been used to describe tests to confirm carrier or affected status for rare genetic

diseases, or for paternity tests. More recently, the idea of using genome-wide data for risk assessment for a range of common diseases has become popular.

Much of the language and some of the discussion of “genetic tests” simply does not fit these new developments, and does not apply to applications of results from genetic association studies. These latter are not genetic tests in the sense of providing a definitive answer as to whether an individual has a particular genetic condition, or will go on to develop a particular disease (like the current genetic test for Huntington’s disease). Instead, they allow a refinement of estimates of individual’s risk of developing a particular disease. For common human diseases, the disease outcome is due partly to genetics and partly to environmental and lifestyle factors, and we have not yet learnt all of these risk factors nor how they interact. What is possible now, and potentially interesting as we discuss below, is that knowledge of genetic risk factors can be used to update risk assessments for particular individuals, and potentially to identify sets of individuals at significantly increased risk.

Another important distinction relates to the consumer of the information. In the case of traditional genetic tests, these were usually ordered by health-care professionals and the results returned to them, for passing on, in a careful way, to their patient. My view is that current knowledge of genetic variants associated with common diseases is not yet suited for routine medical use. However, as we discuss below, individuals themselves may be interested and motivated enough to pay for these analyses.

Regarding this use of genomic data, I believe there is a crucial point which is not yet widely appreciated.

For a particular disease, and a typical individual, their genetics is unlikely to have a major impact on their disease risk. Put another way, these genetic markers are not yet good predictors of disease outcome, as has been widely noted. But there is another perspective.

Across the population as a whole there will be some individuals at greatly increased risk of disease, based on their genetics. For example, based on our recent research using known genetic risk factors for Crohn’s Disease^{*}, the top 5% of the population have a disease risk which is 5-8 times greater than the average, while the top 1% have 9-15 times increased risk, and the top 0.1% have 17-29 times the average risk of disease. The analogous findings for Type 2 Diabetes are a 2-3 increase fold for the top 5%, ~4 times increase for the top 1% and ~6 times greater risk than the average for the unluckiest 0.1% of the population. In each case these elevations in risk will increase as our understanding of the genetic basis of disease improves. Crohn’s Disease and Type 2 Diabetes are currently amongst the diseases with the most advanced knowledge.

^{*} Under plausible assumptions about disease model and actual effect size, and allowing for error in estimating these effect sizes.

Here genetics has the potential to identify individuals at greatly increased risk. It is not yet known how best to intervene to improve outcomes, and more research is urgently needed, but it seems wrong to ignore a method for identifying such high risk groups – many of the risk factors of current focus in medicine identify subgroups with much less increased risk.

From an individual's point of view, for most diseases the genetic risk is about average, but over 50 or more diseases the chance of being in the top 5% for one or more is 95%. So a potentially helpful perspective on "consumer genomics" is that for the individual it can identify the small subset of diseases for which their genetics puts them at much increased risk.

There is currently no regulation of the sale of genetic tests to consumers, and no distinction between the various kinds of test and analysis available, whether for rare disease alleles or risk modifiers for common diseases. The Council for Europe has taken the lead in this area, introducing draft regulations for consideration and adoption by member nations that would, amongst other things, broaden the definition of a genetic test, mandate medical supervision of the testing process, and require that tests meet standards of utility and reliability. These seem to be wholly sensible aims, so long as they are framed in terms that recognize the massive potential for technological and analytical development.

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?

This is an area of active research, via large collaborative projects such as MolPAGE (Molecular Phenotyping to Accelerate Genetic Epidemiology) a Europe-wide consortium coordinated from Oxford University that aims to bring together large-scale data resources including genomics and genomic variation, transcriptomics (RNA), metabonomics (small molecules) and proteomics (proteins) in tackling diabetes and related disorders. Even where there is no formal integration of genomic data and biomarker searches, successful genome-wide association studies inevitably lead to biomarker leads. Thus, a potentially important outcome of genomic studies may be the identification of a key biomarker that is more predictive and hence clinically useful than any single genetic variant.

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?

Currently, disease heterogeneity is a potential problem in successfully identifying genetic bases for common diseases (this is where multiple clinical entities are mistakenly lumped together). Conversely, successful identification of genetic risk factors may allow more accurate designations to be developed, and may also lead to a better understanding of common pathways. This has emerged in the results of the WTCCC and other recent studies, where some genes are implicated in multiple diseases, and particular variants are associated with both increased and decreased risk, in different diseases. Currently, the main application of genetic association information is in understanding aetiologies by identifying genes associated with disease that may then become the targets for translational research.

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?

Genomic information is already used in clinical management and personal medical advice, for rare genetic diseases, where risk information is relatively straightforward to convey, but it may still prove challenging to many patients. Its usefulness in formulating individualized advice concerning common diseases like heart disease and type 2 diabetes is not yet clear, for several reasons. (i) we have only just begun to obtain reliable information about genetic associations, so the total proportion of genetic risk that can be accounted for is still low; (ii) the usefulness and effectiveness of advice is unproven – does the genetic information change the advice we give to a particular individuals, and will it be more likely to lead to a change in behaviour than non-genetic information? (iii) Challenges remain in conveying complex probabilistic information to ordinary people.

We remain at an early stage in understanding risk for common diseases; we can identify risk factors, but we don't yet know how combine them with environmental effects and with each other. We also don't yet know how to use this kind of information to make decisions – that is something that will change but needs urgent research.

The number of potentially serious, relatively common diseases for which useful genomic information exists will increase rapidly over the next few years, and the proportion of risk accounted for and the accuracy of predictions for individuals will improve. In the face of many competing dangers, perhaps the most useful way of thinking about personal genomic information may be in identifying the one or two diseases that comprise the main threats to health for each person. Spread over up to 50 diseases, the cost would be modest and comparable to other investigations, and genomic profiling could well become economical for general use. Currently, family history tells us more than genomic information about the

genetic component of risk for most diseases; eventually genomic information will be more informative. For people with limited knowledge of family medical history, genomic data will be particularly useful.

In summary, we can view data from genome-wide scans as potentially interesting for individuals, but not yet appropriate for the NHS. No doubt this will change with time.

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

There are issues about both test availability and advice. In the absence of reliable associations between genetic variants and common diseases before the last year or so, it must be assumed that many established tests do not provide useful information. Advice about genetic risk for common diseases will always be probabilistic, and never absolute. Many people will need assistance in interpreting results of genomic scans or single gene / disease tests, and this will require re-training of personnel. Together, these observations support the introduction of minimum standards for conveying the results of new kinds of genetic tests that might include one-to-one consultation with an appropriately trained genetic consultant.

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?

The most obvious need for training is in providing advice to individuals, either by medically qualified persons, or by genetic counsellors. There will be a need, both for larger numbers of qualified professionals and for changes and updates in knowledge, because of the changing nature of the information. Currently, most tests confirm carrier or affected status for a single disease, with clear implications for treatment, outcome and having children. Probabilistic information about competing risks and the uncertainty of predictions will require careful handling and specialized training. More generally there is a major need to educate the healthcare professionals themselves about the new findings in genomic medicine.