Impact case study (REF3b)

Institution: University of Oxford
Unit of Assessment:10 – Mathematical Sciences
Title of case study: Driving clinical genetic testing and biotechnology development based on the International HapMap Project

1. Summary of the impact

The International HapMap project was a major international research collaboration to map the structure of common human genetic variation across populations from Europe, Asia and Africa. Mathematical Scientists from the University of Oxford played key roles in the development of statistical methods for the project, along with its overall design and management of the International HapMap Project.

Companies have used HapMap as the primary resource to design genome-wide microarrays to make novel discoveries in, for example, pharmacogenetic studies. The size of this market is estimated at $1.25 billion.

One novel discovery has led to a genetic test that is predictive of sustained viral suppression in patients treated for chronic hepatitis C. An estimated 2.7 to 3.9 million people are affected by HCV infection. This test is sold commercially by the company LabCorp and is a significant contributor to the company's testing volume. Finally, the project has been important in widening the public understanding of genetic variation.

2. Underpinning research

The International HapMap Project was an international collaboration of research institutions from the UK, US, Canada, Japan, China and Nigeria (for details see www.hapmap.org). The project assayed genetic data on a sample of individuals from around the world in order to map the patterns of common genetic variation. The HapMap project had two main strands: data production and data analysis. The data analysis group was co-chaired at different times by Professor Peter Donnelly and Professor Gil McVean and they and other statisticians from the University of Oxford (Dr Jonathan Marchini, Dr Simon Myers) played key roles in developing statistical methods that were used in the project to generate and analyse the data; they made a very substantial contribution to the analysis work. Furthermore, Professor Donnelly sat on the main HapMap committee and played a pivotal role in the feedback from analysis to experimental design.

The human genome consists of a sequence of 3 billion base pairs, but only a small fraction of these positions vary between individuals. The experiments carried out by the HapMap project discovered many millions of these variable positions, known as Single Nucleotide Polymorphisms (SNPs). The analysis of the data provides an understanding of the structure of common human genetic variation, in particular, how SNPs at nearby locations in the genome are arranged into common combinations or haplotypes. The project showed that well-powered genome-wide analyses of SNPs can be carried out by assaying only a fraction of all SNPs in the genome, through exploiting the correlation structure of nearby SNPs.

The main product of the HapMap project was the set of haplotypes on the 270 individuals. Each individual's genome consists of two copies of each chromosome, one from each of their parents. These copies are known as haplotypes. Haplotypes are not directly observed by the genotyping technologies that were used by the project, so statistical methods were needed to infer the haplotypes from the genotypes. In a precursor paper [1], Donnelly and Dr Matthew Stephens (a PDRA) developed a method to achieve this. As the HapMap project progressed, it became necessary to select from a number of competing methods to perform this inference. Marchini and Donnelly led the effort within the project to compare different methods and develop new methods for haplotype estimation. This work was published in the American Journal of Human Genetics in

2006 [2] and the methods of Donnelly and Stephens, as further developed by Marchini and Donnelly, were used to estimate haplotypes in the project's two main papers in Nature [3,4].

In parallel, Myers, McVean and Donnelly developed a method for estimating fine-scale recombination rates from the project data. The recombination maps provided an insight into the underlying evolutionary forces that shape the patterns of genetic variation found in the project data [5]. This work has further led directly to the identification of the first sequence motifs that are associated with hotspot activity in humans and evidence that these same motifs mark sites of recurrent disease-causing genomic rearrangements in humans [6]. The fine-scale genetic maps produced by Myers, McVean and Donnelly and their research groups have been used in effectively all the subsequent genome-wide association studies (GWAS) to delimit regions of associations, and to pinpoint the natural candidate genes underpinning the association findings.

The first phase of the project was completed and published, with very substantial public interest, in 2005 [3]. Donnelly was co-chair of the Analysis Group during the first phase, co-wrote [3] and is joint corresponding author. A second phase of the project, published in 2007 [4], produced genotypes at 3.1 million SNPs from 270 people of European, African, Japanese and Chinese ancestry. McVean was co-chair of the Analysis Group in Phase 2, wrote [4] and is joint corresponding author.

Work was carried out during the period 1999-2008. All the key researchers were at the University of Oxford when the research was carried out. Donnelly has been Professor of Statistical Science since 1996. From 2007 he has been Director, Wellcome Trust Centre for Human Genetics. McVean was a Royal Society University Research Fellow from 2000-2004. Since then he has been a University Lecturer in Mathematical Genetics and has been on secondment as Head of Bioinformatics and Statistical Genetics at the Wellcome Trust Centre for Human Genetics, since 2010. Marchini held a Wellcome Trust Postdoctoral Fellowship from 2002-2005. In 2005 he became a University Lecturer in Statistical Genetics. Myers was a Nuffield Trust Fellow in Medical Mathematics from 2002-2005. He was then a Broad Fellow at the Broad Institute of MIT and Harvard, USA from 2005-2007. Since 2007 he has been a University Lecturer in Bioinformatics.

Stephens was a postdoctoral researcher at the University of Oxford from 1997-2000.

3. References to the research

*[1] M Stephens, N Smith and P Donnelly (2001) A New Statistical Method for Haplotype Reconstruction from Population Data. *American Journal of Human Genetics* 68:978-989. DOI: 10.1086/319501

*[2] J Marchini, D Cutler, N Patterson, M Stephens, E Eskin, E Halperin, S Lin, Z Qin, H Munro, G Abecasis, P Donnelly, and International HapMap Consortium (2006) A Comparison of Phasing Algorithms for Trios and Unrelated Individuals. *Amercan Journal of Human Genetics*, 78 437-450. DOI:10.1086/500808

*[3] The International HapMap Consortium. (2005) A haplotype map of the human genome. *Nature* 437, 1299-1320 DOI:10.1038/nature04226

[4] The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, (7164):851-61. DOI:10.1038/nature06258

[5] GA McVean, SR Myers, S Hunt, P Deloukas, DR Bentley, and P Donnelly. (2004) The fine-scale structure of recombination rate variation in the human genome. *Science.* Apr 23; 304 (5670): 581-4. DOI:10.1126/science.1092500

[6]   SR Myers, C Freeman, A Auton, P Donnelly, G McVean. (2008) A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat Genet.* 40, 1124-1129. doi:10.1038/ng.213

The three asterisked outputs best indicate the quality of the underpinning research. All six papers are in high quality internationally refereed journals.

4. Details of the impact

The HapMap project has led to three areas of impact. Firstly, it has made an economic impact on biotech and diagnostic companies by facilitating the introduction of new products and services, which have led to substantial wealth creation for these companies since 2008 (the official release of HapMap was 20 December 2007, see hapmap.ncbi.nlm.nih.gov). Secondly, the project has led to a new diagnostic test that has been widely adopted that can guide doctors and patients when considering drug treatment for hepatitis C virus. Finally, the project has had an impact on society by increasing the public interest in, and engagement with, science. Specific details of these impact areas are as follows.

Economic impact: facilitating biotech companies to develop new products
Prior to the HapMap Project, genetic variants at known locations in the genome were typically analysed, in both research and clinical labs, through small-scale experiments. Such technologies did not scale (in terms of cost or throughput) to genome-wide analysis. The HapMap Project demonstrated that most of the common genetic variation in the genome could be 'captured' through selected use of a few hundred thousand SNPs. Realising the major potential for such genome-wide products, a number of companies developed technologies for enabling massively parallel genotyping. These included Perlegen Sciences (which was awarded the contract to genotype the Phase 2 SNPs for the project) and Illumina Inc. (currently a NASDAQ listed company). Products from Illumina Inc. were specifically designed from the HapMap project (for example, the HumanOmniExpress chip [A], [B]) so as to maximise the power of studies using these 'SNP chips' (a chip is a collection of microscopic DNA spots attached to a solid surface). Their customers include genomic research centers, pharmaceutical companies, academic institutions, clinical research organizations and biotechnology companies. The Associate Director for Scientific Research at Illumina states [B] "*This letter is to outline the utility of the HapMap project as a vital resource for developing the products used for genome wide association studies (GWAS). The array portion of the GWAS market is estimated around $1.25 billion USD over the past 5 years. The data generated by the HapMap project was the primary resource used to develop these arrays.* […] *Illumina's current line of GWAS array products includes the OmniExpress which consists of over 700k of SNP content derived solely from the HapMap.* […] *Without the data available from the HapMap project these arrays would be significantly less powerful for detecting regions of the genome association*". The number of individuals genotyped using chips that trace back to the HapMap project is well over 1 million [C].

Health impact: Providing a framework for discoveries of genetic risk factors that have made clinical impacts
The HapMap Project did not directly analyse the genetic contribution to human disease. However, it provided a framework for the wealth of discoveries about the genetic contribution to common complex disease and pharmacological risk via the GWAS approach. Evidence of the success and scope of GWAS is documented at the National Human Genome Research Institute GWAS Catalog [C] which lists 1,449 published GWA in 237 different diseases and traits. For many diseases GWAS has led to the discovery of multiple disease genes. These discoveries have led to a greater understanding of the disease etiology, and functional work that might lead to a clinical impact is still ongoing. As one example, the details of a GWAS discovery that has led to a substantial clinical impact are given below.

A certain polymorphism of the IL28B gene was found in individuals infected with the most common type of hepatitis C virus (HCV), HCV genotype 1, to aid in identifying those patients who are twice as likely to eliminate the HCV virus on a sustained basis when treated with pegylated interferon-

ribavirin combination therapies [D]. This study used the Illumina Human610-quad BeadChip that was designed using the HapMap resource.

A test for these genetic variants is sold by LabCorp [E, F] and used by clinicians when treating HCV and in clinical trials. HCV is the most common chronic blood-borne infection in the US. The Senior Vice President for Science and Technology at LabCorp confirms [E] "*I understand that the Oxford team were key players in the HapMap project, which underpins downstream genome-wide association studies, and that you are interested in a practical example of where association studies have led to clinical impact.* […] *The IL-28B test is a significant contributor to the testing volume we do in the area of pharmacogenetics and has been since its launch in 2010, continuing to today. Recent estimates indicate that an estimated 2.7 to 3.9 million people are affected by HCV annually.*"

Impact on society: widening the public understanding of human genetic variation
The last ten years has seen a major growth in the public interest and understanding of genetic variation, both in relation to disease, but also more generally in relation to ancestry and origins. The publication of the first phase of the HapMap Project was widely reported in international non-specialist media including, in the UK, interviews with Donnelly on Newsnight and Radio 4's Today programme [G]. The contribution of the HapMap Project was acknowledged in the House of Lord's report into Genomic Medicine [H] (published in 2009; Donnelly was interviewed as part of the committee enquiries), which in turn has resulted in the founding of the Human Genomics Strategy Group [I], which advises the government and NHS on how genomics can be integrated into a national healthcare programme.

5. Sources to corroborate the impact

[A]   Product description for Illumina HumanOmniExpress BeadChip that describes the use of the HapMap project data in product design. Copy held by University of Oxford.
www.illumina.com/documents/products/datasheets/datasheet_human_omni_express.pdf

[B]   Letter from the Associate Director for Scientific Research at Illumina that confirms how the HapMap Project data was used in the design of Illumina's SNP-chips. Copy held by University of Oxford.

[C]   National Human Genome Research Institute (NHGRI) Genome-Wide Association (GWA) Catalog, which provides evidence of the success and scope of GWAS. The catalog lists 1,449 published GWAS in 237 different diseases and traits. www.genome.gov/26525384

[D]   Main paper on the discovery of the IL28B variants. This study used the Illumina Human610-quad BeadChip that was designed using the HapMap resource:

Ge, D., Fellay, J., Thompson, A. J., Simon, J. S., Shianna, K. V., Urban, T. J., et al. (2009). Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature*, *461*(7262), 399–401. doi:10.1038/nature08309

[E]   Letter from the Senior Vice President for Science & Technology at LabCorp providing details of the IL-28B test that they sell. Copy held by University of Oxford.

[F]   Press release on LabCorp's IL28B test. Copy held by University of Oxford.

[G]   Non-academic media coverage of the publication of the first phase of the project. See, for example, news.bbc.co.uk/1/hi/health/4378624.stm. Copy held by University of Oxford.

[H]   Discussion of the role of the HapMap Project in medical genetics. House of Lords report on Genomic Medicine (2008-2009). Copy held by University of Oxford.

[I]   Human Genomics Strategy Group Report. Copy held by University of Oxford.