

Institution: University of Southampton
Unit of Assessment: 01 Clinical Medicine
Title of case study: 01-09 New gene mapping tools
<p>1. Summary of the impact</p> <p>Research carried out by the University of Southampton into the genetic causes of diseases, and the gene mapping techniques and applications derived from this research, has benefited patients worldwide through improved prediction, diagnosis and treatment for common diseases with a complex genetic basis. A particularly striking example is age-related macular degeneration which is a common cause of blindness. Commercially, the research provides cost-effective strategies for genotyping DNA samples, and marker-based selection strategies for economically relevant animal species, such as cattle. The work underpins the development of the personal genomics industry, which specialises in individual genetic risk profiling.</p>
<p>2. Underpinning research</p> <p>With the race to find the genetic causes of diseases gathering pace, there is less focus on diseases caused by rare single gene mutations, such as cystic fibrosis, and more on common diseases with a complex genetic basis, such as breast cancer and Type 2 diabetes. Research at the University of Southampton, led by Professors Newton Morton (1988- retired in 2010) and Andrew Collins (Professor of Genetic Epidemiology and Bioinformatics, 1989-date), has delivered gene mapping techniques which enable the identification of specific chromosome sections where the causes of diseases lie.</p> <p>Collins and Morton published an early paper on allelic association mapping which adapted the Malecot model for genetic isolation by distance to accurately fine-map a Mendelian gene (for cystic fibrosis) as proof of principle. The model was extended for association mapping in 'common' disease using single nucleotide polymorphisms (SNPs) [3.1].</p> <p>Morton and Collins were the first to quantify the advantages and increased power achieved by case-control (contrasting disease cases with disease-free individuals) versus family-based gene mapping strategies [3.2]. This work facilitated rapid adoption of this simpler, more powerful and cost-effective format by many research groups, including the Wellcome Trust Case Control Consortium, which identifies disease genes in Genome-Wide Association Studies (GWAS: testing many genetic variants in different individuals to find associations with disease).</p> <p>Collins and Morton were the first to show cost-effective GWAS can be achieved using a small panel of marker SNPs (just ~1% of all SNPs) because of extensive linkage disequilibrium (LD, the population association between genetic markers in close proximity on a chromosome) [3.1]. Their early findings of extensive LD were in stark contrast to the (incorrect) assertions of other influential authors and informed the design and development of the 'HapMap' project enabling construction of genome-wide SNP screening panels.</p> <p>Collins and Morton developed LD unit (LDU) maps that represent patterns of LD and, with commercial funding from Applied Biosystems, the LDMap software for their construction [3.3]. They demonstrated that LDU maps increase power for gene mapping and are invaluable for characterising human population structure. LDU maps are used in research projects by many different groups internationally including Nelson Freimer (UCLA) [3.4], Leena Peltonen (Finland) [3.4] and Alec Jeffreys (Leicester). Collaborative work with the Freimer group revealed the value of isolated populations for cost-effective gene mapping with reduced genetic heterogeneity.</p> <p>Using LDU maps, Morton and Collins identified remarkably numerous and extensive regions of homozygosity in outbred (and not just isolated) populations [3.5]. They demonstrated the persistence of unbroken ancestral haplotypes in regions with low recombination rates. This work opened a new international research area exploiting homozygosity mapping in outbred populations to identify novel recessive disease genes.</p> <p>Collaboration with the University of Sydney (Professors Frank Nicholas, and Herman Raadsma) underpinned development of SNP panels for marker-based selection in the bovine genome, enabling commercial applications.</p>

Impact case study (REF3b)

The many applications of case-control based association carried out at the University of Southampton, exploiting LDU maps for mapping genes underlying common diseases, include identification of significant metabolic genes in large birth cohorts [3.4] and genes causing age-related macular degeneration (AMD) [3.6].

3. References to the research

3.1 Collins, A., Lonjou, C. and **Morton, N.E.** (1999) Genetic epidemiology of single nucleotide polymorphisms. *Proc Natl Acad Sci USA* 96(26), 15173-7. (279 citations).

3.2 Morton, N.E. and **Collins, A.** (1998) Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci USA* 95, 11389-11393. (248 citations in GS).

3.3 Maniatis, N., **Collins, A.**, Ku, X-F., McCarthy, L.C., Hewett, D.R., Tapper, W., Ennis, S., Ke, X., **Morton, N.E.** (2002) The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by diplotype analysis. *Proc Natl Acad Sci USA*, 99(4), 2228-2233. (164 citations).

3.4 Sabatti C, Service SK, Hartikainen A-L, Pouta A, Ripatti S, Brodsky J, Jones CG, Varilo T, Kaakinen M, Sovia U, Ruokonen A, Laitinen J, Jakkula E, Coin L, Hoggart C, Elliot P, **Collins A**, Turunen H, Gabriel S, McCarthy MI, Daly MJ, Järvelin M-R, Freimer NB, Peltonen L (2008). Genomewide association analysis of metabolic phenotypes in a birth cohort from a founder population. *Nature Genetics*, 41:35-46. (314 citations).

3.5 Gibson, J., **Morton, N.E.**, **Collins, A.** (2006) Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics* 15 (5), 789-95. (114 citations).

3.6 Ennis S, Jomary C, Mullins R, Cree A, Chen X, MacLeod A, Jones S, **Collins A**, Stone E, Lotery A (2008). Association between the SERPING1 gene and age-related macular degeneration: a two-stage case-control study. *Lancet*, 372:1828-34. (102 citations).

Grants

1998-2003. **Collins A.** Medical Research Council, Career Establishment Grant+supplement: Integration of maps to support disease gene mapping. £456,568.

2001-2005. **Morton NE** and **Collins A.** Medical Research Council. Linkage disequilibrium in the human genome. £157,259.

2004-2005. **Collins A.** Applied Biosystems, Foster City, California. Linkage disequilibrium maps. £59,466. [**Commercial funding for development of LDU maps used in their SNPbrowser™ software.**]

2004-2006. **Collins A** and **Morton NE.** BBSRC. Linkage disequilibrium maps for human populations. £274,260.

2004-2008. **Morton NE** and **Collins A.** NIH. Development and application of methods to map disease genes by allelic association. ~£300,000.

2006-2009. Lotery A, **Collins A**, Ennis S. Macular Vision Research Foundation. Linkage disequilibrium mapping of genes associated with age-related macular degeneration. ~£150,000

2011-2013. Eccles D, **Collins A**, Tapper W. Breast Cancer Campaign. Does inherited genetic variation influence breast cancer biology and prognosis?. £128,222.

4. Details of the impact

The research by Collins and Morton (CAM) has substantial impact in translational medicine, contributing new ways to predict disease and diagnose and treat patients. The work enables creation of new, cost-effective, technologies, commercial exploitation of livestock genomes and development of the personal genomics industry.

These methods provide a basis for >450 genome-wide association studies undertaken worldwide since 2008 which have identified many novel disease variants and opened new avenues of research highlighting unanticipated disease pathways and mechanisms. The research has identified disease genes underlying cancer, metabolic traits, ophthalmic traits and others, enabling risk prediction and therapeutic interventions.

Age-related macular degeneration (AMD), the most common cause of blindness in developed

Impact case study (REF3b)

countries, is an excellent example. The mapping by GWAS (Genome-Wide Association Studies) of a key gene, Complement Factor H (CFH) paved the way for identification of related genes. Application of CAM's gene mapping methods, in collaboration with Southampton Professor of Ophthalmology, Andrew Lotery, identified the SERPING1 gene and established a virtually complete understanding of the genetic causes of AMD by 2010 [5.1]. The impact of an individual's genetic makeup is now quantified as risk to develop AMD, and at least five commercial genetic testing kits now predict patient risk [5.2], with clinical trials of genetic therapy underway [5.3]. About 20% of the population is at risk of AMD and genetic models have 83% predictive value [5.2].

The work enables commercial development of cost-effective strategies for genotyping DNA samples in association studies, including GWAS. Powerful GWAS with very large sample sizes are therefore economically feasible and identify genetic factors underlying different diseases, improving disease prediction and identifying potential routes to therapy. The power of CAM's work was recognised by Life Technologies, a US company which funded the development of LDU maps for incorporation in their SNPbrowser™ software [5.4]. With subsequent LDU map updates the browser reached its full commercial potential from 2008, continuing to the present day. The software enables effective genotyping strategies in cases and controls for the identification of disease genes.

The methodologies developed by CAM are applicable in commercially significant animal genomes. Collaboration with the University of Sydney enabled development of cost-effective commercial genotyping of DNAs from dairy cattle, focusing on the identification of superior cattle strains to increase milk yields [5.5]. A direct result of this research led to the development of LDU maps for cattle which enabled the construction of non-redundant 'reduced' SNP panels which have practical application in genetic marker-based selection. Beneficiaries of commercial livestock genotyping are primarily cattle breeders – an important market in Australia alone, which is the world's second largest beef exporter. As a result of this work several companies focus on genetic profiling of livestock. One example is the Neogen Corporation which has a new GeneSeek® Genomic Profiler™ to maximise the genetic potential of stock animals to increase profitability [5.6]. Neogen reported \$6.6 million net income for the third quarter of the 2013 financial year.

The personal genomics industry, which has developed since 2008, has become possible through disease gene identification in GWAS for which CAM's research was pivotal [5.7]. Personal genomics is concerned with the analysis of individual genomes and particularly the characterisation of genomic variation known to be implicated in disease. One example is the 23andme service and offers genetic risk profiles for more than 40 diseases, with the aim of detecting risk profiles early, motivating lifestyle changes and focussing medical screening. Users of these new commercial services include insurance companies and individuals interested in their own risk profiles.

An additional impact of the research is through contribution to pharmacogenomics, which deals with the influence of genetic variation on drug response in patients. Profiling is important to avoid dangerous adverse drug responses and wasteful prescription of medicines unsuitable for a specific patient. Collated information available on the genetic basis of drug response has become available via the PharmGKB database [5.8]. This provides data on around 170 drug-gene relationships that are valuable in clinical practice. The case of AMD, described above, is an example in which a patient's genetic profile strongly determines disease risk and provides invaluable information for tailoring screening and treatment regimens. The database identifies AMD genetic variants mapped by association studies/GWAS as targets for the drug Ranibizumab (Lucentis) which has been approved to treat 'wet' AMD. The ultimate beneficiaries of the research, the patients can, therefore, be treated more effectively using this Southampton-led approach to personalised/stratified medicine.

In summary, these methods enable the mapping of genes involved in human diseases with significant impact on disease prediction and the development of personalised medicine. The work also underpins genetic profiling in cattle and other genomes of commercial importance. The continuing development of disease and trait genomics will have far-reaching impact in the coming years.

5. Sources to corroborate the impact

5.1 Genes identified using methodologies developed by Collins and Morton for case-control based GWAS, include the SERPING1 gene. It is one of six AMD genes which account for 45% of the risk of developing age-related macular degeneration (87% population attributable risk). Ref.: Gibson J, Cree A, **Collins A**, Lotery A, Ennis S, 2010, "Determination of a gene and environment risk model for age-related macular degeneration". *Br J Ophthalmol*, 94(10):1382-1387). Metabolic trait related genes identified (reference 3.4) include high-density lipoprotein with NR1H3, low-density lipoprotein with AR and FADS1-FADS2, glucose with MTNR1B, and insulin with PANK1.

5.2 The mapping of many of the genes underlying AMD has enabled the development of commercial kits for genetic prediction such as RetnaGene <http://www.sequenomcmm.com/AMD> (Sequenom). Ref.: Hageman GS et al., Clinical validation of a genetic model to estimate the risk of developing choroidal neovascular age-related macular degeneration. *Hum Genomics*. 2011; 5(5). The Macula risk model has 83% predictive value and stratifies individuals into five categories with increased risk of AMD, representing 20% of the population:

http://www.revoptom.com/continuing_education/tabviewtest/lessonid/107647/

5.3 An example of a clinical trial program (2010) for AMD is Oxford Biomedica, a gene therapy company in the U.K., which has received FDA authorization to launch a clinical trial of its RetinoStat® gene therapy for the treatment of wet age-related macular degeneration:

http://www.blindness.org/index.php?option=com_content&view=article&id=2360:clinical-trial-for-wet-amd-gene-therapy-to-begin-in-december-2010&catid=64:macular-degeneration&Itemid=120

5.4 The SNPbrowser™ software now owned by Life Technologies, version 4.0 was published on March 25 2012: <http://www.mybiosoftware.com/population-genetics/5641>. It is a freely available tool enabling knowledge-guided selection of over six million TaqMan® or SNiPlex™ System SNP Genotyping Assays, including 650 million genotypes generated for over three million SNPs validated by the International HapMap Project or Applied Biosystems in five major populations. It includes visualization of SNPs integrated with the physical genome maps and haplotype block information. The tool features the linkage disequilibrium unit (LDU) maps developed by Collins and Morton to enable selection and purchase of cost-effective SNP panels.

5.5 Collins established a strong collaboration with the University of Sydney Cooperative Research Centre (CRC) for Innovative Dairy Products. The CRC is a seven-year, \$80 million, research consortium set up by the dairy industry and the Commonwealth Government, and involves a number of Australia's leading research institutes and dairy companies:

<http://www.ausbiotech.org/directory/details.asp?companyid=%7B295B6313-9CFA-4912-9CEC-82642C6BABC7%7D&returntourl=%2Fdirectory%2Fsearch.asp%3Fpg%3D15>. The collaboration

established the first LDU maps for cattle and the first comprehensive understanding of the LD structure of cattle genomes. Ref.: Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. MS Khatkar, FW Nicholas, **AR Collins** et al., *BMC genomics*, 2008, 9 (1), 187.

5.6 The LDU maps of cattle, and the development of reduced panels of marker SNPs, enable genetic-marker based selection for traits of importance. Companies which now exploit marker-based selection include: the Canadian Dairy Network which established commercial genetic testing in 2012: <http://www.ciaq.com/news/news-items/2012/alliance-gives-dairy-producers-access-to-new-genetic-testing-services.html> and Neogen which has a comprehensive cattle genomic test: <http://www.neogen.com/Corporate/PR2013/2013-05-06.html>

5.7 Our development of association mapping methods for the identification of genes underlying disease has enabled the development of personal genomics companies such as 23andMe (<https://www.23andme.com/>) whose personal genome test kit was named "Invention of the Year" in 2008: "TIME's Best Inventions of 2008". Time magazine. 2008-10-29. The company is aiming to achieve a one million customer base in 2013.

5.8 The Pharmacogenomics Knowledge Base (PharmGKB) <http://www.pharmgkb.org/> provides data on ~170 drug-gene relationships that are valuable in clinical practice. Web service for pharmacogenetic study published Gong L, Owen RP, Gor W, Altman RB, Klein TE. (2008). PharmGKB: an integrated resource of pharmacogenomic data and knowledge. *Curr Protoc Bioinformatics*. Chapter 14:Unit14.7