



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Genomics of common diseases

**Citation for published version:**

Wright, AF 2011, 'Genomics of common diseases: approaching the tipping point' *Genome Medicine*, vol. 3, no. 10, pp. 70. DOI: 10.1186/gm286

**Digital Object Identifier (DOI):**

[10.1186/gm286](https://doi.org/10.1186/gm286)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Genome Medicine

**Publisher Rights Statement:**

© 2011 BioMed Central Ltd. Europepmc open access link

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## MEETING REPORT

# Genomics of common diseases: approaching the tipping point

Alan F Wright\*

### Abstract

A report on the Wellcome Trust Scientific Conference 'The Genomics of Common Diseases 2011', held at the Wellcome Trust Conference Centre, Hinxton, Cambridge, UK, 30 August to 2 September 2011.

The study of common human diseases has made considerable progress since 2007 when the first 'The Genomics of Common Diseases' conference was held. There are now over 1,300 genome-wide associations reported between SNPs and human diseases or related traits (<http://www.genome.gov/GWASStudies>). Despite some initial skepticism, these results are now beginning to open doors on new disease pathways and mechanisms, but it remains to be seen whether the next phase of the Human Genome Project, utilizing massively parallel next-generation sequencing to discover rare genomic variants, will have even more explanatory power, as some have predicted. The 2011 conference was held at the Wellcome Trust Genome Campus in Hinxton and many talks focused on whether this new technology will help to turn the corner in complex diseases, by elucidating their heritability, most of which is currently unexplained, and providing further mechanistic insights.

### Limitations of genome-wide association studies

To the extent that genome-wide association studies (GWASs) are not providing all of the answers, keynote speaker Mark Daly (Broad Institute, USA) pointed to the genetic architecture as the main culprit. Most complex traits are likely to be underpinned by both common and rare variants, but the role of the latter has been relatively untouched by GWASs. As the architectural fog begins to lift, there appears to be two extreme scenarios: some diseases involve a small number of genetic pathways,

while others appear to show an almost limitless complexity. An example is in GWASs of inflammatory bowel disease (IBD) that, in the course of 12 years, have revealed >100 susceptibility loci, an extraordinary achievement, although still only explaining about 25% of the heritability. Despite the apparent complexity, the identified genes predominantly fall into two main functional categories: innate immunity and autophagy. However, the preliminary results of whole genome sequencing in IBD patients and controls is throwing up rare variants in new genes, as well as in known genes such as *CARD9*, where a whole spectrum of allele frequencies and effect sizes is apparent. In contrast, the diversity of rare copy number and other structural variants in neuropsychiatric disorders, such as schizophrenia or autism, suggests a very large number of pathways and mechanisms. Previous GWASs in these disorders have been disappointing. In schizophrenia, for example, there were few genome-wide significant hits and the results pointed to thousands of common alleles of very small effect (odds ratio: approximately 1.1 or less), consistent with a polygenic model, but identifying no clear pathways or mechanisms. However, the sample sizes may simply have been too small to overcome the genetic heterogeneity and small allele effect sizes involved. Recent studies using substantially more cases and controls (>50,000 individuals) look more promising, with the suggestion of a neural development pathway involving the *MIR137* microRNA and some of its targets.

### Searching for rare variants associated with common diseases

Several speakers reported new susceptibility loci, harboring an excess of rare variants, found by whole genome or exome sequencing in patients and controls at the phenotypic extremes. Sekar Kathiresan (Massachusetts General Hospital, USA) described exome sequencing of 1,100 patients with early-onset myocardial infarction and 1,100 'hypernormal' controls. They used 'allelic burden' tests to detect clustering of individually rare functional variants within candidate genes and also 'imputation' (genotype prediction) into GWAS datasets to replicate the results. An excess of apparently functional rare

\*Correspondence: [Wright.alan.wright@hgu.mrc.ac.uk](mailto:Wright.alan.wright@hgu.mrc.ac.uk)  
MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine,  
Crewe Road, Edinburgh EH4 2XU, UK

variants (<1% to 5% frequency) was found both in novel susceptibility genes (for example, *CHRM5*, *DKK2* and *LRIG2*) and in known genes (for example, *PCSK9*). Replication of rare variant associations is demanding, however, requiring tens of thousands of individuals. Goncalo Abecasis (University of Michigan, USA) described exome sequencing those at the extremes of the low-density lipoprotein cholesterol (LDL-C) distribution in an isolated Sardinian population (approximately 6,000 individuals). So far, rare variants have been found mostly in known genes, such as *LDLR*, *SORT1* and *APOB*. Cristen Willer (University of Michigan, USA) and colleagues are also exome sequencing those at the LDL-C extremes (top and bottom 1%) but in this case from four general population-based cohorts. They also found an excess of rare variants at one or other extreme, mostly in known genes, suggesting either that a point of diminishing returns has been reached or that extreme samples from much larger populations are needed to identify new pathways.

Michael Boehnke (University of Michigan, USA) described a similar 'extremes' genome sequencing strategy in diabetes, using leaner, younger and family-history-positive type 2 diabetics compared with heavier, older, ancestry-matched controls. He also discussed a family-based sequencing project, which highlighted the advantages of families not only for establishing haplotypes and confirming heterozygous sequence variants but also for establishing phenotype-genotype correlations. As many carriers of a rare variant (frequency  $10^{-4}$ ) could be identified in a single family as in a screen of 15,000 unrelated individuals. Currently, only about 10% of the heritability of type 2 diabetes has been explained by the >50 GWAS associated genes but it remains to be seen whether rare variants explain more of the heritability.

The role of rare variants in trait architecture was also raised by Richard Durbin (Wellcome Trust Sanger Institute, UK), who summarized progress with the 1,000 Genomes and UK10K Projects. Cladograms of the major human population expansions over the past 10,000 years show exponential growth and remote common ancestry, whereas isolate populations show recent common ancestry, facilitating the search for shared haplotypes, long-range phasing and imputation of rare variants identified by whole genome sequencing. Exponential growth also leads to an excess of rare or private mutations, which may therefore figure prominently in the genetic architecture of many traits.

### The problem of locus heterogeneity

Several speakers emphasized another aspect of genetic architecture: locus heterogeneity (many genes influencing the trait). David Goldstein (Duke University, USA) described studies of the sustained virological response (SVR)

following hepatitis C virus infection, in which GWASs identified the *IL28B* gene, explaining half of the difference in SVR between Europeans and Africans. A similar study of HIV-1 resistance showed no genome-wide significant associations but a follow-up whole genome sequencing study is pointing to several promising candidates. However, the sample size being sequenced may be critical to reliably detect an allelic burden, because of the extreme locus heterogeneity.

Next-generation sequencing is also facilitating a quiet revolution in Mendelian genetics. Debbie Nickerson (University of Washington, USA) described the advantages of exome over whole-genome sequencing, namely larger allelic effect sizes, easier interpretation and functional follow-up, lower cost and computational savings. Sequencing a small number of well-chosen individuals can produce disproportionate information on physiological as well as disease mechanisms. The role of the *DLX5/DLX6* pathway and its important role in jaw development in humans (auriculocondylar syndrome), mice and zebrafish was discovered in this way. Exome sequencing a small number of parent-child trios has also facilitated the discovery of highly penetrant *de novo* mutations. Nickerson's central message was again the need for strategies to circumvent the locus heterogeneity, including the use of isolates, families with extreme phenotypes, trait or disease subgroups, or intermediate quantitative traits.

### Determining the functional role of new variants

One of the problems raised by next-generation sequencing is evaluating the functional significance of newly identified rare variants. Nicholas Katsanis (Duke University, USA) presented an elegant approach to resolve the functionality of candidate genes and variants using morpholino-induced translation suppression in zebrafish. He described its use in systematically resolving which of the 29 candidate genes at a locus influencing head size was causal (*KCTD13*). The approach has limitations, however, and seems to be most suitable for anatomical traits.

The role of genomic sites regulating gene expression was discussed by Len Pennacchio (Lawrence Berkeley National Laboratory, USA) and John Stamatoyannopoulos (University of Washington, USA). Pennacchio described the use of chromatin immunoprecipitation (ChIP) sequencing P300 transcriptional coactivator targets to specifically detect enhancer sequences. These represent a substantial (20% to 50%) subset of all enhancers but show relatively poor sequence conservation despite being active across species as diverse as human and mouse. They are also hard to mutate, suggesting the importance of secondary structure. Stamatoyannopoulos presented the results of systematically surveying

regulatory sites marked by DNase I hypersensitivity (DHS) in a broad range of cells and tissues (<http://www.roadmapepigenomics.org/>). There are over 2.4 million human DHS sites, which commonly influence chromatin states and are exposed whenever a gene is being expressed in a particular cell type. A majority of GWAS-identified SNPs are either in DHS sites or strongly associated with them. Many of these sites cluster within physiologically relevant transcription factor binding sites. Stamatoyannopoulos emphasized that the GWAS 'story' in common diseases is more concerned with identifying such pathways than with individual, small effect susceptibility genes. An old question is perhaps finally being answered: the polygenic basis of complex traits may principally result from such expression differences.

### **Are we reaching the 'tipping point' for genomics of common diseases?**

The field is fast moving but, like an old daguerreotype photograph that only captures solid unmoving objects, a few conclusions can be drawn. Complex traits consist of an unpredictable mix of common, uncommon and rare genetic variants with effect sizes that are inversely related to their frequency. Only members of the first group (common variants) have been well captured by GWASs and about three-quarters of these occur in non-coding regulatory regions. The associated genetic pathways have

accordingly been hard to discern, leading to patchy progress in unraveling pathogenetic mechanisms. Next-generation sequencing of whole genomes or exomes is now beginning to capture the uncommon and rare, if not the very rare, variants influencing common diseases. New arrays with a higher proportion of low frequency variants will help to replicate the results of genome sequencing studies. These resources should also allow more researchers to reach the 'tipping point' or critical sample size required to overcome the extent of locus and allelic heterogeneity and small individual genetic effect sizes that are an emerging feature of most complex traits. Time and technology will tell but one thing we can be sure of is a tsunami of genomic sequence data over the next few years.

#### **Abbreviations**

DHS, DNase I hypersensitivity; GWAS, genome-wide association study; IBD, inflammatory bowel disease; LDL-C, LDL cholesterol; SNP, single nucleotide polymorphism; SVR, sustained virological response.

#### **Competing interests**

The author declares that he has no competing interests.

Published: 28 October 2011

doi:10.1186/gm286

**Cite this article as:** Wright AF: Genomics of common diseases: approaching the tipping point. *Genome Medicine* 2011, **3**:70.