

Meeting report

Complex disease genetics: present and future translational applications

Michael V Holmes*, Sonia H Shah†, Aspasia Angelakopoulou‡, Tauseef Khan‡, Daniel Swerdlow*, Karoline Kuchenbaecker*, Reecha Sofat§ and Tina Shah§

Addresses: *Genetic Epidemiology Research Group, UCL Research Department of Epidemiology and Public Health, University College London, 1-19 Torrington Place, London, WC1E 6BT, UK. †UCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, Kathleen Lonsdale Building, Gower Place, London, WC1E 6BT, UK. ‡Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK. §Centre for Clinical Pharmacology, Division of Medicine, University College London, Rayne Institute, 5 University Street, London, WC1E 6JJ, UK.

Correspondence: Tina Shah. Email: t.shah@ucl.ac.uk

Abstract

A report on the British Atherosclerosis Society autumn meeting 'Genetics of Complex Diseases', Cambridge, UK, 17-18 September 2009.

Introduction

Complex disease genetics is at a critical turning point. Genome-wide association studies (GWASs) have generated an abundance of data, resulting in the use of advanced analytic methods and raising many questions. This common platform has brought together various scientific disciplines, including genetics, epidemiology, bioinformatics, statistics and medicine, reflected in the diverse backgrounds of speakers and delegates at this meeting. Here, we summarize two principal themes that emerged in the meeting: first, the success of GWASs in the discovery of novel disease loci using emerging new analytical methodologies, and second, the current and future translational applications of GWASs.

Genome-wide association studies - discoveries, limitations and future directions

GWASs enable a hypothesis-free approach to finding novel genes associated with diseases and traits. Facilitated by the HapMap project (<http://www.hapmap.org>), chips with probes for up to one million single nucleotide polymorphisms (SNPs) can be used to capture variation across the entire human genome. Mark Caulfield (Barts and The London School of Medicine, London, UK), Sekar Kathiresan (Massachusetts General Hospital and Broad Institute, Boston, USA) and Nilesh Samani (University of Leicester, UK) communicated results on novel loci arising from recent GWASs conducted on cardiovascular diseases

(CVDs). They also highlighted the importance of collaborative analyses in reliably identifying signals that might otherwise be missed owing to small sample sizes. This was exemplified by the finding of novel genes associated with blood pressure and the discovery of SNPs on chromosome 1 associated with CVD that alter low-density lipoprotein (LDL)-cholesterol.

Generation of large volumes of data brings with it analytical challenges, resulting in methodological development. Bayesian approaches that enable direct comparison among SNPs both within and between studies were described by David Balding (Imperial College, London, UK). These methods are in contrast to classical ('frequentist') methods, which compute a *P*-value as evidence for association without incorporating any information about minor allele frequency (MAF) and study size, both factors that affect the power of the test. Hence, the same *P*-value computed at different SNPs or in different studies may not provide the same level of confidence for true association. To partly avoid this issue, it has become the norm to discard low-MAF SNPs when using classical methods, but this may result in detectable association being discarded. In Bayesian analysis prior knowledge is incorporated into the model. The outcome is the posterior probability of association, which can be directly compared between SNPs and studies and also avoids the problem of multiple testing.

John Whittaker (London School of Hygiene and Tropical Medicine, London, UK) further described how the Bayesian approach allows meta-analysis to be performed at the gene level rather than just at the SNP level. This approach facilitates the pooling of data from different studies of the same gene that have investigated a partially overlapping

CNV, copy number variation; CVD, cardiovascular disease; GWAS, genome-wide association study; HDL, high-density lipoprotein; LDL, low-density lipoprotein; MAF, minor allele frequency; MI, myocardial infarction; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes.

range of SNPs, by incorporating information on linkage disequilibrium between SNPs (Newcombe *et al.*, *Am J Hum Genet* 2009, **84**:567-580). This helps to enhance power and also facilitates inference on likely causal variation that could be used to inform functional experiments.

Unexpected GWAS findings were another focus: the first generation of GWASs did not show the expected 'Manhattan skyline' of dense signals, and loci reported by GWASs so far explain only a small proportion of the observed phenotypic variation. Speakers discussed approaches to overcome this and thus to account for the 'missing heritability'. These include fine mapping around GWAS hits, deep sequencing, identification of rare variants, gene-centric approaches, identification of copy number variations (CNVs), gene-gene and gene-environment interactions, and epigenomics.

Peter Donnelly (University of Oxford, UK) presented fine-mapping approaches of loci uncovered by the initial survey done by the Wellcome Trust Case-Control Consortium (WTCCC1; <http://www.wtccc.org.uk>) for CVD, type 2 diabetes (T2D) and autoimmune thyroid disease. Re-genotyping with denser SNP coverage, paralleled with analysis using Bayesian statistical tools, reduced the number of likely causal signals to single SNPs in some cases, though not all; for the association between the fat mass and obesity-associated (*FTO*) gene and body mass index, almost all SNPs in the region had an equal chance of being or marking the causal site, and for the transcription factor 7-like 2 (*TCF7L2*) gene and T2D the signals could be reduced to just two SNPs. The success of such re-genotyping depended on the strength of the initial signal and on the successful tagging of causal variants. The ongoing 1000 Genomes Project (<http://www.1000genomes.org>) was discussed by several speakers as an extension to fine-mapping approaches and as a project that will aid further identification of common and rare variants.

Brendan Keating (Penn Cardiovascular Institute, University of Pennsylvania, Philadelphia, USA) and others presented alternative approaches to fine mapping by capture of variants not always represented on GWAS platforms. Such gene-centric chips enable denser coverage of genes known to be associated with CVD and of genes with high biological plausibility of association. The current IBC Cardiochip (<http://bmic.upenn.edu/cvdsnp>) includes denser SNP coverage of approximately 435 genes (with MAFs of over 0.02) than current GWAS platforms; the next generation, the 200K Cardio-MetaboChip, was also described.

Alex Blakemore (Imperial College, London, UK) presented her experience of re-genotyping CNVs in a subset of an initial T2D GWAS from WTCCC1. Technologies for reliable typing, replication and analysis of CNVs were described as lagging behind those for SNPs. Donnelly presented

ongoing CNV analyses from WTCCC1, suggesting that a large number of SNPs can already be used to type stable CNVs (75% of CNVs with MAFs of over 10%, $r^2 > 0.8$); he also reiterated current pitfalls associated with the computation platform used.

This year's Hugh Sinclair Lecture was delivered by Leena Peltonen (Sanger Centre, Hinxton, UK), who described the unique properties of highly conserved populations for mining the genome for complex traits. Finland's integrated health care system provides detailed and uniform data used by scientists to delineate genetic contributions to disease. She gave the specific example of *GLE1*, which encodes a nuclear-pore-associated mRNA export factor, and fetal motor neurone disease. Population-based cases and controls facilitated functional analysis that revealed expression patterns of *GLE1* in the anterior motor neurons. The principle of reverse genomics and its role in identifying CNVs in patients with cognitive deficits was also described. Peltonen concluded by forecasting that genetics may form the basis of future health care decision-making, but first the functional annotation of the genome through large international consortia using richly characterized, prospective cohorts to expose novel genes relevant for human health is required. She added that the future of GWASs may lie in identifying rare (perhaps population-specific) high-impact variants and more complex structural variants, characterizing effects of genes and lifestyle, and extensive DNA functional studies.

Describing the newly established BGI-Hong Kong, Jun Wang (Beijing Genome Institute, Shenzhen, China) reported on the sequencing of extreme genotypes (such as the Giant Panda Genome Project, the Extreme Environment Animals Genome project and the Asian Human Genome project). Work from his institution has identified 5 Mb of novel sequence in humans that vary across populations, their presence mapping to population migration patterns. Future work will focus on epigenomics (especially the human methylome), gene expression, imprinted genes, network analyses and metagenomics.

Using cancer as an example, Phil Stevens (Sanger Centre) presented data on characterizing structural variation by massive parallel sequencing technology to identify structural rearrangements in cancer cell lines using genome-wide screening approaches. This enables the identification and characterization of base-pair-level deletions, tandem duplications, inverted duplications, inversions and inter-chromosomal rearrangements, including fusion genes, as well as providing copy number information.

Application of genome-wide association studies - are we there yet?

Steve Humphries (University College London, UK), proposing the motion 'this house believes that genetic testing

for CVD has clinical utility', described key variants in the LDL receptor that are currently used in diagnosis and management of familial hypercholesterolemia. He argued the case for the use of common polymorphisms in diagnosis and risk prediction at a population level, describing the incorporation of at least three SNPs identified from GWASs into risk models as a way of increasing their predictive utility for CVD. This is important given that existing algorithms fail to capture 86% of all events (with a 5% false positive rate). He further explained that 7% of the population may have eight or more CVD risk alleles, conferring a CVD odds ratio (OR) of 1.8, similar to smoking (OR = 2).

Tom Dent (PHG Foundation, University of Cambridge, UK), opposing the motion, established the analytical framework for assessing the predictive or diagnostic utility of a test. Discriminative risk scores (such as the Framingham and QRisk scores) were not enhanced by adding SNPs. Potential harms of population screening were described, for example the withholding of beneficial drugs that have the same relative risk reduction in CVD irrespective of genotype.

Aroon Hingorani (University College London) described translational applications of CVD genetics, contrasting the individual approach ('personalized' medicine through prediction of disease risk and response to treatment) with the population approach ('impersonal medicine' through public health improvements). In terms of prediction, he used the prevention paradox as an example, in which most events occur at normal levels of causal (and putative) risk factors. This means that single SNPs with an OR of around 2 are unlikely to be individually informative either in diagnosis or prediction. Rather, a combination of variants (a 'polygene') may perform better, although the cost of this would be the high numbers that would need to be screened. Common variants in CVD may be useful for reliably identifying those at risk of early CVD events, and therefore for enabling preventative strategies. Extending the idea of the use of genetics in public health, Hingorani described potential applications of large-scale genetic data in informing the drug development pipeline. The random allocation of common variation in genotypes (Mendelian randomization) can be used as a proxy for a therapeutic drug trial, which he illustrated with the example of the cholesteryl ester transfer protein gene *CETP* Taq1B variant and the *CETP* inhibitor torcetrapib. Such genetic data may be used with other lines of evidence to avoid, for example, late-stage, high-cost failure of new drugs.

Continuing the theme of application of genetics to human health, Philippa Talmud (University College London) described the incorporation of 20 genes from recent GWASs into two T2D risk algorithms to predict how many individuals in the Whitehall II cohort were reclassified

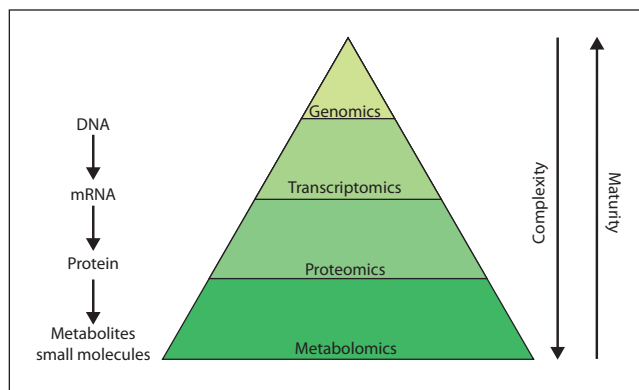


Figure 1

From genome to metabolome: genomics is a relatively mature field but the more complex of the 'omic' or high-throughput technologies, such as systems medicine and metabolomics, are not yet as mature. The pyramid represents the interface between the increasing numbers of interactions at each level.

from low to high T2D risk and vice versa. Addition of these SNPs resulted in more individuals being inappropriately reclassified (net reclassification of -4.7%). Suggested reasons for the lack of improvement included that only 3% of the genetic contribution to T2D has been identified and that further SNPs associated with insulin resistance and fasting glucose measurements were in the pipeline (the presently known SNPs are associated with pancreatic beta-cell function). Very large cohorts (such as the UK Biobank) were described as essential to reliably identify gene-environmental interactions.

Sekar Kathiresan presented work on probing the causality of lipids in myocardial infarction (MI). After an allelic 'dosage score' was constructed, the mean differences in several lipid traits between top and bottom quintiles of the score and their effect on MI risk were compared. For both high-density lipoprotein (HDL) and triglycerides, observed risk was less than predicted. He also discussed the lack of association between SNPs and MI due to pleiotropy (for example, the ATP-binding cassette transporter protein *ABCA1* decreases both HDL and LDL, with a net observed risk of 1.0) and the hypothesis that perhaps only some mechanisms that increase HDL also affect atherosclerosis.

Conclusions

GWASs have facilitated the discovery of novel disease-associated loci, an approach validated by identification of known loci in tandem with new ones. Despite the limitations, GWAS data will inform the next generation of studies and catalyze development of necessary technology and analytical methods. Together these will reveal biological pathways and networks with the hope of benefiting human health through applications in personal and public health.

Anna Dominiczak (University of Glasgow, UK) encapsulated these thoughts in her concluding talk with an inspirational forecast of the integration of genomics with transcriptomics, proteomics and metabolomics (Figure 1). Using peptidomics as an example, she described novel work in which renally excreted peptides differ in patients with CVD, with high discrimination, and ongoing work comparing urinary peptide profiles with coronary angiogram findings. The interface of genetics with downstream experimental fields (Figure 1) may pave the way for future mechanistic studies to elucidate how hits from GWASs exert their effects, filling in the many gaps in our knowledge.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

MVH is funded by a Population Health Scientist Fellowship from the Medical Research Council (G0802432). DS is funded by a Medical Research Council Doctoral Training Award. RS is supported by a British Heart Foundation (Schillingford) Clinical Training Fellowship (FS/07/011). We thank A Dominiczak for allowing reproduction of Figure 1.

Published: 5 November 2009

doi:10.1186/gm104

© 2009 BioMed Central Ltd