Atherosclerosis 208 (2010) 599–602

Contents lists available at ScienceDirect

Atherosclerosis

journal homepage: www.elsevier.com/locate/atherosclerosis

Meeting report

A report on the Genetics of Complex Diseases meeting of the British Atherosclerosis Society, Cambridge, UK, 17–18 September 2009☆

1. Introduction

The study of complex disease genetics is at a critical turning point. Development of high-throughput technology has enabled a change in the investigational approach from traditional candidate gene studies to hypothesis-free genome-wide association studies (GWAS). These experiments have generated an abundance of data, novel analytic methods, some answers but also many questions. The aim of this year's Autumn meeting of the British Atherosclerosis Society was to update members on research in complex disease genetics. This common platform has brought together a number of scientific disciplines including genetics, epidemiology, bioinformatics, statistics and medicine, all of which were reflected in the diverse backgrounds of speakers and delegates at this meeting.

Two principal themes emerged from this meeting and are summarised here. Firstly the success of GWAS in discovery of novel disease loci with emerging novel analytical methodologies to interpret this abundance of data (and how to overcome their limitations) and secondly, the translational application of the information gathered so far from GWAS, and what we can expect in the future. An abbreviated version of this meeting report was first published in *Genome Medicine* [1].

2. Genome-wide association studies – discoveries, limitations and future directions

GWAS are today's currency of complex disease genetics, allowing an unbiased approach to finding novel genes associated with diseases and traits by typing common single nucleotide polymorphisms (SNPs) across the entire human genome in large populations. The use of up to one million-SNP chips in current GWAS was described, largely facilitated by the HapMap project cataloguing common variants and allowing tagging approaches to capture variation across the 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 of a number of such GWAS from the field of cardiovascular disease, and highlighted the validity of the approach through the discovery of known as well as novel loci. They as well as a number of other presenters emphasized the importance of collaborative analyses; this was heralded as a marked change in the landscape of common disease genetics. Given the moderate effect size of SNP associations for complex disease thus far (predictable from the nature of the underlying allele frequency), analyses across multiple studies together have allowed reliable detection of signals that would otherwise be missed owing to sample size. This was exemplified by recent successes in the finding of novel genes associated with blood pressure [2], facilitated only by meta-analyses across 34,433 individuals, and furthermore discovery of novel SNPs on chromosome 1 associated with coronary heart disease (CHD) that alter LDL-cholesterol level [3].

Generation of this volume of data brings with it analytical challenges, which have facilitated discourse between the fields of statistics, bioinformatics and genetics, resulting in new analytical approaches to preliminary and joint data analyses. Bayesian methods were described by David Balding (Imperial College, London, UK) and John Whittaker (London School of Hygiene and Tropical Medicine, London, UK), which complement the current classical ("frequentist") methods. The latter test the null hypothesis of no association between genetic variants and the phenotype (trait or disease) under investigation. However, the same *p*-value for different SNPs' associations within and across studies does not necessarily provide a comparable level of confidence that an association is true, as this is dependent on the underlying SNP and study characteristics (sample size and minor allele frequency). Bayesian approaches provide an additional step that can allow direct comparison among SNPs within and between studies, permitting rational approaches to comparing results across studies (meta-analysis) [4].

Despite their success, a number of limitations of GWAS were discussed with potential strategies to overcome these. The first generation of GWAS hoped for a Manhattan skyline of dense signals, however this was not the case. This apparent shortcoming alone is not problematic if SNPs discovered account for a large proportion of the variation in a trait or disease. However, loci reported by GWAS thus far explain only a small proportion of the observed phenotypic variation. The classic case quoted by many was that of height, where even though estimates of heritability are between 80 and 90%, the approximately 40 variants identified by GWAS in over 30,000 individuals explained only about 5% of height's heritability [5]. A number of speakers discussed alternative approaches that could account for this "missing heritability". These included fine mapping and next generation sequencing to further delineate functional variants within disease- or trait-associated loci, the search for rarer variants and gene-centric approaches (which are currently missed by HapMap based common variation SNP chips). Other structural variations such as copy number variations (CNVs, large blocks of deletions or duplications) were a focus for several speakers who highlighted the analytical challenges faced in this field. Finally, several groups are also investigating effects that, owing to their interaction with other genetic or environmental factors, might not be identified by using standard single-locus tests. These are discussed in more detail below.

☆ This article is a re-publication of an abbreviated version, originally published in *Genome Medicine* [1]. *Genome Medicine* remains copyright holder.

0021-9150/$ – see front matter
doi:10.1016/j.atherosclerosis.2009.11.034

Peter Donnelly (University of Oxford, UK) presented fine mapping approaches of loci uncovered by WTCCC1 [6] for coronary artery disease (CAD), type 2 diabetes (T2D) and autoimmune thyroid disease. Fourteen regions were re-genotyped using denser SNP coverage and analysed using newly developed Bayesian statistical tools, narrowing down two associated to single SNPs, with a further three regions showing multiple signals for independent association (i.e. >1 causal SNP in the sequenced region). The success of these methodological strategies was contingent on the strength of the initial signal and successful tagging of causal variants. This work not only highlighted novel approaches for delineation of causal variants, but also the successful application of Bayesian analyses. In addition to this example a number of speakers alluded to the ongoing 1000 Genomes Project [7] undertaking the re-sequencing of 1000 genomes of individuals from diverse ethnic backgrounds. The project hopes to capture variation that has not been possible with HapMap, particularly rarer variants and variants occuring at different frequencies between ethnic groups. These may give more insight into disease-associated variants.

Brendan Keating (Penn Cardiovascular Institute, University of Pennsylvania, Philadelphia, USA) and others presented alternative approaches to fine mapping and capture of variants not always represented on GWAS platforms through the use of gene-centric chips. The first of these in use is the IBC 50K Cardiochip [8]. The premise of such an approach is for example, genes that are known to be associated with CHD (such as *APOE*) and others with a high biological plausibility of association are not represented in detail on current GWAS platforms. However, chips that home in on these putative loci are able to make a more detailed investigation of their contribution to disease. The current IBC Cardiochip includes a dense SNP coverage in ~435 genes, and encompasses variants which have a minor allele frequency (MAF)>0.02. Results from groups using the platform are currently awaited, with anticipated insights into variants that may account for a larger proportion of the variability in CHD associated traits. The next generation of this chip will be the 200K Metabochip, which is hoped to overcome the limited tagging of genes in its predecessor.

Structural variation and in particular CNVs were the topic of a number of speakers' communications. Alex Blakemore (Imperial College, London, UK) painted a realistic picture of the difficulties in detecting and analysing CNVs and disease associations by posing a number of questions: Which CNVs to prioritise for genotyping? Do CNVs exist? Is the association real? Can it be replicated? How may they exert their biological effect? She addressed these using examples from data aiming to re-genotype CNVs in a subset of the initial T2D GWAS. Overall, she showed that technologies to reliably type, replicate and analyse CNVs is lagging behind that of SNPs, and expectations from the field as a whole for CNVs may be less than were initially thought. Peter Donnelly presented ongoing CNV analyses from WTCCC1, suggesting that in fact a large number of SNPs can be used to type stable CNVs already (75% of CNVs with MAF>10%; $r^2 > 0.8$), although reiterating the current computation platform associated pitfalls.

One further approach was described by Leena Peltonen (Sanger Centre, Hinxton, UK), who delivered this year's Hugh Sinclair Lecture. Using the example of properties of highly conserved populations she described how these could be used to mine the genome for complex traits. Drawing on her experience of research in Finland she described how an integrated health care system which provides detailed and uniform documentation could be used by scientists to delineate genetic contribution to disease, giving the specific example of *GLE1* and foetal motor neurone disease [9]. Association analyses from population-based cases and controls facilitated subsequent functional analysis revealing expression patterns of *GLE1* in the anterior motor neurones, and delineation of function in mRNA export and translation. She summarised the principle of reverse genomics and its role in identifying large deletions in the DNA (CNVs) of patients with cognitive deficits. Professor Peltonen concluded that genetics may provide the basis for future health care decision-making, but that this first requires the functional annotation of the genome, which will occur through large international consortia using richly-characterised, prospective cohorts to expose novel genes relevant for human health. She added that the future of GWAS may be in identifying rare (perhaps population-specific) high impact variants, more complex structural variants, characterising the effects of genes and lifestyle and extensive DNA functional studies.

An insight to technologies and services that will facilitate in overcoming limitations was summarised in captivating communications by Jun Wang (Beijing Genome Institute, Shenzhen, China) and Phil Stephens (Sanger Centre). Describing the newly established BGI-Hong Kong, Jun Wang related the success of the Beijing Genomics Institute in sequencing extreme genotypes: for example completion of the Giant Panda Genome, launch of the extreme environment animals genome project as well as the Asian human genome [10]. He highlighted the importance of information from diverse ethnic populations and how these would illustrate the nature of structural variation within the genome and differences between these genomes per se. Work from his institution has identified 5 Mb of novel sequence in humans that vary across populations, their presence mapping to population migration patterns. The sequence is also conserved in chimpanzee, macaque and mouse, and such efforts have enabled them to estimate that individuals' genomes may vary from one another by ~4.8–8 Mb. He alluded to future work to expect from this energetic facility, staffed by individuals with an average age of 25, in the fields of epigenomics (particularly the human methylome), gene expression, imprinted genes, network analyses and metagenomics. He gave practical examples of his work including the gene sequencing of stool in patients with obesity and Crohn's disease, and the effects of altitude on increased expression of genes promoting stem cell differentiation into erythrocytes. Using cancer as an example, Phil Stevens presented data on characterising structural variation by massive parallel sequencing technology to identify structural rearrangements in cancer cell lines using genome-wide screening approaches. Thus far this technology enabled the identification and characterisation of base-pair level deletions, tandem duplications, inverted duplications, inversions and interchromosomal rearrangements, including fusion genes as well as obtaining copy number information, and may be applied to other diseases and traits in the future.

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

The second theme of this meeting is how discoveries from GWAS can be applied to human health now and in the future. The debate "This house believes that genetic testing for cardiovascular disease (CVD) has clinical utility" established some key concepts early on in the meeting, by those both for and opposing the motion, for the utility of genetic testing for CHD in its current state in the UK. The latter was highlighted for by Steve Humphries (University College London, UK), proposing the motion, describing the identification of key variants in the LDL receptor which have been genotyped and are used in the diagnosis and management of familial hypercholesterolaemia. He further presented evidence of how SNPs identified from GWAS could be incorporated into common CHD risk models on a population basis, firstly explaining that existing algorithms based on phenotypic measures are poor at predicting CHD in middle-aged men in the UK, missing 86% of CHD events when a 5% false positive threshold is employed [11]; addition of 3 SNPs to existing algorithms increases the area under the ROC curve from 0.62

to 0.7 [12], and using a risk allelic dosage score, 7% of the population will possess 8 or more risk alleles, conferring an odds ratio (OR) of 1.8 (similar to smoking, OR = 2). Professor Humphries stressed that since many SNPs from recent CHD GWAS tag genes of unknown function, they may associate with quantitative traits already included in existing algorithms and thus may be of questionable clinical utility.

However, a more cautionary note was presented by Tom Dent (PHG Foundation, University of Cambridge, UK), opposing the motion. He first established the analytical framework used to assess the predictive or diagnostic utility of a test (genetic or otherwise), and the importance of making changes in healthcare policy evidence-based. He distinguished between current discriminative powers of risk scores (such as Framingham and QRisk), which are in themselves poor, and cited evidence which demonstrated SNPs identified by GWAS did not add to the discriminative power of these models. The prevailing evidence base for complex disease such as CHD would therefore not support population-based testing of common variants. He cited specific examples from the literature to support his argument [13,14]. He proceeded to describe potential harms of such applications, which could mean individuals who would otherwise be eligible, would not receive statins, thus denying them of CHD risk reduction. This is at odds with current observational and randomised trial evidence, in that lipid lowering across the full range of LDL-levels will result in the same proportional risk reduction, and given that most events occur at average levels of cholesterol and LDL, such testing would probably mean that individuals who would benefit from treatment would not receive it. He was, however, optimistic for the future, suggesting that risk prediction, and perhaps prevention, may be the major utility of genetics.

His sentiments were echoed later in the meeting by Aroon Hingorani (University College London) who described the translational applications of CVD genetics contrasting the individual to the population. Whereas in the individual, genetics can help predict disease risk and response to treatment ("personalised" medicine), in the population, its utility lay more in improvements in public health (through identifying the cause of disease) and through drug development ('impersonal' medicine). Professor Hingorani described the difficulty in predicting CVD due to the normal distribution of causal biomarkers (such as BP) across the population, and the associated linear/log-linear risk and that most cardiovascular events occur in those at the 'normal' part of the distribution curve (Rose's prevention paradox). As most SNPs conveying risk of CVD have an OR of <2, the key may lie in combining variants (polygene), however, the number needed to screen to detect one person carrying 10 variants could be >1 million! He described the emerging shift away from measurement of risk factors towards their modification (e.g. through the proposed panacea of the polypill), and suggested that the clinical application of genetics may lie in predicting those who have early incident CVD (proposing the genetic prediction of "time to event"). Genotype as a model for predicting drug response through the principles of mendelian randomisation was described, and its applications in improving the late-phase attrition rate of emerging drugs, citing work with torcetrapib as an example.

The lack of the predictive utility of SNPs was further described by Philippa Talmud (University College London) using the example of T2D. Her team examined the calibration of two risk algorithms, the Framingham Offspring risk score, and the Cambridge risk score, both based on several blood markers, in the Whitehall II cohort of civil servants. Twenty SNPs identified from recent GWAS (most of which associated with beta cell function and insulin secretion) were incorporated in a gene score model. With addition of genes, the net reclassification was −4.7%, suggesting that use of those SNPs does not enhance the predictive utility at present. Suggested reasons included that only 3% of the genetic contribution to T2D has been identified (thus much 'missing heritability' remains, possibly to be found in copy number variation, differential methylation or epigenetic effects and rare mutations of large effect) and that further SNP associated with insulin resistance and fasting glucose measurements were in the pipe-line. She mentioned that gene-environmental interactions may be important in diabetes – and the need for very large cohorts such as the UK Biobank to reliably identify them.

Sekar Kathiresan mentioned similar limitations when probing the causality of lipids in myocardial infarction. After constructing an allelic "dosage score", the mean differences in lipid trait between top and bottom quintiles of the score and their effect on MI risk were compared. For both HDL and triglycerides, the observed risk was less than predicted as not all SNPs were associated with MI risk. Some of the explanation for the null association of SNPs include pleiotropy (e.g. *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.

Need for the inclusion of recent advances in genetics and genomics in medical curricula was made by Professor Humphries, highlighted in a recent UK House of Lords Science and Technology Committee report. The technologies and discoveries of the last decade are little known to many of those teaching the doctors of tomorrow. Within postgraduate education, and perhaps especially applicable to CVD (in which a large proportion of care is delivered by general practitioners) continuing professional education should focus on training in the concepts of complex disease genetics. With a field advancing at such rapid pace and the true implications of discoveries still uncertain, the educational challenge is considerable.

4. Conclusions

What did we learn about the current landscape in complex disease genetics? GWAS have facilitated the discovery of novel disease-associated loci, an approach validated by tandem identification of known loci, moving the field of complex genetics to the next level. Despite limitations, GWAS data will inform the next generation of studies and catalyse the development of necessary technology and analytical methods. Together these will reveal biological pathways and networks with the hope of benefitting human health and the optimism of genetics as a benevolent tool when applied to personal and public health will be realised.

Anna Dominiczek (University of Glasgow, UK) encapsulated these thoughts in her concluding summation of the meeting with an inspirational forecast of the integration of genomics with transcriptomics, proteomics and metabolomics (Fig. 1). Using peptidomics as an example, she described novel work in which peptides excreted in

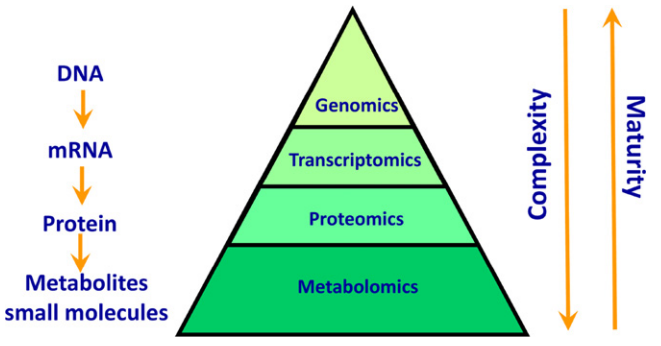


Fig. 1. From genome to metabolome: genomics is a relatively mature field but the more complex of the 'omic' or high-throughput technologies, such as systems medicine and metabolomics, are not yet as mature. The pyramid represents the interface between the increasing numbers of interactions at each level.

the kidney differ in patients with CAD [15] with high discrimination (area under the ROC curve 0.97) and how this is being extending into clinical studies to compare urinary peptide profiles with coronary angiogram findings. She highlighted that interface of genetics with the downstream experimental fields may pave the future for mechanistic studies to elucidate how GWAS hits exert their effects, filling in the many gaps in our knowledge.

Acknowledgements

Michael V. Holmes is funded by a Population Health Scientist Fellowship from the Medical Research Council (G0802432). Daniel Swerdlow is funded by a Medical Research Council Doctoral Training Award. Reecha Sofat is supported by a British Heart Foundation (Schillingford) Clinical Training Fellowship (FS/07/011).

We thank Professor Dominiczak for allowing reproduction of Fig. 1.

References

[1] Holmes MV, Shah SH, Angelakopoulou A, et al. Complex disease genetics: present and future translational applications. Genome Med 2009;1(11):104.
[2] Newton-Cheh C, Johnson T, Gateva V, et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet 2009, in press.
[3] Linsel-Nitschke P, Heeren J, Aherrahrou Z, et al. Genetic variation at chromosome 1p13.3 affects sortilin mRNA expression, cellular LDL-uptake and serum LDL levels which translates to the risk of coronary artery disease. Atherosclerosis 2009, in press.
[4] Newcombe PJ, Verzilli C, Casas JP, et al. Multilocus Bayesian meta-analysis of gene-disease associations. Am J Hum Genet 2009;84(5):567–80.
[5] Maher B. Personal genomes: the case of the missing heritability. Nature 2008;456(7218):18–21.
[6] Wellcome Trust Case Control Consortium. Available at: http://www.wtccc.org.uk.
[7] 1000 Genomes Project. Available at: http://www.1000genomes.org.
[8] Keating BJ, Tischfield S, Murray SS, et al. Concept, design and implementation of a cardiovascular gene-centric 50k SNP array for large-scale genomic association studies. PloS One 2008;3(10):e3583.
[9] Nousiainen HO, Kestila M, Pakkasjarvi N, et al. Mutations in mRNA export mediator GLE1 result in a fetal motoneuron disease. Nat Genet 2008;40(2):155–7.
[10] Wang J, Wang W, Li R, et al. The diploid genome sequence of an Asian individual. Nature 2008;456(7218):60–5.
[11] Cooper JA, Miller GJ, Humphries SE. A comparison of the PROCAM and Framingham point-scoring systems for estimation of individual risk of coronary heart disease in the Second Northwick Park Heart Study. Atherosclerosis 2005;181(1):93–100.
[12] Humphries SE, Yiannakouris N, Talmud PJ. Cardiovascular disease risk prediction using genetic information (gene scores): is it really informative? Curr Opin Lipidol 2008;19(2):128–32.
[13] Kathiresan S, Melander O, Anevski D, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med 2008;358(12):1240–9.
[14] Talmud PJ, Cooper JA, Palmen J, et al. Chromosome 9p21.3 coronary heart disease locus genotype and prospective risk of CHD in healthy middle-aged men. Clin Chem 2008;54(3):467–74.
[15] Zimmerli LU, Schiffer E, Zurbig P, et al. Urinary proteomic biomarkers in coronary artery disease. Mol Cell Proteom 2008;7(2):290–8.

Michael V. Holmes
Genetic Epidemiology Research Group, UCL Research Department of Epidemiology and Public Health, University College London, 1 - 19 Torrington Place, London, WC1E 6BT, UK

Sonia H. Shah
UCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, Kathleen Lonsdale Building, Gower Place, London, WC1E 6BT, UK

Aspasia Angelakopoulou
Tauseef Khan
Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK

Daniel Swerdlow
Karoline Kuchenbaecker
Genetic Epidemiology Research Group, UCL Research Department of Epidemiology and Public Health, University College London, 1 - 19 Torrington Place, London, WC1E 6BT, UK

Reecha Sofat
Tina Shah*
Centre for Clinical Pharmacology, Division of Medicine, University College London, Rayne Institute, 5 University Street, London, WC1E 6JJ, UK

* Corresponding author.
E-mail address: t.shah@ucl.ac.uk (T. Shah)

19 November 2009
Available online 26 November 2009