EDITORIAL

Genome-wide association studies: past, present and future

When the time came to decide on the topic for the next in the series of *Human Molecular Genetics* reviews, on this occasion, there was little need for debate. The advent of genome-wide association (GWA) technology has transformed the landscape of human genetic research. It has enabled those in the field to move beyond the limitations of small-scale candidate gene studies, and well over 200 loci influencing a wide range of complex phenotypes have now been identified. Although, for many conditions, these variants still explain only a small proportion of individual differences in disease predisposition, there is a growing confidence that identifying most of the remaining genetic variance now represents a finite and tractable challenge. We may not yet have all the technology, the samples or the analytical power that we need, we may not even agree where is the best place to start looking for the ‘missing’ heritability, but recent developments have shown that none of these obstacles is likely to be insurmountable. The reviews in this issue, therefore, as well as describing some recent achievements of GWA studies, set out some of the opportunities and challenges with which the field is presented as it seeks to expand susceptibility-gene discovery, and, crucially, to translate those findings into advances in clinical management.

Three of the disease areas in which there has been the greatest yield of novel complex trait-susceptibility genes from GWA studies include metabolic/cardiovascular, autoimmune and cancer. In their summary of advances in the metabolic and cardiovascular arena, Mohlke *et al.* (1) detail over 50 novel loci now known to modify individual risk of type 2 diabetes and cardiovascular disease, to influence circulating levels of lipids or to alter energy balance and thereby body mass index and potential for obesity. There has been a similar explosion in respect of autoimmune diseases, and Lettre and Rioux (2) summarize how the total numbers of loci implicated in predisposition to celiac disease, inflammatory bowel disease, multiple sclerosis, rheumatoid arthritis, lupus or type 1 diabetes have more than quadrupled in the past 3 years, with several of those loci predisposing to more than one autoimmune disease. GWA studies into cancer predisposition tell the same story: Easton (3) documents over 20 new loci for breast cancer, prostate cancer, colorectal cancer or melanoma discovered in the last 2 years. The findings in cancer also provide an early glimpse of the possible complexity of susceptibility loci, with multiple alleles at the same 8q24 locus influencing the risk of different combinations of cancers.

The search for further association signals is taking the field in a variety of directions. One dominant theme at the moment is the use of meta-analysis approaches to combine GWA data from multiple samples informative for the same phenotype of interest, and thereby overcome the limitations of power that can compromise any individual study. The size and complexity of the data, and a variety of sources of potential error and bias, have required a number of challenges to be overcome. De Bakker *et al.* (4) provide a timely exposition on the steps that need to be taken to implement and interpret such a GWA meta-analysis which should be extremely useful for those contemplating such endeavours. Another powerful approach integrates genome-wide data on sequence variation with global transcript profiling information to understand the genetic basis of variation in expression levels. Nica and Dermitsakis (5) describe both the potential power and the current limitations in using gene expression as a phenotype to help ascribe functional annotation to the associated loci.

The GWA approach, as implemented so far, has its limitations, and there remain important sources of bias and error that can complicate inference. Looking back, the danger of generating spurious signals due to population stratification effects was seen as one of the most prominent challenges. That concern was legitimate, but as Tian *et al.* (6) describe, methods have been developed that largely eliminate the danger of false attribution of association. As it turns out, the rich information content of GWA data itself provides the basis for powerful solutions to the detection of, and correction for, subtle structure effects.

One of the explanations for the relatively limited proportion of phenotypic variation ascribable to the variants so far identified is the fact that the GWA studies conducted to date have focused on European-descent samples, and have explored only that part of genome variation captured by the common single nucleotide polymorphisms represented on the commercial genotyping arrays. Cooper *et al.* (7) set out the case for extending these efforts to a wider range of ethnic groups, not only to support the recovery of additional susceptibility variants, but also to drive subsequent fine-mapping efforts. In his review, McCarroll (8) summarizes the state-of-the-art as regards determining the contribution made by copy number variants to disease predisposition: as he shows, this area is rapidly reaching technical and analytical maturity, and is likely to be the focus of much activity in the coming months.

The primary outputs of GWA scans are merely association signals. Rarely is the causal variant revealed by these studies, and at many loci, it is not even possible to make confident assignments of the causal gene. The task of moving from association signal to causal variant, and from causal variant to an appreciation of the molecular and cellular mechanisms involved in generating phenotypic effects, is emerging as
one of the major roadblocks to translation. There are no easy solutions here, but some of the options that groups are pursuing are discussed in a review on these next steps (9).

The high profile of GWA discoveries, in both the scientific and lay media, has heightened the expectations about the capacity of this research to generate tangible translational benefits. There seems little doubt that the biological insights derived from locus identification will feed into new treatments, new biomarkers, new ways of preventing disease, but such translation takes time, and is dependent on first bridging the gap from association signal to mechanism (9). However, the use of information gathered from GWA studies to provide measures of individual disease predisposition through ‘genetic profiling’ approaches requires no such mechanistic understanding, and several companies have made high-profile entries into this particular market. In their reviews, Janssens and van Duijn (10) set out the limitations of profiling based on the currently available set of variants, and Kaye (11) explores legal, ethical and social issues that arise from the direct-to-consumer marketing of such tests. In the final review, Wang and Weinshilboum (12) describe some of the ways in which GWA information can be used to support both the development and the use of pharmaceutical agents, for example, through the identification of the mechanisms involved in the generation of adverse events.

The past 2 years have seen substantial strides in the field of human genetics, with unprecedented numbers of associated loci identified for numerous common diseases and quantitative traits. The achievements have been spectacular, but it is all too clear how much more remains to be done. It will be interesting to look back at this review issue in a couple of years time and see how far we have travelled in overcoming the challenges described in these pages and realizing the expectations raised by these initial successes.

REFERENCES


Mark I. McCarthy1,2,3

1Oxford Centre for Diabetes, Endocrinology and Metabolism University of Oxford, Churchill Hospital Oxford OX3 7LJ UK

Tel: +44 1865857298; Fax: +44 1865 857299

Email: mark.mccarthy@drd.ox.ac.uk

2Wellcome Trust Centre for Human Genetics University of Oxford Old Road, Headington Oxford OX3 7BN UK

3Oxford NIHR Biomedical Research Centre Churchill Hospital Oxford OX3 7LJ UK

Joel N. Hirschhorn4,5,6

4Program in Genomics and Divisions of Genetics and Endocrinology Children’s Hospital Boston, MA 02115 USA

Tel: +1 6179192129; Fax: +1 6177300253

Email: joelh@broad.mit.edu

5Broad Institute of Harvard and MIT Cambridge, MA 02142 USA

6Department of Genetics Harvard Medical School Boston, MA 02115 USA