COMMENTARY

Promise and pitfalls of the Immunochip

Adrian Cortes and Matthew A Brown*

Abstract

Genomewide association studies (GWAS) have proven a powerful hypothesis-free method to identify common disease-associated variants. Even quite large GWAS, however, have only at best identified moderate proportions of the genetic variants contributing to disease heritability. To provide cost-effective genotyping of common and rare variants to map the remaining heritability and to fine-map established loci, the Immunochip Consortium has developed a 200,000 SNP chip that has been produced in very large numbers for a fraction of the cost of GWAS chips. This chip provides a powerful tool for immunogenetics gene mapping.

Genomewide association studies (GWAS) have made a phenomenal contribution to our understanding of common heritable diseases over the past 4 years. Immunogenetics research in particular has been highly successful in identifying large numbers of genetic loci. These findings have greatly advanced our understanding of the basic causes of autoimmune and inflammatory conditions, and have provided a solid foundation for hypothesis-driven research into disease mechanisms. As the boundaries of GWAS have been tested, however, limitations of the approach have become more apparent.

It is clear that a substantial fraction of the heritability of common diseases, even in diseases for which quite large GWAS have been performed, has not been mapped, raising questions as to where the missing heritability lies [1]. Theories regarding the location of the unmapped heritability include: residual unidentified common variant associations (common disease–common variant model), rare variant associations not mapped because they are poorly captured by common tagSNPs (common disease–rare variant model), copy number variants (CNVs), epigenetic effects, gene–gene interactions and gene–environment interactions.

*Correspondence: matt.brown@uq.edu.au
University of Queensland Diamantina Institute, Princess Alexandra Hospital, Ipswich Road, Woolloongabba, Brisbane, Queensland, 4102 Australia

Further, the true associated variants are uncertain for most identified loci – even though GWAS have far better resolution than the linkage studies preceding the GWAS era. Even high-density mapping with common SNPs has in most cases not been able to distinguish an association signal due to direct association with disease risk from an indirect association signal due to linkage disequilibrium effects.

Common CNVs are an unlikely source of much missing heritability. Of the 95 loci known by SNP studies at the end of 2009 to be associated with Crohn's disease and type 1 and type 2 diabetes, only three harbored CNVs that may explain the association [2]. In an extensive study of the role of CNVs in eight common diseases, the Wellcome Trust Case Control Consortium identified just three CNV associations, each of which had already been identified by tagSNP studies [2]. The study concluded that 'common CNVs which can be typed on existing platforms are unlikely to contribute greatly to the genetic basis of common diseases'. Whether epigenetic effects can contribute to heritability of common diseases is unclear, as the evidence for heritable transmission of epigenetic marks from generation to generation is limited in humans [3] – although definitive studies are awaited, and they may be tagged by SNP studies anyway [4]. Most heritability studies report narrow-sense heritability, which is heritability excluding gene–gene interaction; thus gene–gene interaction does not contribute to missing narrow-sense heritability. Gene–environment interaction studies in most diseases are in their infancy, and the contribution of such interactions to heritability is unknown.

Recent modeling studies suggest that the missing heritability lies in a mixture of unmapped common and rare variants [5]. Rare variants may have larger functional effects than common variants, which can only become common in a population if they do not have a significance adverse effect on survival/health, or if they are removed from populations by natural selection. Rare variants may also have higher genetic resolution, helping to pinpoint the key regions underlying genetic associations.

Current genotyping chips used for GWAS are not well suited to either picking up the remaining common variants or identifying rare variants. The sample size required to identify the remaining common variants in

most common diseases once the low-hanging fruit have been identified is massive. For example, a recent meta-analysis of GWAS data on the model phenotype height studied 183,727 individuals and identified 180 loci; these contributed just 20% of the heritable component of height variation [6]. At a rough GWAS genotyping cost of US$250 per sample nowadays, this type of study is clearly unaffordable for most diseases even if there were enough cases available. Most of the remaining common variants are thought to probably be contained amongst the most strongly associated SNPs, however, even if they have not yet achieved definite levels of association.

The current crop of GWAS chips does not identify rare variants very well either. Genotyping companies are now racing to increase rare variant coverage on genotyping chips, but even very high-density chips such as the 5 million SNP chips in the Illumina pipeline will only sample a small fraction of the 3.3 billion bases in the human genome. In the dbSNP database there are currently ~12 million annotated SNPs, and a further 32 million awaiting annotation. Ultimately, this coverage issue will be solved by whole genome sequencing studies, but these remain too expensive for widespread use. Further, the sample sizes required to map rare variants are much higher than for common variants, unless those rare variants have quite large individual effects. Adequately powered rare variant mapping studies using these new, denser, GWAS chips are therefore going to be very expensive.

At least part of the answer to these problems lies in the development of custom genotyping chips such as the Immunochip designed for immunogenetics studies, the Metabochip designed for studying metabolic diseases, and a cardiovascular disease chip [7]. Immunochip is an Illumina Infinium genotyping chip, containing 196,524 polymorphisms (718 small insertion deletions, 195,806 SNPs) designed both to perform deep replication of major autoimmune and inflammatory diseases, and fine-mapping of established GWAS significant loci. Initiated by the Wellcome Trust Case–Control Consortium, Immunochip was designed by a consortium of leading investigators covering all of the major autoimmune and seronegative diseases, many of interest to rheumatological researchers, including rheumatoid arthritis, ankylosing spondylitis and systemic lupus erythematosus, as well as the related autoimmune conditions type 1 diabetes, autoimmune thyroid disease, celiac disease and multiple sclerosis, and the seronegative diseases ulcerative colitis, Crohn's disease, and psoriasis. SNPs for deep replication were also included from the findings of GWAS performed on non-immunological diseases that were studied as part of the Wellcome Trust Case–Control Consortium 2 [8]. For each disease, ~3,000 SNPs were selected from available GWAS data for deep replication,

as well as to cover strong candidate genes. The chip will thus enable deep replication studies to identify which amongst the top-ranked SNPs in GWAS studies are truly disease associated. Further, because these diseases are genetically related, the chip will lead to pleiotropic genes being identified, which are associated with more than one of the diseases for which the chip was designed.

At loci with established disease association, the chip contains all known SNPs in the dbSNP database, from the 1000 Genomes project (February 2010 release), and from any other sequencing initiatives that were available to the consortium. This enables cost-effective fine-mapping of loci for both rare and common variants. This fine-mapping would only be possible otherwise if each individual disease produced custom genotyping chips to investigate their particular disease-associated loci, a much more expensive proposition due to the far smaller production runs this would entail.

The chip also contains a dense set of SNPs in the MHC, which will enable imputation of the major classical HLA loci. Although this approach has been previously validated in white Britons, and in African and non-African samples from the HapMap database [9], further confirmation in additional cohorts is being performed by the Immunochip Consortium. A dense SNP set across the KIR/LILR complex is also included to allow imputation of KIR and LILR alleles. Ancestry informative markers are included to allow identification and control of population stratification effects.

The cost of the Immunochip is far lower than GWAS chips (~US$39/sample) because it has been produced in very large numbers (>150,000 ordered in the initial batch). This has enabled groups to finance genotyping of very large cohorts – for example, the International Genetics of Ankylosing Spondylitis Consortium will complete a case study of 12,000 participants by early next year, something unaffordable should it be attempted using GWAS chips. The Immunochip Consortium are sharing control data that will be available for most ethnic groups; more than 20,000 white European controls are expected to be available. The study sample size will thus be sufficient to map rare variants without blowing the bank.

Weaknesses of the Immunochip approach include the following. The chip is designed for use in white European populations and will therefore be less informative for other ethnic groups, although the chip will still be informative particularly where disease-associated variants and haplotypes are shared between white Europeans and the specific ethnic group studied. Another weakness is that many rare variants have yet to be identified and are thus not represented on the chip. Third, genotyping rare variants is a difficult process – and although early indications are that the chip performs well, a proportion of particularly the rarer variants will

probably not be accurately genotyped by the chip. The Immunochip also does not type rare CNVs, which are not well captured by tagSNP studies. A final weakness is that the chip does not cover the whole genome, and depends on the power of the initial GWAS studies for its marker selection. The chip, particularly for diseases where fewer cases have had GWAS performed, will therefore miss residual associated loci.

The Immunochip will thus enable some very valuable and relatively inexpensive studies. For complex problems, however, there is rarely a single comprehensive solution, and genetics is no exception to this rule. Future progress in gene mapping will probably involve a range of different methods, including GWAS, sequencing, and targeted, informed genotyping strategies such as the Immunochip.

Abbreviations
CNV, copy number variant; GWAS, genomewide association studies; HLA, human leucocyte antigen; KIR, killer-cell Immunoglobulin-like receptor; LILR, leukocyte Immunoglobulin-like receptor; MHC, major histocompatibility complex; SNP, single nucleotide polymorphism.

Competing interests
The authors declare that they have no competing interests.

Published: 1 February 2011

References

1. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM: Finding the missing heritability of complex diseases. *Nature* 2009, **461**:747-753.
2. Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D, Barnes C, Conrad DF, Giannoulatou E, Holmes C, Marchini JL, Stirrups K, Tobin MD, Wain LV, Yau C, Aerts J, Ahmad T, Andrews TD, Arbury H, Attwood A, Auton A, Ball SG, Balmforth AJ, Barrett JC, Barroso I, Barton A, Bennett AJ, Bhaskar S, Blaszczyk K, *et al.*: Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 2010, **464**:713-720.
3. Kaminsky ZA, Tang T, Wang SC, Ptak C, Oh GH, Wong AH, Feldcamp LA, Virtanen C, Halfvarson J, Tysk C, McRae AF, Visscher PM, Montgomery GW, Gottesman II, Martin NG, Petronis A: DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet* 2009, **41**:240-245.
4. Slatkin M: Epigenetic inheritance and the missing heritability problem. *Genetics* 2009, **182**:845-850.
5. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM: Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010, **42**:565-569.
6. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segre AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Magi R, Pastinen T, Liang L, Heid IM, Luan J, Thorleifsson, G, *et al.*: Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010, **467**:832-838.
7. Keating BJ, Tischfield S, Murray SS, Bhangale T, Price TS, Glessner JT, Galver L, Barrett JC, Grant SF, Farlow DN, Chandrupatla HR, Hansen M, Ajmal S, Papanicolaou GJ, Guo Y, Li M, Derohannessian S, de Bakker PI, Bailey SD, Montpetit A, Edmondson AC, Taylor K, Gai X, Wang SS, Fornage M, Shaikh T, Groop L, Boehnke M, Hall AS, Hattersley AT, *et al.*: Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS One* 2008, **3**:e3583.
8. Wellcome Trust Case–Control Consortium 2 [http://www.wtccc.org.uk/ccc2/wtccc2_studies.shtml]
9. Leslie S, Donnelly P, McVean G: A statistical method for predicting classical HLA alleles from SNP data. *Am J Hum Genet* 2008, **82**:48-56.