

## MEETING REPORT

# Understanding complex traits: from farmers to pharmas

Doug Speed\* and David J Balding

### Abstract

A report on the 4th International Conference on Quantitative Genetics (ICQG4), Edinburgh, UK, June 17-22, 2012.

**Keywords** BLUP, complex traits, epistasis, genomic architecture, prediction, quantitative genetics, selection.

In his welcome, Bruce Walsh (University of Arizona, USA) illustrated the breadth of quantitative genetics by asking for a show of hands as he listed various areas of interest. When he reached human genetics, he joked that others should take note of the raised hands, as those people would be buying the drinks later. His jest reflects the feeling of many working on other organisms that their research is underfunded relative to human disease genetics. The fact that the meeting was such a success reflects its timing; many of the major strands of genetics are coming together in a way that has not occurred for decades. There are pressing shared interests, for example, in prediction and in understanding the roles of epistasis and epigenetics; advances made by transferring insights and techniques across discipline boundaries were evident throughout the week. While human geneticists can access funding for sophisticated technologies, they envy the possibilities for extensive experimental design and control of environments available to, say, crop geneticists. Conversely, the success of global collaboration and data sharing, championed in human genetics by the Wellcome Trust, among others, is being replicated in other species: we heard about large-scale international collaborations in mice, cattle, *Arabidopsis*, maize and *Drosophila*.

### Genomic architecture of complex traits

A unifying methodological theme of the conference was the mixed regression model, which includes both fixed effects (individual SNPs or other covariates) and a random polygenic term with a correlation structure dictated by pedigree relatedness. Developed long ago by animal breeding geneticists - we still heard references to the 'animal model' - such models gained a new lease of life in 2006 when plant breeders began to use kinships estimated from marker data: it is sharing of causal variants that matters, and pedigree can give only expected and not realized genome sharing. The approach burst into human genetics in 2010 when the group of Peter Visscher (University of Queensland, Australia) proposed using the small differences in genome sharing between unrelated individuals to capture polygenic effects tagged by SNPs. This opened the way to estimating contributions to heritability tagged by selected SNP-sets, such as those associated with pathways, genomic regions or minor allele frequency (MAF) classes. For example, Cornelis Albers (University of Cambridge, UK) used mixed model analysis to estimate how much heritability could be attributed to open chromatin genomic regions in human blood cells. But the assumptions of mixed model analysis can be violated by environmental effects or forces of selection: Magnus Nordborg (Gregor Mendel Institute, Austria) reported that the latter caused problems for association mapping in *Arabidopsis*.

Visscher started by updating on his earlier work exploiting differences in genome sharing between siblings - which can range beyond even the interval 0.4 to 0.6 - to obtain heritability estimates free of ascertainment bias. He then moved on to the recent 'SNP heritability' work, which has tended to support a view that much of the 'missing heritability' of human complex traits lies in common variants of weak effect. He compared progress in understanding the genetic determinants of human height and schizophrenia, both highly heritable but for which little variance can be explained by current genome-wide association study (GWAS) hits. Both traits appear to be highly polygenic: the proportion of variance attributable to a chromosome is strongly correlated to its size. Both traits show a similar inverse relationship

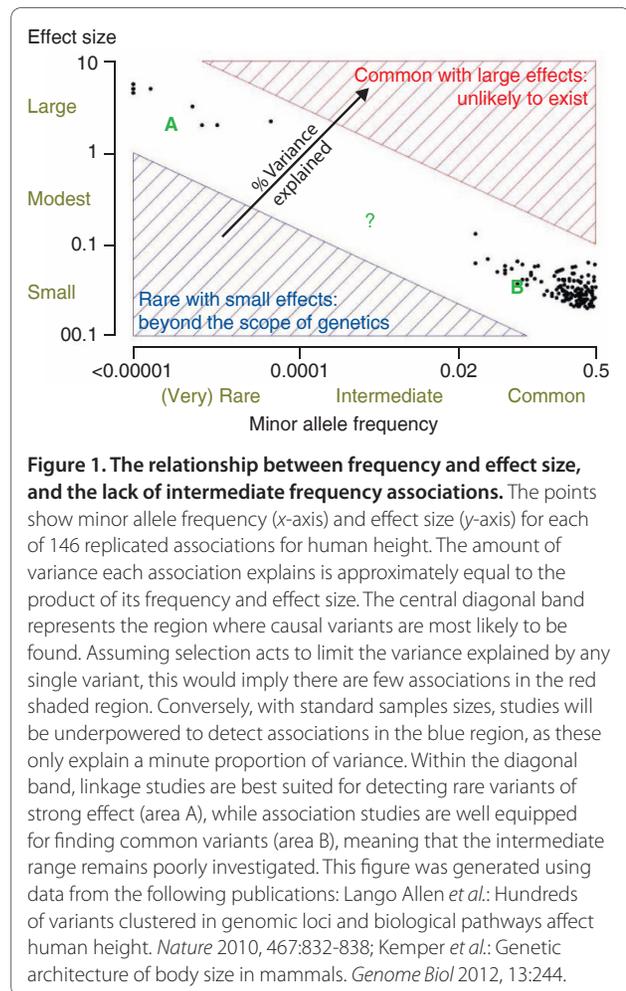
\*Correspondence: doug.speed@ucl.ac.uk  
UCL Genetics Institute, University College London, Gower Street, London WC1E 6BT, UK

between effect size and MAF for validated findings, a pattern consistent with some models of selection against functional variants (Figure 1 demonstrates this relationship for height). Moreover, for both traits there are yet to be found any associated variants of intermediate frequency (roughly between 0.05% and 2%), but hopefully the availability of high-throughput sequencing should begin to fill this gap. Visscher concluded that the only essential difference between progress made on the two traits owed to sample size, which is much higher for height than for schizophrenia.

Rare genetic variants are a topic of much interest in human genetics, and the nature of the MAF-effect size relationship is important for genome-wide prediction models. Sebastian Zöllner (University of Michigan, USA) described a large sequencing study with 14,000 subjects, allowing detailed analyses of rare variation and its implications for study design. Among Europeans, sharing of rare variants declines, as might be expected, with increasing distance (and with Finland showing less sharing) and with decreasing frequency. Matt Hurles (Sanger Institute, UK) gave insights and quantification of the *de novo* mutations that can become rare variants. Richard Durbin (Sanger Institute, UK) promoted the view that quantitative cellular phenotypes are where the action is for genetic association studies, and advances in induced pluripotent stem cells look promising for overcoming the problems of maintaining the quality of cell lines or obtaining fresh, relevant tissue samples. This elicited a question from the floor as to whether GWAS had been barking up the wrong tree!

### Prediction

Genomic selection is currently revolutionizing plant and animal breeding. At its core is prediction of phenotype: even when phenotype is observed, there are advantages in using a predicted 'true' phenotype adjusted for environmental and noise effects. Statistical approaches to prediction from genome-wide SNP data start with the animal model, and much focus remains on BLUP (best linear unbiased prediction) of individual polygenic terms (sometimes called G-BLUP to emphasize that kinship is estimated from SNPs and not the traditional pedigrees). Gustavo de los Campos (University of Alabama, USA) showed that predictive performance depends strongly on the pairwise relatedness between test and training samples: the SNP heritability analysis of Visscher therefore pays a large price in efficiency by using unrelateds, in return for the flexibility of allowing genome partitioning. The success of G-BLUP in plant and animal breeding is not yet reflected in human genetics: response to drugs is an obvious phenotype that it would be enormously helpful to predict, but the larger effective population size of humans, complex breeding patterns and more



**Figure 1. The relationship between frequency and effect size, and the lack of intermediate frequency associations.** The points show minor allele frequency (x-axis) and effect size (y-axis) for each of 146 replicated associations for human height. The amount of variance each association explains is approximately equal to the product of its frequency and effect size. The central diagonal band represents the region where causal variants are most likely to be found. Assuming selection acts to limit the variance explained by any single variant, this would imply there are few associations in the red shaded region. Conversely, with standard samples sizes, studies will be underpowered to detect associations in the blue region, as these only explain a minute proportion of variance. Within the diagonal band, linkage studies are best suited for detecting rare variants of strong effect (area A), while association studies are well equipped for finding common variants (area B), meaning that the intermediate range remains poorly investigated. This figure was generated using data from the following publications: Lango Allen *et al.*: Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010, 467:832-838; Kemper *et al.*: Genetic architecture of body size in mammals. *Genome Biol* 2012, 13:244.

heterogeneous environments mean that predictive accuracy for human traits remains disappointing, as reviewed by Pak Sham (University of Hong Kong, China). More sophisticated statistical modeling is one avenue for improvement. G-BLUP is equivalent to the statistical technique of ridge regression, which in effect assumes a normally distributed effect with the same variance for every SNP. Its simplicity seems to offer scope for more sophisticated regression models with effect sizes that are more realistic and that differ over SNPs, but currently G-BLUP seems hard to beat.

### Epistasis and variance-altering genes

Epistasis (multiple genes affecting a phenotype in a non-additive way) undoubtedly exists in nature, but debate continues as to how important it is in understanding mechanisms and predicting phenotypes. Eric Lander (Broad Institute, USA), in his masterly overview of where we've come in the past two decades, put it in the category of important things that we just don't know. He pointed out that it is now becoming feasible to check since

heritability, measured as a regression of phenotypic correlation on genome-sharing, should give different estimates for different levels of kinship if epistasis plays an important role. Trudy Mackay (North Carolina State University, USA), working with *Drosophila*, reported little consistency between mapped quantitative trait loci (QTLs) in an original and replication experiment, an observation that she attributed to epistasis: a complex interaction system could generate many QTLs, of which only a few might be identified in each replicate. Orjan Carlborg (Swedish University of Agricultural Sciences, Sweden) in his Genetics Society Balfour Lecture described an evolutionary role for genes that alter the variance of a trait rather than its mean, another recurring topic of the meeting. Hugues Aschard (Harvard School of Public Health, USA) described a non-parametric statistical test allowing for any change in phenotypic distribution over genotypes.

### Computational advances

Alan Gray (University of Edinburgh, UK) used some computer science tricks to perform a mixed model analysis on 300,000 SNPs in 9,000 individuals that had taken 17 hours using standard software, in under 15 minutes. Similarly, Gibram Hemani (University of Queensland, Australia) described his use of graphics processing units to provide out-of-the-box and cost-effective computer systems for repetitive low-memory tasks, such as mind-bogglingly large numbers of pairwise tests for SNP interactions: upwards of 10 million tests per second can be achieved. Jun Zhu (Zhejiang University,

China) also unleashed powerful computing resources to analyze network models of SNPs and transcripts.

Overall, the conference affirmed the large advances made over the past decade, and provided optimism for even better things to come. However, there remains no simple answer to the problem of missing heritability: there is no doubt that rare, intermediate and common variants all play some role, together with epistasis and epigenetics, but currently no one of them obviously dominates. There is excitement about the cross-fertilization of ideas among disease, production and evolutionary genetics. All that, together with Edinburgh's beautiful sights, friendly venues and characteristically poor weather means ICQG5, provisionally scheduled for Michigan in 4 years time, has much to live up to.

### Abbreviations

BLUP, best linear unbiased prediction; G-BLUP, genomic best linear unbiased prediction; GWAS, genome-wide association study; MAF, minor allele frequency; QTL, quantitative trait locus; SNP, single nucleotide polymorphism.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

Both authors have been involved in drafting the manuscript and revising it critically, and have given final approval of the version to be published.

Published: 27 July 2012

doi:10.1186/gm360

Cite this article as: Speed D, Balding DJ: Understanding complex traits: from farmers to pharma. *Genome Medicine* 2012, 4:59.