

LETTER

GWAS: heritability missing in action?

European Journal of Human Genetics (2010) **18**, 859–861;
doi:10.1038/ejhg.2010.35; published online 17 March 2010

We have followed with interest the discussion in your journal and elsewhere of the contribution of genome-wide association studies (GWAS) to the elucidation of the aetiology of complex disorders. Although GWAS have been very successful in identifying specific loci and/or genomic regions that contribute to the causation of certain disorders, there has been some disappointment that only a small proportion of the heritability of many conditions has been accounted for.^{1,2} Although fully aware of the inherent limitations of GWAS, we nevertheless believe that the ‘failure’ of this approach may well have been overstated owing to previous misconceptions regarding available measures of heritability.

Back in 2007, Bourgain *et al*³ advised caution against placing excessive expectations on the likely findings of GWAS, suggesting that they might well complement but would not completely supplant family-based linkage studies. Estivill and Armengol⁴ also acknowledged the limitations of contemporary GWAS but hoped that the improved recognition of copy number variants (CNVs) – often in candidate genes or specific chromosomal regions – would overcome some of these difficulties. Subsequent studies of common CNVs of >1 kb in size have not, however, accounted for much more of the ‘missing heritability’.^{1,2} Could rare CNVs conceivably make a significant contribution? This has been suggested by Seng and Seng⁵, who recommended more and larger GWAS rather than a conceptual reassessment. Although Bodmer and Bonilla⁶ opined that rare variants could indeed be of great importance in the context of common (complex) disorders, they advocated a wholly different approach – the study of candidate genes and the development of efficient sequencing strategies.

In his recent personal perspective on GWAS, van Ommen⁷ recommended the massive accumulation of data, without previous hypotheses, as the way forward to solve the puzzle of complex disorders. Others have, however, taken issue with his gentle teasing of the classical Popperian approach to scientific endeavour.⁸ We would like to strengthen the defence of Popper by pointing out that massive data collection merely shifts the point at which hypotheses have to be considered from pre-data collection (at the stage of experimental design) to post-data collection (at the data analysis stage). Popper need not feel too vulnerable just yet!

Several studies have now shown that CNVs can contribute to both schizophrenia and autism, often but not always through *de novo* mutations.^{9–11} Despite the recognition of this latter point, the question remains as to why these conditions are so heritable.¹² One recent review of this topic proposes a set of research strategies to pinpoint the ‘missing heritability’ that appears not to have been accounted for by any of the approaches used so far.¹³

So, where is this ‘missing heritability’? We respond to this question in two different ways. First, we believe that complex disorders are

indeed complex and that genetic studies of complex disorders in humans face a number of challenges including gene–gene and gene–environment interactions and epigenetic modification of the genome. Second, we shall argue that high estimates of heritability have been misinterpreted as showing that a predisposition to such a condition (one with high heritability) must have been transmitted through the family from parent to child. The complexity of these common conditions is apparent from the range of factors that need to be considered as potentially contributing to the ‘missing heritability’. These can be rare variants whose significance is not yet recognised, less uncommon variants of small effect, or common variants of very small effect (very weakly penetrant). The suggested factors that are likely to account for at least some of the missing heritability include the following:

(i) Locus heterogeneity or ‘multiple unilocus’ disorders,¹⁴ with each case resulting from mutation in one gene but, there being many such genes, each only accounting for a small fraction of the affected individuals in the population.

(ii) Gene×gene and gene×environment interactions. If these factors do indeed account for much of the missing heritability, then it may be manifest in the polymorphism maintained by various processes of selection, such as frequency-dependent selection, balancing or disruptive (antagonistic) selection (especially in relation to sex), heterozygote advantage, and environment-dependent effects; these processes collectively maintain high levels of polymorphism – as in *Drosophila*.¹⁵ Moore and Williams¹⁶ have considered the issues in some depth and it is encouraging that evidence for such interactions has emerged from reworking data from the Wellcome Trust Case–Control Consortium data sets.¹⁷ Moore has commented that a biological appreciation of the effects of selection may be more readily apparent when phenotypes are grouped according to the pathway of development, metabolism, or function that is involved.¹⁸

Within the field of gene×environment interactions, we must not omit to consider epigenetic processes that may contribute substantially to a range of complex disorders, often through processes that permit the body to adapt over decades, or even across generations, to the environmental circumstances confronting the individual.

The question of natural selection operating in modern humans so as to maintain a high degree of genetic polymorphism in the population is highly contentious yet important if we are to understand complex disorders. If little selection is operating, then one potential explanation for the ‘missing heritability’ will be removed. Some commentators have searched for evidence of recent¹⁹ or more ancient²⁰ selection and found few traces; more recent studies, however, have had more success in that the levels of polymorphism maintained by selection may be substantial.^{21–23} Polymorphism maintained by selection may therefore contribute substantially to the aetiology of complex disorders.

(iii) Genetic variation in regulatory sequences modifying levels or patterns of gene expression, in regulatory gene sequences, and in regulatory genes (including almost certainly microRNA genes).

A number of other possible factors may account for some of the difficulty in identifying the genetic basis of the ‘missing heritability’, in particular, the narrow range of populations studied in any depth, the scope for (so far undetected) rare variants (including structural variants) to contribute to disease causation, and environmental factors.

Finally, we come to the misinterpretation of the high estimates of heritability found in numerous studies of complex diseases in humans. These have been taken as demonstrating that such conditions must have been transmitted through the family from parent to child. When we recognise that this need not be so, the problem of the missing heritability may appear less daunting. The word 'heritability' has a tendency to seduce us into thinking that any trait with an abundance of this particular quality must necessarily be inherited; the idea that it could arise *de novo* seems counter-intuitive, although the situation is obvious once pointed out.

The diseases for which GWAS have had some success in identifying numerous contributory genetic loci (such as diabetes mellitus type 1 and Crohn's disease) are precisely those conditions in which one would expect natural selection to have been operating over the millennia, trading one effect on survival or reproductive success against another. As a result, substantial amounts of genetic polymorphism will have accumulated through the action of either antagonistic or balancing selection, ensuring that extensive polymorphism will have been present during the time required for GWAS to be able to contribute productively to today's research programmes. By contrast, those disorders in which most genetic variation has been predominantly harmful – either to physical or cognitive development, or to reproductive fitness – will tend to have a much smaller pool of ancient genetic variation capable of contributing to current disease. It follows that most of the genetic variation, which contributes to such disorders in contemporary humans, will be comparatively young – either *de novo* mutations or else recent mutations originating over the last few generations, which will consequently be invisible to GWAS.

The above considerations are especially pertinent in the context of developmental disorders and psychiatric disease. It has long been clear that many disorders of physical and cognitive development and serious psychiatric conditions arise *de novo* in the absence of any previous family history. The finding that *de novo* CNVs are involved in the aetiology of psychiatric disease (eg, Xu *et al*²⁴), however, strongly suggests that other new mutational events – probably including different types of subtle microlesion – may account for many additional cases. The proportion of observable CNVs that arise *de novo* is as yet unclear, but is significant enough to suggest that selection against these *de novo* variants is likely to be substantial.

So, where is the 'missing heritability'? Most estimates of the heritability of serious psychiatric conditions have been derived from twin and adoption studies – ie, from comparisons between identical and fraternal twins or between identical twins brought up together or apart. In those cases caused by newly arisen mutations of large effect, the heritability of these disorders would appear to be fairly substantial because both of the identical twins will be affected (as the penetrance of the mutations will be high), whereas both of the fraternal twins will not usually be affected as only one of them will carry the mutation. (NB: It should also be appreciated that recurrence of a disorder within a sibship, or the occurrence of more severe disease in a child than in the less seriously affected parent, may both be consequences of parental mosaicism.) These assessments of heritability have in the past been interpreted as if the polygenic model of quantitative traits was equally applicable to these different disease contexts. However, it is now clear that we must allow for multiple, newly arisen Mendelian mutations (or contiguous gene CNVs). Although the measurements on which the original assessments of heritability were based may well have been accurate, and though the results did not conflict with the polygenic threshold model of complex disorders, the estimates of heritability obtained could nevertheless still have been spurious because the model of inheritance assumed to be applicable was

incorrect. The heritability assessments did not allow for the frequent occurrence of new mutations.

The 'missing' heritability may therefore not be so much elusive as absent. To return to our earlier allusion to the philosophy of science, this is very much an instance of the crucial distinction between the induction- and verification-oriented Baconian on the one hand, and the falsification-oriented Popperian modes of thought on the other. Finding a substantial heritability for a disorder was taken, in the Baconian tradition, as confirmation (verification) of the explanatory model. The (Popperian) search for falsification is a much more powerful approach, however, as the heritability studies were not set up so as to falsify an alternative explanatory model. The measures of heritability used have therefore completely failed to distinguish between these two major possibilities.

In conclusion, and in sharp contrast to the views of Maniolo *et al*,¹³ we believe that *de novo* CNVs, occurring in some cases at remarkably high frequencies,² could make a substantial contribution to disease heritability. Where they occur *de novo* in association with neurodevelopmental and psychiatric disorders, CNVs will almost certainly contribute to measurements of heritability in both twin and adoption studies, and perhaps even in some family-based studies incorporating parents and children. It is the biological interpretation of these estimates that has so often been flawed, not the measurements of heritability themselves. We believe that it is quite unreasonable to expect that GWAS can make a meaningful contribution to understanding those psychiatric and developmental disorders in which the major genetic contributions have arisen *de novo* as mutations of large effect.

Finally, we should remember that many smaller CNVs have been excluded from consideration in studies of chromosome structural variation and studies of disease–gene association, despite the fact that they seem to occur at higher frequencies than the CNVs of > 1 kb. These smaller CNVs, which may or may not be located within the bounds of known, annotated genes (and which could encompass some of the myriad non-coding RNA genes), will have provided many opportunities for selection to ensure that disease-relevant variation has not been maintained in a population over time. The widespread existence of such under-researched genetic variation relevant to disease can only strengthen the case for which we are arguing here.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Angus J Clarke and David N Cooper
 Department of Medical Genetics, Haematology & Pathology,
 Institute of Medical Genetics, Cardiff University School of Medicine,
 Heath Park, Cardiff CF14 4XN, Wales, UK
 E-mail: ClarkeAJ@Cardiff.ac.uk

- 1 Maher B: The case of the missing heritability. *Nature* 2008; **456**: 18–21.
- 2 Conrad DF, Pinto D, Redon R *et al*: Origin and functional impact of copy number variation in the human genome. *Nature* 2009; doi:10.1038/nature08516.
- 3 Bourgain C, Génin E, Cox N, Clerget-Darpoux F: Are genome-wide association studies all that we need to dissect the genetic component of complex human diseases? *Eur J Hum Genet* 2007; **15**: 260–263.
- 4 Estivill X, Armengol L: Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet* 2007; **3**: 1787–1799.
- 5 Seng KC, Seng CK: The success of the genome-wide association approach: a brief story of a long struggle. *Eur J Hum Genet* 2008; **16**: 554–564.
- 6 Bodmer W, Bonilla C: Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 2008; **40**: 695–701.

- 7 van Ommen G-J: GWAS here, last year. *Eur J Hum Genet* 2008; **16**: 1–2.
- 8 Nothnagel M, Lu TT, Krawczak M: Hypotheses in genome-wide scans. *Eur J Hum Genet* 2008; **16**: 1174–1175.
- 9 Stefansson H, Rujescu D, Cichon S *et al*: Large recurrent microdeletions associated with schizophrenia. *Nature* 2008; **455**: 232–236.
- 10 International Schizophrenia Consortium: Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008; **455**: 237–241.
- 11 Weiss LA, Shen Y, Korn JM *et al*: Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 2008; **358**: 667–675.
- 12 McCarroll SA: Extending genome-wide association studies to copy-number variation. *Hum Mol Genet* 2008; **17**: R135–R142.
- 13 Manolio TA, Collins FS, Cox NJ *et al*: Finding the missing heritability of complex diseases. *Nature* 2009; **461**: 747–753.
- 14 Weiss KM: Tilting at quixotic trait loci (QTL): an evolutionary perspective on genetic causation. *Genetics* 2008; **179**: 1741–1756.
- 15 McKay TFC: The genetic architecture of quantitative traits: lessons from *Drosophila*. *Curr Opin Genet Devel* 2004; **14**: 253–257.
- 16 Moore JH, Williams SM: Epistasis and its implications for personal genetics. *Am J Hum Genet* 2009; **85**: 309–320.
- 17 Emily M, Mailund T, Hein J, Schauer L, Schierup MH: Using biological networks to search for interacting loci in genome-wide association studies. *Eur J Hum Genet* 2009; **17**: 1231–1240.
- 18 Moore JH: From genotypes to genomotypes: putting the genome back in genome-wide association studies. *Eur J Hum Genet* 2009; **17**: 1205–1206.
- 19 Bubb KL, Bovee D, Buckley D *et al*: Scan of human genome reveals no new loci under ancient balancing selection. *Genetics* 2006; **173**: 2165–2177.
- 20 Asthana S, Schmidt S, Sunyaev S: A limited role for balancing selection. *Trends Genet* 2005; **21**: 30–32.
- 21 Kelley JL, Swanson WJ: Positive selection in the human genome: from genome scans to biological significance. *Annu Rev Genomics Hum Genet* 2008; **9**: 143–160.
- 22 Baysal BE, Lawrence EC, Ferrell RE: Sequence variation in human succinate dehydrogenase genes: evidence for long-term balancing selection on *SDHA*. *BMC Biology* 2007; **5**: 12.
- 23 Hillmer AM, Freudenberg J, Myles S *et al*: Recent positive selection of a human androgen receptor/ectodysplasin A2 receptor haplotype and its relationship to male pattern baldness. *Hum Genet* 2009; **126**: 255–264.
- 24 Xu B, Roos JL, Levy S *et al*: Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nat Genet* 2008; **40**: 880–885.