### **GUEST EDITORIAL**

www.nature.com/mp

# Genome-wide association studies in psychiatry: lessons from early studies of non-psychiatric and psychiatric phenotypes

*Molecular Psychiatry* (2008) **13**, 649–653; doi:10.1038/mp.2008.45; published online 27 May 2008

Recent advances in high-throughput genotyping now permit genome-wide association studies (GWASs) in which hundreds of thousands of DNA polymorphisms spread across the genome can be assayed in a large set of individuals rapidly and for a realistic cost.<sup>1</sup> The GWAS approach has now been successfully applied to many non-psychiatric diseases. In this editorial, we consider the key lessons from those studies for GWASs of psychiatric disorders, using type 2 diabetes (T2D) as a particularly instructive example.

#### What is a GWAS?

The defining feature of a GWAS is that a suitably large number of genetic polymorphisms is examined to provide an acceptable level of association information across the whole genome.<sup>1,2</sup> The technical advance that has made this possible is the availability of genotyping chips that can characterize DNA-sequence variation at hundreds of thousands of single-nucleotide polymorphisms (SNPs). A perfect tool would provide complete information at every variable point in the genome. Reality falls short of the ideal, with current designs typically capturing a high proportion of the information for around 65-80% of variant sites where the minor allele frequency is above 5%, assuming most of the SNPs on the array pass appropriate quality control measures.<sup>3</sup> Some regions of the genome are covered well, others less well, and low-frequency alleles (minor allele frequency <1%) are generally not interrogated with current study designs. Thus, the power of current GWAS is constrained not only by the size of the sample, but also by the technical properties of the genotyping chip used (in terms of coverage of genomic location and spectrum of DNA variants that can be detected). It is extremely important to recognize these shortcomings, and in particular that GWASs are not well powered to detect rare variation that influences disease susceptibility (even rare variants of large effect)-such variants require approaches based on sequencing. Also the study design is not optimal for detecting common alleles at risk loci where there are multiple risk variants on multiple haplotypes within the same genomic region. It follows, therefore, that in general, GWASs are unable to provide definitive data for 'excluding' a gene from involvement in susceptibility to illness, and there continues to be an important role for focused and detailed molecular genetic analysis of genes that are the subject of specific biological or positional hypotheses.

Genome-wide association studies can be used in either of the two main genetic association study designs: case-control based on unrelated subjects and family-based association designs of many sorts. In practice, most GWASs are of the unrelated casecontrol design. One reason is that adequately powered GWASs as applied to complex diseases require very large sample sizes, and unrelated case-control samples are usually much easier and cheaper to collect than family-based samples. Another is that case-control designs can exploit a single large common set of controls, the allele frequencies in which can be contrasted with many different disorders.<sup>4</sup> This is clearly more economic than family designs in which the controls are unique to that study.

#### Lessons from GWASs of non-psychiatric diseases

#### Lesson 1: GWASs work

Proof of principle for GWASs in human disease was provided by the identification of the gene encoding complement factor H as a risk locus for age-related macular degeneration.<sup>5</sup> However, it should be noted that this study was atypical in that the risk variant identified had a relatively large effect size that was detectable in a mere 96 cases and 50 controls typed for only ~116 000 SNPs. Subsequently, GWASs have resulted in the identification of alleles that have been confidently associated with common diseases including coronary artery disease, atrial fibrillation, asthma, Crohn's disease, rheumatoid arthritis, type 1 (T1D) and T2D, obesity, prostate cancer, breast cancer and coeliac disease.<sup>4.6.7</sup>

## Lesson 2: effect sizes are usually small, so big samples are needed

Acknowledging the possibility of success in a small sample, the overwhelming message from the many GWASs of complex diseases so far is the importance of large samples powered to detect small effect sizes. Self evidently, the true effect size for any given risk allele cannot be known in advance of its identification, but theoretical considerations lead to the expectation of a spectrum of effect size and risk allele frequency, with alleles of small effect being much Editorial

more frequent than alleles of large effect.<sup>8</sup> Consistent with the theoretical predictions, with few exceptions, the effect sizes that have been identified in studies of non-psychiatric diseases have been in the small range. For example in the Wellcome Trust Case Control Consortium (WTCCC) GWASs of seven common diseases,<sup>4</sup> per-allele odds ratios of identified loci were in the range 1.2–1.5, and even these are likely to be somewhat inflated, a phenomenon known as the so-called 'winners curse' (see below). To have reasonable power to detect such loci, it requires samples of the order of 2000 cases and 2000 controls or larger. If a sample of 1000 cases and 1000 controls for each disease had been used in WTCCC, it is estimated that only six rather than 16 signals would have been detected at the most-stringent significance threshold.<sup>4</sup>

The importance of large samples is further illustrated by the recent follow-up and meta-analysis of T2D that used four case–control samples, each with case numbers ranging from 1924 to 6529, resulting in a total of 14 586 cases and 17 968 controls.<sup>9</sup> Within the WTCCC study, three susceptibility loci for T2D were identified at the most stringent threshold of statistical significance whereas the meta-analysis allowed the identification of nine robustly associated SNPs  $(P=1.2 \times 10^{-7}-1 \times 10^{-48})$ . Small genetic effects were the rule with all but one having an odds ratio  $\leq 1.20$ .

It is important to remember that in most cases, larger samples are required for replicating a finding than can be predicted from the effect size estimated from the discovery sample. This 'winners curse' arises because most discoveries benefit from, indeed in small samples require, a favorable constellation of factors that amplify the effect size in the discovery sample above the true effect size in the population from which it is drawn.<sup>10</sup> Examples of factors that contribute to this include chance fluctuations in allele frequencies in cases and controls that maximize the distinction between the two groups or genotyping errors operating in a similar manner. Given that we require large samples for initial detection of an effect, we need to recognize that, on their own, even a series of failures to replicate findings in modestly sized replication samples do not constitute refutation of a finding, although small samples can contribute to assessing the involvement of a locus when incorporated into meta-analyses that summarize the overall state of evidence from all samples.

An issue that is often not appreciated is that the confidence attributable to a 'significance level' is influenced by sample size.<sup>4</sup> Ignoring the impact of errors or poor design (for example, population stratification), any set of results that cross a certain threshold of statistical significance can be expected to be a mixture of 'true' positives reflecting a genuine disease association and false positives reflecting chance. The rate of false positives due to chance is approximately constant for all (nontrivial) sample sizes and is simply the significance level used. However, the rate of true positives will increase with

sample size because power to detect true effects will increase. Thus, the proportion of true associations among significant findings is expected to be greater in larger samples than in small samples, and in general, we should be wary of apparently impressive findings in samples that have very limited power under plausible models of effect size. In contrast, we should place increasing confidence in highly significant findings in large samples that are well-powered to detect plausible effect sizes.

#### Lesson 3: rigorous quality control is paramount

The importance of quality control is not unique to GWASs. However, the enormous data sets (samples and SNPs) in GWASs provide a large number of opportunities for spurious 'associations' to emerge with high levels of statistical significance. One important source of spurious findings is systematic differences in allele calling between cases and controls,<sup>11</sup> a phenomenon that has particular impact on haplotype-based tests.<sup>12</sup> It is crucial, therefore, that the data are thoroughly cleaned to remove low-quality DNA samples, genotype calls and individual samples. It is sobering that in uncleaned (and partly cleaned) data, it was the experience within WTCCC that the best predictor of an SNP with poor QC was a highly significant difference in genotype distributions between cases and controls.

Lesson 4: GWASs may fail to detect susceptibility genes It should be clear from the foregoing considerations that the GWAS approach is poorly powered to detect any susceptibility gene that is not well covered by SNPs on the array used. Further, the approach is not designed to detect a susceptibility allele that is rare, even if that gene is well tagged to capture common variation. An example of this is provided by the WTCCC study of T1D where the insulin gene (*INS*) was not detected because of poor coverage of the gene.<sup>4</sup> This provides a clear illustration of the importance of not assuming that absence of an association signal within a GWAS is a strong evidence against, or worse can 'exclude', a gene from involvement in illness.

## *Lesson 5: it is important to look well beyond the top few 'hits'*

There is a natural tendency to focus attention on the few strongest findings in any particular study. However, this is not the best way to exploit GWASs, the strength of which is the provision of genome-wide data in a reasonably unbiased way. It can be shown theoretically (and this is borne out in practice) that for susceptibility loci of plausible effect sizes, even large studies cannot reliably place these loci at the very top of the list of hits for each study. The reason is simply that for such loci, even large samples do not have sufficient power to identify loci at very stringent levels of statistical significance, whereas they often are well-powered to identify the loci at more modest levels of significance.<sup>4</sup> In practical terms, this means that the true risk loci will, in suitable sample sizes, usually fall within the top few hundred or thousand hits, but which of these make it to the top of the list is largely the result of the factors that drive the 'winners curse'. These factors will differ between studies and so we cannot expect the most significantly associated SNPs to match across studies. We would, however, expect overrepresentation of the risk loci among the top few hundred or few thousand.

Clear examples of this are provided by studies of T2D. One of the previously known robust associations, a polymorphism within PPARG, was detected within WTCCC but the significance level was  $P = 1.3 \times 10^{-3}$ . The key point here is that at the level of a specific hypothesis concerning this gene, the WTCCC is positive because this significance level survives correction for the multiple SNPs tested within the gene. However, this result does not even approach significance if correcting for testing at a genome-wide level, the signal ranking only just within the top 1000 hits. Moreover, the polymorphism at PPARG did not even achieve (within the context of the genome) compelling evidence for significance in the large combined meta-analysis sample of 14586 cases and 17968 controls  $(P=1.7\times10^{-6})$ .<sup>9</sup> Thus, it can be expected that many-in fact, almost certainly most-of the true associations will be contained within the large set of hits of modest evidence rather than being concentrated in the top few highly significant hits. The traditional way of thinking has been to focus attention on the very top one or few hits because the evidence from the study itself is most persuasive for these hits. However, as soon as one moves to thinking of combining other data sources—be they other GWASs or non-genetic sources of data-it is much more appropriate to use approaches that consider a much broader range of hits.

#### Lesson 6: collaboration is important

Given the modest effect sizes and need for large samples, it should be intuitively obvious that very substantial benefits can be gained from collaborations that increase total sample sizes and test consistency and generalizability of findings. As mentioned above, in T2D, 'aggressive data sharing' (that is, very early and proactive) was key to rapid and efficient identification of several susceptibility loci that were not evident in any single study alone. This model has been used to great effect for several other diseases including T1D,<sup>13</sup> Ankylosing Spondylitis<sup>14</sup> and coronary heart disease.<sup>15</sup>

#### Lesson 7: phenotype/selection is important

It has long been acknowledged in theory, but widely ignored in practice, that sample ascertainment and, consequently, variation in phenotype (case or control) between studies can have a dramatic effect on the ability to detect a susceptibility locus. A striking empirical example of this fact is provided by the gene FTO, which was shown in the collaborative meta-analysis9 to be associated with risk of T2D  $(P = 1.3 \times 10^{-12})$ . However, association at FTO was not significant in the  $\sim 14\,000$  subjects comprising the 'DGI' sample. In fact the estimated odds ratio for the risk polymorphism in the diabetes genetics initiative (DGI) sample was close to unity (that is, no effect). In contrast, association was highly significant in the similarly sized UK sample ( $\vec{P} = 7 \times 10^{-14}$ ). Evaluation of study design led to the realization that the difference was caused by important phenotypic differences in design and analysis: in the DGI sample, analyses matched for measures of obesity whereas no such criterion was imposed on the WTCCC study. Subsequent work has shown that fat mass and obesity associated (FTO) influences risk of T2D through a primary effect on body mass.<sup>16</sup>

This shows that phenotype variation can be critical to the ability to identify susceptibility variants and that taking account of phenotype variation across samples has the potential to aid the understanding of the mode of action of a susceptibility locus.

#### Are psychiatric disorders different from non-psychiatric disorders?

Findings from genetic epidemiology, such as familial recurrence risks and estimates of heritability, show that many types of major psychiatric illness are among the most genetically influenced of human traits and diseases.<sup>17</sup> As for other disorders, it is likely that a range of mechanisms may influence genetic risk including common polymorphisms, rare mutations and structural rearrangements. There seems to be no strong reasons to expect that the genetic mechanisms underlying major psychiatric illness will be qualitatively different from those underlying non-psychiatric disorders. There are, perhaps, reasons to expect that the effect sizes of common susceptibility variants might be at the lower end of the range of effect sizes for complex diseases as a whole. The rationale for this suggestion is that major disturbances in behavior and social functioning, the hallmark of major psychiatric illness, would usually be expected to adversely influence an individual's ability to reproduce and pass genes on to future generations. It follows that, in the absence of balancing selection, variants conferring a high level of risk would rapidly become lost from the population and would not establish themselves as common polymorphisms.

However, the most obvious issue that marks out psychiatric genetics as being more challenging than genetic investigation of non-psychiatric disorders is the phenotype.<sup>18,19</sup> This is more difficult to define and measure than for most non-psychiatric disorders. Further, we have less knowledge of the causes and mechanisms of pathogenesis. Our current official classification systems, Diagnostic and Statistical Manual of Mental Disorders and International Classification of Diseases, are descriptive systems. They were developed to have acceptable reliability but with no expectation that the categories represented valid Editorial

entities. Many of the diagnostic categories have a degree of genetic validity from genetic epidemiology, but it is also clear that there is likely to be genetic overlap between categories<sup>19,20</sup> and much heterogeneity within them. This will have a number of consequences for gene-finding studies. First, the inherent heterogeneity means that diagnostic categories will be more complex genetically than for nonpsychiatric disorders and consequently the effect sizes of individual loci might be expected to be smaller than those for non-psychiatric disorders. The corollary of this is that we might expect genetic effects to be greater for more narrowly defined phenotypes. The problem is that for many disorders, there are no clear-cut ways of subdividing the phenotype *a priori*. This means that we have few alternatives to empirical approaches with the consequent burden of multiple testing. However, an approach that can help with this problem is the use of phenotype refinement and a subsequent iterative approach toward a more biologically valid clinical phenotype.<sup>21</sup> Second, the genetic overlap between disorders suggests that it might be fruitful to explore the relationship between specific genetic findings and specific symptom profiles and dimensions between as well as within diagnostic groupings. Third, the fact that diagnostic categories are not anchored to an underlying pathophysiology suggests that even quite subtle differences in ascertainment and diagnosis could alter the constellation of alleles conferring risk to the samples in question, with damaging consequences for consistency and replication. It follows that if psychiatric genetics is to harness fully the power of GWASs, we must pay close attention to how we define the phenotype and expect a high degree of 'co-morbidity' and heterogeneity.

Those working in the field of psychiatric research are usually aware, at least at a theoretical level, of the limitations imposed by psychiatric phenotypes. However, the reality is that the researchers often do not have the training or experience to deal with the phenotype issues, and so they get substantially overlooked with the tacit assumption that using Diagnostic and Statistical Manual of Mental Disorders categories provides as useful a classification as blood sugar for diabetes, blood pressure for hypertension or histopathology for Crohn's disease or breast cancer. We know, as discussed above, that even for an illness like T2D, clinical covariates like obesity can have a dramatic influence on the effect size of risk alleles. For psychiatric phenotypes, as presently defined, 'comorbidities' (that is, overlaps in clinical syndromes) are the rule and we must expect that effect sizes of risk alleles will vary greatly between and within samples according to the phenotypic characteristics of the individuals within the sample. Obvious examples of psychiatric scenarios that may be similar to the obesity-diabetes situation include presence or absence of prominent psychotic features in bipolar disorder or prominence of anxiety in recurrent depression. The phenotype issue, though not restricted to psychiatry, is likely to be the issue that most distinguishes psychiatry from the non-psychiatric diseases.<sup>22</sup> Consequently, it is the area that needs particular attention in psychiatry to ensure that maximum benefit and efficiency is gained from the ongoing major investments of time and money in GWASs.

In summary, the phenotype issues provide a particular challenge for psychiatric genetics. It seems quite possible that effect sizes for common susceptibility variants will be at the lower end of the range found for complex diseases but there is nothing to suggest that they are qualitatively different from other complex disorders.

## What are the principles we should apply to GWASs of psychiatric disorders?

There is no reason to suspect that the approaches that have been so successful for non-psychiatric diseases are not capable of being successful in psychiatry. However, successful application of the tools of molecular genetics in general, and GWASs in particular, for the benefit of psychiatric patients will require us to fully embrace the lessons from studies of non-psychiatric disorders while taking account of those issues that may be specific to psychiatric phenotypes.

Box 1 summarizes key principles that should be applied to conduct and interpret GWASs of psychiatric phenotypes.

Given that most of these data sets will be made available for analysis by independent researchers at an early stage, it is crucial that all those analyzing the data are aware of these issues and take account of them in interpreting and reporting analyses.

#### Experience to date in GWASs of bipolar disorder

In the WTCCC, the pattern of association signals in bipolar disorder when compared with most of the other disease phenotypes, showed fewer hits exceeding the highly significant benchmark,  $P < 5 \times 10^{-7}$  but more signals within the more modest range (P = $10^{-4}-5 \times 10^{-7}$ ) (Table 1). This is consistent with the idea that, at least as currently defined, most of the common genetic variation that influences susceptibility to bipolar disorder has modest effect sizes and

**Box 1** Key issues of importance in GWAS in general and GWAS of psychiatric disorders in particular

Pay rigorous attention to quality control			
Use appropriately large samples			
Collaborate and share data			
Consider distribution of signals not just top few hits			
Pay close attention to phenotype characteristics and			
selection of sample			
Do not over-interpret a 'negative' finding at a locus			

Abbreviation: GWAS, genome-wide association study.

**Table 1** Distribution of association signals in WTCCC

Phenotype	$10^{-4} < P < 10^{-5}$	$5 \times 10^{-7} < P < 10^{-6}$	$P < 5 \times 10^{-7}$
BD	76	14	1
CAD	57	9	1
CD	85	9	9
HT	56	6	0
RA	49	9	2
T1D	55	8	5
T2D	47	10	3

Abbreviations: BD, bipolar disorder; CAD, coronary artery disease; CD, Crohn's disease; HT, hypertension; RA, rheumatoid arthritis; T1D, type 1 diabetes; T2D, type 2 diabetes.

Summary of the numbers of independent association signals where at least one SNP achieved the significance threshold shown within the Wellcome Trust Case Control Consortium (WTCCC) study.<sup>4</sup>

that there are very few (perhaps none) common susceptibility alleles of large effect such as ApoE in Alzheimer's disease or HLA in T1D and other autoimmune disorders. The most significant finding was at rs420259 in a gene rich region at 16p12 (genotypic  $P = 6.3 \times 10^{-8}$ ). Consistent with the discussions above and with experience in non-psychiatric disorders, this locus was not identified by the other individually typed large-scale GWASs of bipolar disorder for which data are available (STEP-UCL sample),<sup>23</sup> nor the published GWASs using a pooling approach in samples from US and Germany.<sup>24</sup> However, if rather than focusing on the top one (or few) hits one considers the distributions of association signals, there is a very clear, highly significant overlap in the association shown within the WTCCC and US/ Bonn data sets  $(P = 7 \times 10^{-5})$ .<sup>25</sup> Similarly, meta-analysis of the WTCCC and STEP/UCL data sets demonstrated a strong signal with the same risk allele within the gene CACNA1C (meta-analysis  $P = 6.9 \times 10^{-7}$ ).<sup>23</sup> There is a substantial amount of work required to take such observations forward but the pattern of findings to date is that expected theoretically for a complex disease and is similar to that observed in studies of similar-sized samples for non-psychiatric phenotypes.

#### Conclusion: the first step on a long journey

We are at the beginning of GWASs in psychiatry. The approach has a great deal to offer although it is important to remember that GWAS forms one (albeit a major) component of the genetic approaches that can help us better understand psychiatric illness. It will be necessary also to pursue approaches designed to detect rare variants and structural variations that contribute to illness. There is also a continuing need to undertake detailed and extensive study of specific genes to test specific biological hypotheses.

The initial experience with GWAS in bipolar disorder is consistent with the experiences in nonpsychiatric disorders and suggests we can be optimistic that careful application of the GWAS approach across large, phenotypically well-characterized samples, including those of the GAIN collaborations,26 will make an important contribution to delineating the etiology and pathogenesis of the disorders that can devastate the lives of our patients.

#### Acknowledgments

We are grateful to all our colleagues and collaborators within the Wellcome Trust Case Control Consortium and are indebted to the many individuals who have participated in, and helped with, our research.

#### **Declaration of interest**

None.

N Craddock, MC O'Donovan and MJ Owen Department of Psychological Medicine, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff, UK *E-mail: craddockn@cardiff.ac.uk* 

#### References

- 1 Hirschhorn JN et al. Nat Rev Genet 2005; 6: 95-108.
- 2 Barrett JC, Cardon LR. Nat Genet 2006; 38: 659-662.
- 3 International HapMap Consortium. Nature 2007; 449: 851-861.
- 4 Wellcome Trust Case Control Consortium. Nature 2007; 447: 661-678
- 5 Klein RJ et al. Science 2005; 308: 385-389.
- 6 Todd JA. Nat Genet 2006; 38: 731-733.
- 7 Petretto E et al. Nat Genet 2007; 39: 1299-1301.
- 8 Wang WY et al. Nat Rev Genet 2005; 6: 109-118.
- 9 Zeggini E et al. Science 2007; 316: 1336-1341. E-pub April 26 2007. Erratum in: Science. 2007;317(5841):1035-6.
- 10 Zollner S, Pritchard JK. Am J Hum Genet 2007; 80: 605-615.
- 11 Clayton DG et al. Nat Genet 2005; 37: 1243-1246.
- 12 Moskvina V et al. Hum Hered 2006; 61: 55-64.
- 13 Todd JA et al. Nat Genet 2007; 39: 857-864.
- 14 Wellcome Trust Case Control Consortium. Nat Genet 2007; 39: 1329-1337.
- 15 Samani NJ et al. N Engl J Med 2007; 357: 443-453.
- 16 Frayling TM et al. Science 2007; 316: 889-894.
- 17 Psychiatric genetics and genomics. McGuffin P, Owen MJ (eds). Gottesman II. Oxford University Press: Oxford, 2002, 472pp.
- 18 Angst J. World Psychiatry 2007; 6: 94-95.
- 19 Craddock N, Owen MJ. World Psychiatry 2007; 6: 84-91.
- 20 Kendler KS et al. Arch Gen Psychiatry 1992; 49: 716-722.
- 21 Craddock N et al. Mol Psychiatry 2006; 11: 446-458.
- 22 Craddock N et al. Br J Psychiatry 2007; 190: 200-203. Erratum in: Br J Psychiatry. 2007;190:365.
- 23 Sklar P et al. Mol Psychiatry 2008; e-pub ahead of print 4 March 2008.
- 24 Baum AE et al. Mol Psychiatry 2008; 13: 197-207.
- 25 Baum A et al. Mol Psychiatry (in press).
- 26 GAIN Collaborative Research Group. et al. Nat Genet 2007; 39: 1045-1051.

