

## Editorial

Dissecting the phenotype  
in genome-wide association  
studies of psychiatric illnessCross-Disorder Phenotype Group of the Psychiatric  
GWAS Consortium\***Summary**

Over the past 2 years genome-wide association studies have made major contributions to understanding the genetic architecture of many common human diseases. This editorial outlines the development of such studies in psychiatry and highlights the opportunities for advancing understanding of

the biological underpinnings and nosological structure of psychiatric disorders.

**Declaration of interest**

None.

The International Cross-Disorder Phenotype Group is undertaking analyses in over 80 000 samples to inform understanding of the biological underpinnings and nosology of major psychiatric illness.

Genome-wide association studies involve genotyping hundreds of thousands of common DNA variants (single nucleotide polymorphisms, SNPs) spread throughout the genome in large numbers of individuals with illness and a similar number of comparison individuals with a low prevalence of illness ('controls').<sup>1</sup> Over the past 2 years such studies have made major contributions to advancing our understanding of many common diseases, including diabetes, heart disease, inflammatory bowel disease, various cancers and rheumatoid arthritis.<sup>2</sup> In Crohn's disease, 30 different genes have already been robustly shown to influence risk and this has pointed to novel biological pathways involved in illness pathogenesis.<sup>3</sup>

This editorial outlines the development of genome-wide association approaches in psychiatry and highlights opportunities for advancing knowledge of the biological underpinnings and nosological structure of psychiatric disorders.

**Genome-wide association studies in psychiatric illness**

Although relatively few studies have so far been published, large-scale collaborative studies have started to deliver genome-wide significant genetic associations for bipolar disorder and schizophrenia. Studies of approximately 10 000 individuals have shown strong evidence for association with susceptibility to bipolar disorder at variants within two genes involved in ion channel function: *ANK3* (encoding the protein ankyrin-G) and *CACNA1C* (encoding the alpha-1<sub>C</sub> subunit of the L-type voltage-gated calcium channel).<sup>4</sup> A similar study in close to 20 000 individuals has shown strong evidence for association with susceptibility to schizophrenia at a variant within *ZNF804A* (encoding a zinc finger transcription factor).<sup>5</sup> Recent independent data provide further support for the involvement of *ANK3* in bipolar disorder and suggest the existence of at least two distinct susceptibility variants at this locus.<sup>6</sup> Although further study and replication of the findings is important, these initial results suggest that the study of even larger samples will identify additional reliable associations

and thereby extend knowledge of the proteins and biological pathways involved in illness.

**Psychiatric GWAS Consortium (PGC)**

One of the key lessons from the studies conducted to date is the necessity for very large samples (ie. thousands of cases and controls).<sup>7</sup> Investigators and funders have recognised that scientists need to cooperate and share data rather than compete and restrict data access and consequently collaborative consortia have been forming. In psychiatry, the largest among these includes a large proportion of the world's research groups working with psychiatric genome-wide association study data: the Psychiatric Genomewide Association Study Consortium (PGC: <http://pgc.unc.edu>). The disorders that are currently represented in the PGC are attention-deficit hyperactivity disorder (ADHD), autism, bipolar disorder, major depressive disorder, and schizophrenia. It is expected that during 2009, the PGC will include over 80 000 individuals each with about 500 000 SNP genotypes. The aims of the PGC are to coordinate and facilitate the necessary large-scale collaborative analyses using both (a) traditional disorder categories, and (b) non-traditional analyses that cut across diagnostic categories. This latter aim is the focus of this editorial.

**Phenotype importance**

The importance of phenotype definition and selection on genome-wide association findings is demonstrated strikingly by work on type 2 diabetes where the gene *FTO* was robustly associated with illness in a collaborative meta-analysis.<sup>8</sup> However, association at *FTO* was not present at all in one of the three samples in the meta-analysis although it was highly significant in one of the other samples of similar size. The difference was caused by phenotypic heterogeneity: in the sample showing no association, cases were not included if the individuals were obese. No such exclusion criterion was present in the sample with the strong effect. Subsequent work showed that *FTO* influences diabetes risk through an effect on body mass.<sup>9</sup> This demonstrates that phenotype variation can be critical to the ability to identify susceptibility variants. Furthermore, taking account of phenotype variation across samples can provide critical information about the mode of action of a susceptibility locus.

Psychiatric scenarios that might produce results similar to the obesity-diabetes story include presence or absence of prominent psychotic features in bipolar disorder or prominence of anxiety in recurrent depression.

\*The following international researchers are members of the Cross-Disorder Phenotype Group: Nick Craddock (UK), Kenneth Kendler (USA), Michael Neale (USA), John Nurnberger (USA), Shaun Purcell (USA), Marcella Rietschel (Germany), Roy Perlis (USA), Susan L. Santangelo (USA), Thomas Schulze (USA), Jordan W. Smoller (USA) and Anita Thapar (UK).

## Genetic dissection of psychiatric phenotypes

What about psychiatric phenotypes? Psychiatric diagnoses can be considered 'the weak component of modern research',<sup>10</sup> defined solely by descriptive, usually behavioural, criteria. Although these phenotype definitions are highly heritable, and hence are valid and sensible starting points for genetic research, it is generally agreed that the most useful biological categories and/or dimensional definitions and measures are still unknown. The strikingly high level of co-occurrence of different diagnoses within the same individual (comorbidity) almost certainly reflects a substantial overlap in the underlying biology of currently defined syndromes. For example, the five psychiatric phenotypes represented in the PGC are unlikely to identify completely distinct disease entities and there may be overlaps in genetic susceptibility across the disorders. Justification for this assertion includes: (a) the existence of clinical symptom/item overlap across several of the phenotypes; (b) the non-independence of multiple diagnoses within the same individuals;<sup>11</sup> and (c) the observation that the same structural genetic variants have been described in association with differing phenotypes. For example, deletion of chromosome 22q11 has been associated with childhood autism and ADHD as well as adult mood disorders and psychosis.<sup>12</sup>

Molecular genetics will not provide a simple, gene-based classification of psychiatric illness (as it will not for other common familial illnesses).<sup>13</sup> The notion that there is a gene for one or more psychiatric disorders is inappropriate and unhelpful. Rather, there is a complex relationship between genotype and phenotype that involves multiple genes and environmental factors, together with stochastic variation. Nonetheless, molecular genetic findings can be expected to help delineate the relationship between specific biological pathways/systems and broad patterns, or domains, of psychopathology.<sup>14</sup> A precedent for such insights from genetic studies is already emerging from genome-wide association studies in other areas of medicine that have revealed unforeseen biological relationships among different autoimmune diseases.<sup>15</sup> We anticipate that genetic findings will not map cleanly onto current diagnostic categories and that genetic associations may point to more useful and valid nosological entities. To address this, the Cross-Disorder Phenotype Group was established within the PGC to coordinate, develop and lead different types of analyses both across and within the different phenotype sample sets.

## Types of analyses that may be relevant

We here briefly consider a range of analytic approaches that may be relevant to understanding the relationship between genotype and phenotype for psychiatric traits<sup>16</sup> (see Appendix). These include approaches designed both to discover new pathologically relevant genetic variants and also to characterise the phenotypic spectrum associated with robustly associated variants.

First, we can explore whether individual genetic variants increase risk across multiple diagnostic categories. For example, genes may exist that alter risk for both schizophrenia and autism, or for schizophrenia and bipolar disorder, or for bipolar illness and recurrent depression. Second, we can attempt to identify risk genes for psychosis, depressed mood or some other domain of psychopathology regardless of the syndrome in which they occur. Third, we can look for disease-modifying effects. For example, genes may exist which do not influence risk for the diagnostic category of schizophrenia but, when an individual has this diagnosis, alters the probability that they have auditory hallucinations or early onset.<sup>17</sup> Fourth, instead of starting with phenotypes and then looking at genotypes, we could reverse the order.<sup>16,18</sup> We might start with a single gene or genotype of interest and study

its phenotypic profile. Fifth, we could apply one of a range of advanced statistical tools to define novel diagnostic entities (whether they are categories or dimensions) that would 'make more sense' from a genetic perspective. Sixth, instead of focusing on single genetic variants, we could consider a large set of polymorphisms (perhaps tens of thousands) and use aggregate measures of their overall contribution to phenotypic susceptibility to seek to define 'signatures' of genetic variants, the patterns of which could be compared across phenotypes. This approach, which will be particularly useful if psychiatric phenotypes are highly polygenic (i.e. many, many risk genes, each of small effect on risk), has recently been used to demonstrate a substantial overlap in polygenic contribution to schizophrenia and bipolar disorder.<sup>19</sup>

Such large-scale analytic approaches have not been undertaken previously so an important task for the Cross-Disorder Phenotype Group will be developing and validating the methodology and reporting standards for these analyses. The general framework outlined will accommodate analysis of additional psychiatric phenotypes when they become available to the PGC.

## The challenges ahead

There are, of course, substantial challenges to be overcome in undertaking the types of analyses outlined earlier (see Appendix). An obvious, but crucial, logistical issue is ensuring the comparability of the data used across different component sample collections. Where possible, analyses should be robust to the inevitable variations in clinical measurement. Important statistical considerations include making efficient use of samples with incomplete data and taking account of multiple testing. This requires both a sensible analytic design to minimise the number of tests, together with an appropriate control for false-positive findings that might emerge simply from conducting many tests. Inevitably, any exploratory analyses will require independent replication of positive findings.

## Conclusion

The ongoing major investments of time and money in genome-wide association studies for psychiatric disorders has the potential to contribute to the identification of pathways involved in illness and help psychiatry move towards approaches to diagnosis and treatment that are grounded in a better understanding of pathogenesis. This would be of great benefit to patients.

**Nick Craddock**, PhD, FRCPsych, Department of Psychological Medicine, School of Medicine, Cardiff University, UK; **Kenneth Kendler**, MD, **Michael Neale**, PhD, Virginia Institute of Psychiatric and Behavioral Genetics, and Departments of Psychiatry, and Human and Molecular Genetics, Medical College of Virginia/Virginia Commonwealth University, Richmond, Virginia, USA; **John Nurnberger**, MD, PhD, Department of Psychiatry, Indiana University School of Medicine, Indianapolis, Indiana, USA; **Shaun Purcell**, PhD, Center for Human Genetic Research, Massachusetts General Hospital, Boston, and Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts, USA; **Marcella Rietschel**, MD, Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Mannheim, Germany; **Roy Perlis**, MD, Bipolar Clinical and Research Program, and Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA; **Susan L. Santangelo**, ScD, Harvard Medical School, Department of Psychiatry and Harvard School of Public Health, Department of Epidemiology, and Massachusetts General Hospital, Center for Human Genetic Research, Boston, Massachusetts, USA; **Thomas Schulze**, MD, Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Mannheim, Germany, and Unit on the Genetic Basis of Mood and Anxiety Disorders, NIMH, NIH, DHHS, Bethesda, Maryland, USA; **Jordan W. Smoller**, MD, PhD, Department of Psychiatry, Harvard Medical School, and Department of Epidemiology, Harvard School of Public Health, and Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA; **Anita Thapar**, PhD, FRCPsych, Department of Psychological Medicine, School of Medicine, Cardiff University, UK

**Correspondence:** Professor Nick Craddock, Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Health Park, Cardiff CF14 4XN, UK. Email: craddockn@cardiff.ac.uk

First received 22 Dec 2008, final revision 23 Jan 2009, accepted 6 Feb 2009

## Acknowledgements

The PGC Cross-Disorder Phenotype Group is grateful to all members of the PGC and to all those who have participated in, helped with, or provided funding support for the research.

## Appendix

### Types of analysis to delineate the relationship between genotype and phenotype

Here, phenotype refers to the measurable clinical characteristics of individuals, which may be considered at several levels (e.g. disorder, syndrome, factors or domains of psychopathology, or individual symptom items) and genotype refers to the measured genetic variation, which may also be considered at several levels (e.g. individual allele, individual polymorphism (SNP), gene, gene family, biological pathway, or other large set of polymorphisms (including polygenic 'signature')). It is possible to: (a) start with phenotype(s) and seek associated genotype(s) (traditional approach); (b) start with genotype(s) and seek correlated phenotype(s) ('reverse phenotyping' or 'phenotype refinement' approach); or (c) consider all phenotype and genotype data together and seek patterns of genotype–phenotype correlation (an approach that makes minimal prior assumptions about both nosology and pathogenesis).

#### Phenotype → genotype

- (a) Seek susceptibility across traditional diagnostic categories (uses combinations of disorders v. controls).
- (b) Seek susceptibility to specific domains of psychopathology (uses cases with specific clinical features v. controls).
- (c) Seek modifier genes for specific clinical features (uses cases with specific clinical features v. cases without specific clinical feature).
- (d) Look for patterns (or signatures) of large numbers of associated SNPs that can then be compared across samples or diagnoses.

#### Genotype → phenotype

- (a) Identify the phenotypic spectrum associated with a specific genotype of interest.

#### Genotype ↔ phenotype

Look for patterns of correlation in data with minimal prior assumptions (i.e. seek novel, genetically valid diagnostic entities).

## References

- 1 Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science* 2008; **322**: 881–8.
- 2 Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 2008; **118**: 1590–605.
- 3 Mathew CG. New links to the pathogenesis of Crohn disease provided by genome-wide association scans. *Nat Rev Genet* 2008; **9**: 9–14.
- 4 Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet* 2008; **40**: 1056–8.
- 5 O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskva V, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet* 2008; **40**: 1053–5.
- 6 Schulze TG, Detera-Wadleigh SD, Akula N, Gupta A, Kassem L, Steele J, et al. Two variants in Ankyrin 3 (ANK3) are independent genetic risk factors for bipolar disorder. *Mol Psychiatry* 2008; Dec 16 (Epub ahead of print).
- 7 The Psychiatric GWAS Consortium Steering Committee. A framework for interpreting genome-wide association studies of psychiatric disorders. *Mol Psychiatry* 2009; **14**: 10–7.
- 8 Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007; **316**: 1336–41.
- 9 Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007; **316**: 889–94.
- 10 Angst J. Psychiatric diagnoses: the weak component of modern research. *World Psychiatry* 2007; **6**: 94–5.
- 11 Rzhetsky A, Wajngurt D, Park N, Zheng T. Probing genetic overlap among complex human phenotypes. *Proc Natl Acad Sci USA* 2007; **104**: 11694–9.
- 12 Kates WR, Antshel KM, Fremont WP, Shprintzen RJ, Strunge LA, Burnette CP, et al. Comparing phenotypes in patients with idiopathic autism to patients with velocardiofacial syndrome (22q11 DS) with and without autism. *Am J Med Genet A* 2007; **143A**(22): 2642–50.
- 13 Kendler KS. Reflections on the relationship between psychiatric genetics and psychiatric nosology. *Am J Psychiatry* 2006; **163**: 1138–46.
- 14 Craddock N, Owen MJ. Rethinking psychosis: the disadvantages of a dichotomous classification now outweigh the advantages. *World Psychiatry* 2007; **6**: 84–91.
- 15 Lettre G, Rioux JD. Autoimmune diseases: insights from genome-wide association studies. *Hum Mol Genet* 2008; **17**: R116–21.
- 16 Schulze TG, McMahon FJ. Defining the phenotype in human genetic studies: forward genetics and reverse phenotyping. *Hum Hered* 2004; **58**: 131–8.
- 17 Fanous AH, Kendler KS. Genetic heterogeneity, modifier genes, and quantitative phenotypes in psychiatric illness: searching for a framework. *Mol Psychiatry* 2005; **10**: 6–13.
- 18 Craddock N, Jones L, Jones IR, Kirov G, Green EK, Grozeva D, et al. Strong genetic evidence for a selective influence of GABAA receptors on a component of the bipolar disorder phenotype. *Mol Psychiatry* 2008; Epub ahead of print: doi:10.1038/mp.2008.66.
- 19 International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia that overlaps with bipolar disorder. *Nature* in press.