Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis

Cross-Disorder Group of the Psychiatric Genomics Consortium*

Summary

Background Findings from family and twin studies suggest that genetic contributions to psychiatric disorders do not in all cases map to present diagnostic categories. We aimed to identify specific variants underlying genetic effects shared between the five disorders in the Psychiatric Genomics Consortium: autism spectrum disorder, attention deficit-hyperactivity disorder, bipolar disorder, major depressive disorder, and schizophrenia.

Methods We analysed genome-wide single-nucleotide polymorphism (SNP) data for the five disorders in 33 332 cases and 27 888 controls of European ancestry. To characterise allelic effects on each disorder, we applied a multinomial logistic regression procedure with model selection to identify the best-fitting model of relations between genotype and phenotype. We examined cross-disorder effects of genome-wide significant loci previously identified for bipolar disorder and schizophrenia, and used polygenic risk-score analysis to examine such effects from a broader set of common variants. We undertook pathway analyses to establish the biological associations underlying genetic overlap for the five disorders. We used enrichment analysis of expression quantitative trait loci (eQTL) data to assess whether SNPs with cross-disorder association were enriched for regulatory SNPs in post-mortem brain-tissue samples.

Findings SNPs at four loci surpassed the cutoff for genome-wide significance (p<5x10^-8) in the primary analysis: regions on chromosomes 3p21 and 10q24, and SNPs within two L-type voltage-gated calcium channel subunits, CACNA1C and CACNB2. Model selection analysis supported effects of these loci for several disorders. Loci previously associated with bipolar disorder or schizophrenia had variable diagnostic specificity. Polygenic risk scores showed cross-disorder associations, notably between adult-onset disorders. Pathway analysis supported a role for calcium channel signalling genes for all five disorders. Finally, SNPs with evidence of cross-disorder association were enriched for brain eQTL markers.

Interpretation Our findings show that specific SNPs are associated with a range of psychiatric disorders of childhood onset or adult onset. In particular, variation in calcium-channel activity genes seems to have pleotropic effects on psychopathology. These results provide evidence relevant to the goal of moving beyond descriptive syndromes in psychiatry, and towards a nosology informed by disease cause.

Funding National Institute of Mental Health.

Introduction Psychiatric nosology arose in central Europe towards the end of the 19th century, in particular with Kraepelin’s foundational distinction between dementia praecox (schizophrenia) and manic depressive insanity.1 The distinction between bipolar illness and unipolar (major) depression was first proposed in the late 1950s and became increasingly widely accepted. The major syndromes—especially schizophrenia, bipolar disorder, and major depression—were differentiated on the basis of their symptom patterns and course of illness. At the same time, clinical features such as psychosis, mood dysregulation, and cognitive impairments were known to transcend diagnostic categories. Doubt remains about the boundaries between the syndromes and the degree to which they signify entirely distinct entities, disorders that have overlapping foundations, or different variants of one underlying disease. Such debates have intensified with syndromes described subsequently, including autism spectrum disorders and attention deficit-hyperactivity disorder.

The pathogenic mechanisms of psychiatric disorders are largely unknown, so diagnostic boundaries are difficult to define. Genetic risk factors are important in the causation of all major psychiatric disorders,7 and genetic strategies are widely used to assess potential overlaps. The imminent revision of psychiatric classifications in the Diagnostic and Statistical Manual of Mental Disorders (DSM) and the International Classification of Diseases (ICD) has reinvigorated debate about the validity of diagnostic boundaries. With increasing availability of large genome-wide genotype data for several psychiatric disorders, shared cause can now be examined at a molecular level.

We formed the Psychiatric Genomics Consortium (PGC) in 2007, to undertake meta-analyses of genome-wide association studies (GWAS) for psychiatric disorders and, so far, the consortium has incorporated GWAS data from more than 19 countries for schizophrenia, bipolar disorder, major depressive disorder, attention deficit-hyperactivity disorder, and autism spectrum disorders. Previous research has suggested...
Several molecular variants confer risk of both schizophrenia and bipolar disorder. autistic spectrum disorders and schizophrenia, including identification of copy number variants conferring risk of both. Findings from family, twin, and molecular studies suggest some genetic overlap between autism spectrum disorder and attention deficit-hyperactivity disorder.

In this first report from the PGC Cross-Disorder Group, we analyse data on genome-wide single-nucleotide polymorphism (SNP) for the five PGC disorders to answer two questions. First, what information emerges when all five disorders are examined in one GWAS? When risk is correlated across disorders, pooled analyses will be better powered than individual-disorder analyses to detect risk loci. Second, what are the cross-disorder effects of variants already identified as being associated with a specific psychiatric disorder in previous PGC analyses? We aimed to examine the genetic relation between the five psychiatric disorders with the expectation that findings will ultimately inform psychiatric nosology, identify potential neurobiological mechanisms predisposing to specific clinical presentations, and generate new models for prevention and treatment.

Methods
Samples and genotypes
The sample for these analyses consisted of cases, controls, and family-based samples assembled for previous genome-wide PGC mega-analyses of individual-level data. cases and controls were not related. For the family-based samples, we matched alleles transmitted to affected offspring (trio cases) with untransmitted alleles (pseudocontrols). We estimated the identity-by-descent relation for all pairs of individuals to identify any duplicate individuals in the component datasets. When duplicates were detected, one member of each set was retained. We then randomly allocated these individuals, with a random number generator, to a disorder case-control dataset. Sample sizes differ from previous reports because of this allocation of overlapping individuals. All patients were of European ancestry and met criteria from the DSM third edition revised or fourth edition for the primary disorder of interest.

To ensure comparability between samples, raw genotype and phenotype data for each study were uploaded to a central server and processed through the same quality control, imputation, and analysis process (appendix). We analysed imputed SNP dosages from 1250 922 autosomal SNPs.

Statistical analysis
In the primary analysis, we combined effects of each disease analysis by a meta-analytic approach that applied a weighted Z-score, in which weights equalled the inverse of the regression coefficient’s standard error. This strategy assumed a fixed-effects model, with weights indicating the sample size of the disease-specific studies.

In a second analytical approach, we did a five-degree-of-freedom test by summing the χ² values for each individual disease meta-analysis. Unlike our primary analysis, this model did not assume that all diseases had the same direction of effect and could detect allelic effects that increase risk for some diseases and decrease risk for others. The appendix describes statistical methods and results, including the handling of trios and population stratification.

To characterise the specificity of the allelic effects for our main findings, we examined the association evidence in three ways: we generated forest plots of the disorder beta coefficients with 95% CIs; we calculated a heterogeneity p value for the disorder-specific effects contributing to the overall statistics for meta-analytic association; and we undertook a multinomial logistic regression procedure with model selection for each main SNP for all five disorders to assess the pattern of phenotypic effects (appendix pp 8–11). To compare the fit of various models of genotype–phenotype associations, we applied established goodness-of-fit metrics (the Bayesian information criteria and the Akaike information criteria). We report the best-fitting model by Bayesian criteria and show results of both metrics for a range of models (appendix pp 38–45, 51–61).

To examine shared polygenic risk at an aggregate level between pairs of diagnoses, we used risk-score profiling as previously described. For each pair, we selected one disorder as a discovery dataset and the other as a target dataset and calculated the proportion of variance in the target set explained by risk scores from the discovery set with a range of statistical cutoffs for SNP inclusion in the score (appendix p 13). To assess the role of specific biological systems in the pathogenesis of the five disorders, we did pathway and eQTL analyses. Pathway analysis was by interval-based enrichment analysis (INRICH) for the full dataset consisting of linkage disequilibrium segments containing signals with association p<10⁻³ in the primary meta-analysis. INRICH accounts for potential genomic confounding factors, such as variable gene and pathway sizes, SNP density, linkage disequilibrium, and physical clustering of biologically related genes (appendix pp 14–16). We did eQTL enrichment analysis to assess whether SNPs associated with five psychiatric disorders were enriched for regulatory SNPs in post-mortem brain tissue.

For INRICH see http://atgu.mgh.harvard.edu/inrich
samples compared with those with no association. To assess the specificity of this finding, we also examined eQTL datasets from three non-brain-tissue types: liver, skin, and lymphoblastoid cell lines (appendix pp 17–21).

Role of the funding source
The sponsor of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data and had final responsibility for the decision to submit for publication.

Results
The final dataset consisted of 33,332 cases and 27,888 controls (including pseudocontrols formed from non-transmitted alleles) distributed among the five disorder groups: autism spectrum disorders (4788 trio cases, 4788 trio pseudocontrols, 161 cases, 526 controls), attention deficit-hyperactivity disorder (1947 trio cases, 1947 trio pseudocontrols, 840 cases, 688 controls), bipolar disorder (6990 cases, 4820 controls), major depressive disorder (9227 cases, 7383 controls), and schizophrenia (9379 cases, 7736 controls). The results of the primary fixed-effects meta-analysis for all five disorders, incorporating seven multidimensional scaling components as covariates, yielded a genomic control value of $\lambda=1.167$. The $\lambda_{1000}$ ($\lambda$ rescaled to a sample of 1000 cases and 1000 controls) was 1.005 (appendix p 22).

In view of evidence for substantial polygenic contributions to common psychiatric disorders, this estimate probably shows the aggregate small effect of a large number of risk variants, although a moderate degree of population stratification or technical bias cannot be excluded.

Figure 1 shows the Manhattan plot of the primary results. Four independent regions contained SNPs...
with $p<5\times10^{-8}$ (table 1; appendix pp 34–35, 25–33). The strongest association signal was on chromosome 3 at an intronic SNP within ITIH3 (table 1). This SNP is in linkage disequilibrium with SNPs encompassing several genes across a 1 Mb region (appendix p 22). The second strongest signal was in an intron of AS3MT on chromosome 10q24 (table 1). Linkage disequilibrium around this associated region encompasses several genes including...
CNNM2. We also recorded genome-wide significant association within CACNA1C, and finally detected significant association to a second locus on chromosome 10 in an intron of CACNB2 (table 1). We undertook conditional analyses to assess evidence for multirisk loci in a region. In these analyses, we included the most strongly associated or peak SNP plus any SNPs within 1·5 Mb of the peak SNP with association p values less than 10⁻⁴ and r² less than 0·2 with the peak SNP based on HapMap 3 CEU data. For the chromosome 3p21 region, and regions CACNA1C and CACNB2, no additional independent association signals were detected. For the chromosome 10q24 region, an additional SNP (rs11191732), about 600 kb from the peak SNP, showed association after conditioning on the peak SNP (rs11191454) with a p value of 6·60×10⁻⁶ before conditioning. For CACNA1C and CACNB2, no additional regions were detected. For the chromosome 3p21 region, no additional independent association signals were detected. For the chromosome 10q24 region, an additional SNP (rs11191732), about 600 kb from the peak SNP, showed association after conditioning on the peak SNP (rs11191454) with a p value of 6·60×10⁻⁶ before conditioning and 3·88×10⁻⁵ after conditioning. Several loci previously implicated in PGC analyses of schizophrenia and bipolar disorder⁶,⁷ showed evidence for association in the cross-disorder analysis, despite not exceeding the cutoff for genome-wide significance (appendix pp 23–24). These loci include one near MHC region on chromosome 6, and SYNE1 (appendix pp 23–24). The five-degree-of-freedom χ² test did not identify any additional genome-wide significant SNPs with effects in the opposite direction among the five disorders (appendix pp 36–37).

Forest plots for genome-wide significant SNPs showed the same direction of effect for most or all of the five disorders (figure 2). For three of the four associated regions, the meta-analysis heterogeneity p value was not significant and a model in which all five disorders contributed provided the best fit (table 1). The exception was rs1024582, for which the heterogeneity p value was significant and the best-fit model supported an effect limited to bipolar disorder and schizophrenia (table 1).

Appendix pp 38–45 shows the profile of the Bayesian and Akaike information criteria measures for each SNP for a range of models. We examined the association between the five disorders of four SNPs showing genome-wide significant association with bipolar disorder (table 2) and ten associated with schizophrenia (table 2). Appendix pp 46–61 show forest plots and model-fitting results by disorder. The best-fitting model for seven of the 14 risk SNPs suggested disorder-specific effects for either bipolar disorder (three SNPs) or schizophrenia (four SNPs), whereas the rest were consistent with more pleiotropic models (table 2).

Figure 3 shows the proportion of variance explained in target sets (Nagelkerke’s pseudo R² from logistic regression) by risk scores from the discovery sets. We noted highly significant overlap of polygenic risk between all three adult disorders (bipolar disorder, major depressive disorder, and schizophrenia), with the strongest effects noted for bipolar disorder and schizophrenia (figure 3). Overlap was reduced but still significant between aggregate genetic risk for autism spectrum disorder with schizophrenia (minimum p<10⁻⁴) and bipolar disorder (p<0·05; figure 3). No consistently significant polygenic overlap was detected between major depressive disorder and autism spectrum disorder or between attention deficit-hyperactivity disorder and any other disorder.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Base-pair position*</th>
<th>Nearest gene</th>
<th>Alleles</th>
<th>Frequency†</th>
<th>p value</th>
<th>Best-fit model (BIC)‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs9321601</td>
<td>6</td>
<td>TCF4</td>
<td>T/G</td>
<td>0·346</td>
<td>4·27×10⁻⁴</td>
<td>BPD</td>
</tr>
<tr>
<td>rs50994397</td>
<td>10</td>
<td>ANK3</td>
<td>T/C</td>
<td>0·065</td>
<td>7·08×10⁻⁵</td>
<td>BPD</td>
</tr>
<tr>
<td>rs47659145</td>
<td>12</td>
<td>CACNA1C</td>
<td>T/C</td>
<td>0·204</td>
<td>1·52×10⁻⁴</td>
<td>BPD, MDD, schizophrenia</td>
</tr>
<tr>
<td>rs12576775</td>
<td>11</td>
<td>ODZ4</td>
<td>G/A</td>
<td>0·125</td>
<td>4·40×10⁻⁴</td>
<td>BPD</td>
</tr>
</tbody>
</table>

Peak SNPs in associated region after clumping. BIC=Bayesian information criteria. BPD=bipolar disorder. MDD=major depressive disorder. ASD=autism spectrum disorders.

*Detected with University of California Santa Cruz Genome Browser (version hg18). †Risk allele frequency in controls. ‡Best-fit multinomial logistic model by BIC (appendix pp 46–50 [bipolar disorder] and 51–61 [schizophrenia] provide a comparison of BIC and Akaike information criteria across models). The BIC for the best and second best models do not differ significantly (i.e., greater than 2 as previously suggested). ‡Mh4765914 is a proxy SNP for rs4765911 based on linkage disequilibrium (481 base pairs away, r²=0·874).

Table 2: Modelling analysis results for single-nucleotide polymorphisms (SNP) showing genome-wide significant association in previous genome-wide association studies from the Psychiatric Genome-Wide Association Study Consortium.
**Figure 3:** Pair-wise cross-disorder polygene analysis
We derived polygene risk scores for each disorder (discovery sets) and applied them sequentially to the remaining disorders (target sets). Results are grouped by each discovery set. Each pair is shown on the x-axis and the proportion of variance explained for the target disorder (estimated via Nagelkerke's pseudo R²) on the y-axis. For purposes of illustration, three pT cutoffs are shown, but appendix p 62 shows the proportion of variance results for a broader range of cutoffs. pT-training set p value (used to select training set SNPs). Significance of results:

- a=p<0.05; b=p<0.01; c=p<0.005; d=p<0.001; e=p<0.0001; f=p<0.00001

ADHD=attention deficit-hyperactivity disorder.

**Discussion**
This study is the largest genome-wide analysis of psychiatric illness so far and the first to provide evidence that specific SNPs are significantly associated with a range of childhood-onset and adult-onset psychiatric disorders. For the five disorders studied, SNPs at four loci—regions on chromosomes 3p21 and 10q24, and SNPs in two L-type voltage-gated calcium-channel subunits, CACNA1C and CACNB2—exceeded the cutoff for genome-wide significance in the primary analysis. The strongest signal was within a region on chromosome 3p21.1. Aggregate polygenic risk scores for a broad set of common variants showed cross-disorder effects for all the adult-onset disorders (bipolar and major depressive disorder, and schizophrenia) and nominally between autism spectrum disorders and both bipolar disorder and schizophrenia.

In view of extensive linkage disequilibrium in the 3p21.1 region, encompassing more than 30 genes, we could not identify the causal locus. Genome-wide significant association to the 3p21.1 region has previously been reported in GWAS of samples overlapping with ours at rs1042779 (12 kb from our peak SNP) for bipolar disorder,\(^5\) rs736408 (2 kb) for a combined bipolar disorder and schizophrenia phenotype,\(^3\) and rs2251219 (248 kb) for a combined major depressive and bipolar disorder phenotype.\(^6\) Reanalysis of this last combined dataset suggested that the signal was largely attributable to the group with bipolar disorder.\(^7\) Furthermore, the association evidence for our peak chromosome 3 SNP rs2535629 was genome-wide significant in a joint analysis of bipolar disorder and schizophrenia samples done by the PGC schizophrenia group (p=6.1×10⁻⁹).\(^8\)

Our model-selection analysis was designed to characterise the range of phenotypic effects for loci that showed significant association; however, the statistical evidence and effect size for each contributing disorder can vary. Although the best-fit model for the chromosome
3p21 region included all five disorders, interpretation of these results is complicated by evidence from a PGC GWAS mega-analysis of major depressive disorder. In the discovery phase of that analysis, which consisted of 9240 major depressive disorder cases and 9519 controls many of whom overlap with samples reported here, the smallest association p value for this region (rs2535629) was 0.00013. However, no association was noted in a replication dataset of 6783 cases and 50695 controls (p=0.70) for that disorder, and the combined discovery and replication phase p value was 0.0031. Thus, any association between this region and major depressive disorder is unclear.

Two of the four genome-wide significant signals in our analysis localise to introns of brain-expressed genes encoding L-type voltage-gated calcium-channel subunits (CACNAIC and CACNB2). Previous disorder-specific GWAS, overlapping with the samples included here, identified CACNAIC as a susceptibility gene for bipolar disorder,13 schizophrenia,14 and major depressive disorder.15 Gain-of-function mutations in CACNAIC cause Timothy syndrome, a developmental disorder in which the phenotypic range includes autism.16 Consistent with a pleiotropic role, neuroimaging studies have documented effects of CACNAIC variants on a range of structural and functional brain phenotypes, including circuitry involved in emotion processing,17 executive function,18 attention,19 and memory.20 CACNB2 encodes an auxiliary voltage-gated calcium-channel subunit that interacts with L-type calcium-channel subunits (including CACNAIC, CACNAID, and CACNA1S) to promote their trafficking to the plasma membrane, increase their function, and regulate their modulation by other signalling proteins and molecules.21 Although previous PGC analyses (schizophrenia and bipolar disorder) did not identify CACNB2 as a risk gene, a variant in CACNB2 (52 kb from our peak SNP) was one of the main signals in an independent GWAS of bipolar disorder in Han Chinese individuals.18

The pleiotropic effects of voltage-gated calcium channels on childhood-onset and adult-onset psychiatric disorders are underscored by pathway analysis in which calcium-channel activity genes, including our main two L-type subunit genes, showed significant enrichment in the five disorder dataset. The PGC analysis of bipolar disorder reported enrichment of a pathway including CACNAIC and CACNAID; importantly, however, we detected enrichment of these genes after exclusion of the bipolar disorder dataset. Thus, our results suggest that voltage-gated calcium signalling, and, more broadly, calcium-channel activity, could be an important biological process in psychiatric disorders. A fourth region associated with cross-disorder effects was on chromosome 10, encompassing several genes with the peak signal in an intron of AS3MT. Loci previously associated with schizophrenia and bipolar disorder22 had varying evidence of association for the other major psychiatric disorders. For example, a locus previously strongly associated with schizophrenia, encompassing MIR137 and DPYD on chromosome 1, showed similar evidence of association with autism spectrum disorders; this finding is consistent with reports that autism spectrum disorders are related to microdeletions of this region.38,39 Accumulating evidence, including that from clinical, epidemiological, and molecular genetic studies, suggests that some genetic risk factors are shared between neuropsychiatric disorders. Genome-wide studies have identified rare copy-number variants that confer risk of several neuropsychiatric disorders including autism, attention deficit-hyperactivity disorder, epilepsy, intellectual disability, and schizophrenia.9 Our analyses of 14 SNPS previously identified as being genome-wide significantly associated with schizophrenia and bipolar disorder suggest that some loci identified in studies of individual disorders will have broader phenotypic effects.
Our results suggest a diversity of findings, with some SNPs showing diagnostic specificity and others pleiotropic effects on two or more of the five disorders.

These results should be interpreted in consideration of several limitations. First, we compared models of cross-disorder effects on the basis of the most often used goodness-of-fit measures, but other criteria might yield different results. For all four of the risk loci identified in the primary meta-analysis, the selected model had a substantially better fit than any alternative models had. However, for some loci that did not reach genome-wide significance, the difference in fit between our best-fitting and alternative models was moderate (appendix pp 38–45), so more than one model could be consistent with the noted effects. Second, diagnostic misclassification (eg, reciprocal misdiagnosis of cases of schizophrenia and bipolar disorder) could produce spurious evidence of genetic overlap between disorders. However, a substantial degree of misdiagnosis would be needed to account for our findings of loci that affect all or subsets of five disorders whose diagnostic criteria are fairly distinct. Third, the five disorders we examined were limited to those for which large-scale GWAS datasets have been assembled by the PGC and processed through a uniform quality-control process. As further datasets become available, more comprehensive analyses of cross-disorder genetic effects on psychiatric illness should be pursued. Fourth, we restricted our analyses to individuals of European ancestry. Whether our findings apply to other populations is unknown. Finally, GWAS designs are suited to identify common variant aspects of genetic architecture; further studies (including analyses of copy-number variants and rare mutations) will be needed to account more completely for shared genetic contributions across disorders. As in almost all GWAS of complex disorders reported so far, the effect sizes of genome-wide significant loci are individually quite small and the variance they account for is insufficient for predictive or diagnostic usefulness by themselves. However, our study is the first large-scale effort to characterise allelic effects across five psychiatric disorders, incorporating single locus, multilocus, and pathway analyses.

The identification of genetic variants that confer risk of a diverse set of psychiatric disorders parallels findings from other medical specialties. Most notably, GWAS of autoimmune disorders have shown extensive overlap in genetic variants that affect a diverse range of diseases, including rheumatoid arthritis, coeliac disease, multiple sclerosis, systemic lupus erythematosus, psoriasis, Crohn’s disease, and type 1 diabetes. Our results provide insights into the shared causation of psychiatric disorders (panel). In particular, alterations in calcium-channel signalling could represent a fundamental mechanism contributing to a broad vulnerability to psychopathology.

Contributors


Statistical analytic: Stephan Ripke (lead), Kenneth Kendler, Phil Hyun Lee, Benjamin M Neale, Shaun Purcell. Editorial revisions: Richard Amney, Jan Butelaar, Ayman Fanous, Stephen V Faraone, Witte Hoogendijk, Klaus-Peter Lesch, Douglas F Levinson, Roy H Perlis, Shaun Purcell, Marcella Rietschel, Brien Riley, Edmund Sonuga-Barke, Russell Schachar, Thomas G Schulze, Anita Thapar. PGC coordinating committee: Patrick F Sullivan (chair), Patrick Bender, Sven Cichon, Nicholas Craddock, Mark J Daly, Stephen V Faraone, John Kelsoe, Thomas Lehrner, Douglas Levinson, Mick O’Donovan, Pablo Gejman, Jonathan Sebat, Pamela Sklar, Jordan W Smoller. See appendix for PGC Collaborators from Analysis Committee (Mark J Daly, chair), ADHD Workgroup (Stephen V Faraone, chair), Autism Workgroup (Mark Daly, Bernie Devlin, cochairs), Bipolar Disorder Workgroup (John Kelsoe, Pamela Sklar, cochairs), Major Depressive Disorder Workgroup (Patrick Sullivan, chair), Schizophrenia Workgroup (Michael O’Donovan, chair).

Conflicts of interest

We declare that we have no conflicts of interest.

Acknowledgments

We thank the study participants, and the research staff at the many study sites. This work was supported by NIMH grant U01 MH085520. Numerous grants from the National Institutes of Health (USA), and similar numbers of government grants from other countries, and substantial private and foundation support enabled this work.

References


20 Nicolae DL, Garneau E, Zhang W, Duan M, Cox NJ. Trait-related SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* 2010; 6: e1000888.


