A Genome-wide Association Analysis of a Broad Psychosis Phenotype Identifies Three Loci for Further Investigation

Psychosis Endophenotypes International Consortium[a,b] and the Wellcome Trust Case-Control Consortium 2[a,b]

Background: Genome-wide association studies (GWAS) have identified several loci associated with schizophrenia and/or bipolar disorder. We performed a GWAS of psychosis as a broad syndrome rather than within specific diagnostic categories.

Methods: 1239 cases with schizophrenia, schizoaffective disorder, or psychotic bipolar disorder; 857 of their unaffected relatives, and 2739 healthy controls were genotyped with the Affymetrix 6.0 single nucleotide polymorphism (SNP) array. Analyses of 695,193 SNPs were conducted using UNPHASED, which combines information across families and unrelated individuals. We attempted to replicate signals found in 23 genomic regions using existing data on nonoverlapping samples from the Psychiatric GWAS Consortium and Schizophrenia-GENE-plus cohorts (10,352 schizophrenia patients and 24,474 controls).

Results: No individual SNP showed compelling evidence for association with psychosis in our data. However, we observed a trend for association with same risk alleles at loci previously associated with schizophrenia (one-sided $p = .003$). A polygenic score analysis found that the Psychiatric GWAS Consortium's panel of SNPs associated with schizophrenia significantly predicted disease status in our sample ($p = 5 \times 10^{-14}$) and explained approximately 2% of the phenotypic variance.

Conclusions: Although narrowly defined phenotypes have their advantages, we believe new loci may also be discovered through meta-analysis across broad phenotypes. The novel statistical methodology we introduced to model effect size heterogeneity between studies should help future GWAS that combine association evidence from related phenotypes. Applying these approaches, we highlight three loci that warrant further investigation. We found that SNPs conveying risk for schizophrenia are also predictive of disease status in our data.

Key Words: Bipolar disorder, genome-wide association, meta-analysis, polygenic score analysis, psychosis, schizophrenia

Psychotic disorders including schizophrenia, bipolar disorders, and schizoaffective disorders affect approximately 3% of the general population (1–6) and constitute the most severe forms of mental diseases. Characteristic symptoms include hallucinations, delusional beliefs, and severe mood variations and cognitive impairments, all of which can lead to major changes in behavior and ability to function. According to the World Health Organization's World Health Report, these psychotic disorders are ranked within the top seven leading causes of disability in young adults (7).

The genetic architecture of schizophrenia and bipolar disorder has been shown to include common alleles of subtle effect and rare mutations of large effect, often involving genome copy number variation (8–11). Recent large-scale meta-analyses of schizophrenia (12), conducted by the Psychiatric GWAS Consortium (PGC), combined data from more than 50,000 individuals from 17 international cohorts (13–25) and identified seven associated loci. Of these loci, five were new, and two had been previously implicated. The strongest new finding in schizophrenia was within an intron of a putative primary transcript for *MIR137* (microRNA 137), a known regulator of neuronal development. Four other schizophrenia loci with strong statistical support contain predicted targets of *MIR137*, suggesting *MIR137*-mediated dysregulation as a previously unknown etiologic mechanism in schizophrenia. The meta-analysis (12) also confirmed the role of the major histocompatibility complex (MHC) region, as suggested in other studies (23,24,26,27), as well as a marker in intron four of transcription factor 4 (*TCF4*) (24).

The PGC conducted a similar meta-analysis for bipolar disorder (28) including more than 11,000 cases and 51,000 controls from previous association studies (15,29–41). The analysis confirmed an association with *CACNA1C* and identified a new intronic variant in *ODZ4*. An overlap in the polygenic component between schizophrenia and bipolar disorder was also found (42,43). In a combined meta-analysis of both schizophrenia and bipolar disorder, three loci reached genome-wide significance: *CACNA1C*, *ANK3*, and the *ITIH3-ITIH4* region (28).

As data accumulate, there is increasing evidence for overlap in the genetic component to risk between different psychiatric disorders (44–46). When combined with epidemiologic and neuroimaging data (47–50), the shared genetic architecture supports the view of schizophrenia, bipolar disorder, and other psychoses as related rather than etiologically distinct entities (8,12,28,46,51–59). Motivated by these findings, we performed a genome-wide association study (GWAS) of psychotic disorders including patients with schizophrenia, schizoaffective disorder, and bipolar disorder with a history of psychotic symptoms.

Methods and Materials

The Cohort

Before any exclusion, the full data set included 6935 participants with 1820 patients, 1224 of their unaffected relatives, and 3891

[a]A full list of individuals who qualify for authorship, along with their affiliations, is provided in the Acknowledgments.

[b]A full list of members of the PEIC and WTCCC2 is provided in the Acknowledgments.

Address correspondence to Elvira Bramon, M.D., Ph.D., Mental Health Sciences Unit & Institute of Cognitive Neuroscience, University College London, London W1W 7EJ, United Kingdom; E-mail: e.bramon@ucl.ac.uk.

Received Oct 23, 2012; revised Mar 27, 2013; accepted Mar 30, 2013.

0006-3223/$36.00
http://dx.doi.org/10.1016/j.biopsych.2013.03.033

healthy control subjects. These samples were collected through seven centers across Australia and Europe (Germany, Holland, Spain, and United Kingdom). Participants provided written informed consent, and the study was approved by the respective ethical committees at each of the seven participating centers. After quality control, the full sample included 4835 participants of which 1239, 857, and 2739 were patients, their unaffected relatives, and healthy control subjects, respectively. Additional sample and center details are provided in Table S1 of Supplement 1.

Inclusion and Exclusion Criteria and Phenotype Definition

To allow for a DSM-IV (60) diagnosis to be ascertained or ruled out, all participants (including controls and unaffected family members) underwent a structured clinical interview with either the Schedule for Affective Disorders and Schizophrenia or the Structured Clinical Interview for DSM Disorders, or the Schedules for Clinical Assessment in Neuropsychiatry (61–63). Of the cases passing quality control, 784 met criteria for schizophrenia, 113 for bipolar disorder with a history of psychotic symptoms, 110 for psychotic disorder not otherwise specified, 97 for schizophreniform disorder, 64 for schizoaffective disorder, 44 for brief psychotic disorder, 20 for delusional disorder, and 7 for substance-induced psychosis. Participants in all groups were excluded if they had a history of neurologic disease or head injury resulting in loss of consciousness.

DNA Sample Preparation

Genomic DNA obtained from blood for all participants was sent to the Wellcome Trust Sanger Institute, Cambridge, United Kingdom. Samples were processed in 96-well plate format and each plate carried a positive and a negative control. DNA concentrations were quantified using a PicoGreen assay (Invitrogen, Life Technologies, Grand Island, New York) and an aliquot assayed by agarose gel electrophoresis. A sample passed quality control if the original DNA concentration was at least 50 ng/μL and the DNA was not degraded.

Genotyping Methodology and Quality Control

To track sample identity, 30 single nucleotide polymorphisms (SNPs) including sex chromosome markers were typed on the Sequenom platform before entry to the whole genome genotyping pipeline. Of the initial 6935 samples, 347 failed quality control due to degraded or insufficient DNA or incorrect sex classification. The remaining samples were sent for genotyping with the Genome-wide Human SNP Array 6.0 at Affymetrix Services Lab (http://www.affymetrix.com).

Data Quality Control

See Tables S2 and S3 in Supplement 1. Genotype calling was conducted using the CHIAMO algorithm (64,65) modified for use with the Affymetrix 6.0 genotyping array. We excluded 11,610 SNPs with a study-wide missing data rate over 5%. We removed 26,858 SNPs with four or more Mendelian inheritance errors identified with Pedstats (66). Additional exclusion criteria were departure from Hardy-Weinberg equilibrium ($p < 10^{-6}$) or minor allele frequency (MAF) <.02 with 2404 and 145,097 SNPs removed, respectively. A total of 38,895 SNPs from the X or Y chromosomes or mitochondrial DNA were also excluded from the analysis. Finally, 9499 poorly genotyped SNPs were removed following visual inspection of the genotyping intensity plots in the program Evoker (67).

We excluded 214 samples with more than 2% missing data across all SNPs. Another 70 samples were excluded due to divergent genome-wide heterozygosity (inbreeding coefficients were $F > .076$ or $F < -.076$ as estimated with PLINK (42). Chromosomal sharing was inferred from a genome-wide subset of 71,677 SNPs and from each duplicate pair the sample with the most complete genotype data was kept. We removed 70 duplicates and monozygotic twins by excluding one of each pair of individuals showing identity by descent greater than 95%.

After quality control, 4835 individuals remained. Initial analysis of the genotype data identified a high fraction of samples (approximately 30%), which showed poor signal-to-noise ratio in the genotyping assay. Because the experimental source of the problem was unclear and to ensure a robust set of genotype calls, these samples were removed from further analysis. We note that the sample loss was randomly distributed across the three clinical groups (32% of patients, 30% of relatives, and 30% of controls; χ^2 (2 *df*) = 3.2; $p = .20$). Full details on the sample quality control are provided in Table S2 of Supplement 1.

In addition to 3490 unrelated individuals, there were 1345 related individuals clustered in 462 families. The family size ranged between 2 and 5 with an average of 2.9 members. Of the families, 196 were control families, 243 had one affected case only, 21 families included two cases, and another 2 families had three cases. Data from these individuals were analyzed at 695,193 autosomal SNPs.

Population Structure Analysis

To investigate the genetic structure in the data, we performed principal component (PC) analysis (PCA) of unrelated individuals using EIGENSOFT version 3.0 (68) on a thinned set of SNPs (see Methods and Materials in Supplement 1). Owing to the multicenter nature of our study, we assessed the need to include PCs as covariates in statistical tests of association to control for population stratification (69). This was done by using PLINK (42) to calculate the genome-wide distribution of the association test statistic in the unrelated individuals using different numbers of PCs as covariates. Possible inflation in the test statistic was measured by the genomic control parameter λ, which is the ratio of the median of the observed test statistic distribution to that of its expectation under the null hypothesis (70).

Association Analysis in Our Discovery Sample

A genome-wide association analysis was conducted with UNPHASED v3.1.4. (71), which allows a combined analysis of both families and unrelated individuals, thus increasing statistical power. UNPHASED calculates separately the transmitted and untransmitted alleles in families as well as the allelic frequencies in unrelated patients and controls, giving a combined odds ratio, 95% confidence interval, and *p* value. The analysis included three PCs as covariates.

For SNPs showing association with psychosis in our data with $p < 1 \times 10^{-4}$, proxy SNPs were identified using the proxy report routine in PLINK (42). Only those SNPs that were in linkage disequilibrium ($r^2 \geq .5$) with and within 100-kb distance from at least one such proxy SNP that showed association with psychosis with $p < 1 \times 10^{-2}$ were selected for the replication phase. These criteria reduced the possibility that the association signal was driven by an artifact at the most associated SNP.

We attempted to replicate 44 SNPs included in the catalogue of published GWAS (72), accessed in January 2012, for schizophrenia or bipolar disorder with *p* values less than 1×10^{-7}. These SNPs and the studies that identified them are listed in Table 1. If a reported SNP was not genotyped directly in our data, we used the 1000 Genomes Project (73) data to identify the best tag (highest r^2) and orientated the haplotype to the risk allele so that the directions of the odds ratios were matched between our analysis and the previous studies. Where relevant information was not available, the SNP was excluded from analysis.