Priority Communication

Biological Psychiatry

An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis

Evangelos Vassos, Marta Di Forti, Jonathan Coleman, Conrad Iyegbe, Diana Prata, Jack Euesden, Paul O'Reilly, Charles Curtis, Anna Kolliakou, Hamel Patel, Stephen Newhouse, Matthew Traylor, Olesya Ajnakina, Valeria Mondelli, Tiago Reis Marques, Poonam Gardner-Sood, Katherine J. Aitchison, John Powell, Zerrin Atakan, Kathryn E. Greenwood, Shubulade Smith, Khalida Ismail, Carmine Pariante, Fiona Gaughran, Paola Dazzan, Hugh S. Markus, Anthony S. David, Cathryn M. Lewis, Robin M. Murray, and Gerome Breen

ABSTRACT

BACKGROUND: Polygenic risk scores (PRSs) have successfully summarized genome-wide effects of genetic variants in schizophrenia with significant predictive power. In a clinical sample of first-episode psychosis (FEP) patients, we estimated the ability of PRSs to discriminate case-control status and to predict the development of schizophrenia as opposed to other psychoses.

METHODS: The sample (445 case and 265 control subjects) was genotyped on the Illumina HumanCore Exome BeadChip with an additional 828 control subjects of African ancestry genotyped on the Illumina Multi-Ethnic Genotyping Array. To calculate PRSs, we used the results from the latest Psychiatric Genomics Consortium schizophrenia meta-analysis. We examined the association of PRSs with case-control status and with schizophrenia versus other psychoses in European and African ancestry FEP patients and in a second sample of 248 case subjects with chronic psychosis.

RESULTS: PRS had good discriminative ability of case-control status in FEP European ancestry individuals (9.4% of the variance explained, $p < 10^{-6}$), but lower in individuals of African ancestry ($R^2 = 1.1\%$, $p = .004$). Furthermore, PRS distinguished European ancestry case subjects who went on to acquire a schizophrenia diagnosis from those who developed other psychotic disorders ($R^2 = 9.2\%$, $p = .002$).

CONCLUSIONS: PRS was a powerful predictor of case-control status in a European sample of patients with FEP, even though a large proportion did not have an established diagnosis of schizophrenia at the time of assessment. PRS was significantly different between those case subjects who developed schizophrenia from those who did not, although the discriminative accuracy may not yet be sufficient for clinical utility in FEP.

Keywords: Genetics, GWAS, Polygenic score, Psychosis, Risk prediction, Schizophrenia

http://dx.doi.org/10.1016/j.biopsych.2016.06.028

Recent developments in genetics research, particularly genome-wide association studies (GWASs), have greatly improved our understanding of the genetic architecture of complex disorders such as schizophrenia. The additive contributions of hundreds or thousands of polymorphisms, regulating different biochemical pathways related to the phenotype, determine the genetic liability to complex disorders (1). It has been established that schizophrenia is highly polygenic, with many common genetic variants contributing to the risk of the disease. In the latest meta-analysis of GWASs for schizophrenia (2), 108 independent regions associated with the disease were identified.

Risk prediction remains a primary focus of genetic studies. In schizophrenia, this has been largely based on family history, but with the progress in GWASs, an increasing number of susceptibility variants have been found that contribute to risk prediction (3). However, each genetic marker individually explains only a tiny proportion of the genetic variation with insignificant predictive power (4). For this reason, methods have been developed to examine disorder prediction by genetic variants en masse, via summarizing variation across many nominally associated loci into quantitative scores that are tested in independent samples (5). One such approach is the generation of polygenic risk scores (PRSs), which represents a promising technique for predicting risk (6,7).

PRSs have been successfully associated with schizophrenia, and as the size of the discovery sample increases, their accuracy and predictive power improve. For example,

ISSN: 0006-3223

from explaining approximately 3% of the variance of a case-control sample in 2009 (5), they now explain approximately 18% (2). To establish whether PRSs may be a useful tool for risk prediction, replication and further studies in independent samples are necessary. Importantly, these samples should represent the typical patients we see in the clinical practice rather than the severe end of the phenotype that is more easily identified and recruited for research (8). For this reason, in a sample of patients recruited during a first episode of psychosis (FEP) and ancestry-matched control subjects from South London, we measured the ability of PRS to discriminate case subjects from control subjects and among case subjects to discriminate schizophrenia from other psychoses.

METHODS AND MATERIALS

Sample Description

Participants were recruited as part of the Biomedical Research Centre (BRC) for Mental Health Genetics and Psychosis (GAP) study (9). The study systematically recruited patients aged 18 to 65 years who presented to adult psychiatric services in the South London and Maudsley National Health Service Foundation Mental Health Trust between December 2005 and October 2011 with a first episode of nonorganic psychosis (ICD-10 codes: F20–F29 and F30–F33) (10) and unaffected control subjects. This is a multi-ethnic sample, reflecting the demographic characteristics of the area. Clinical diagnoses of case subjects were validated using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interview (11), and control subjects were screened with the Psychosis Screening Questionnaire (12). Case subjects who met criteria for organic psychosis, intellectual disability (IQ $<$ 70), or transient psychosis ($<$7 days of symptoms) and control subjects who reported previous diagnosis of psychosis or had a first-degree relative with psychosis were excluded.

Because the diagnostic evaluation of FEP patients is difficult due to the short history of illness and variable symptoms seen (13), the following diagnostic approaches were used: 1) consensus diagnoses based on discussions between experienced clinicians who interviewed the patients using the SCAN to collect symptoms characteristics, frequency, and duration over the 4 weeks preceding the assessment; using the SCAN, Present State Examination Data, and applying the Operational Criteria Checklists (OPCRIT) computerized algorithms (14) to obtain diagnoses according to 2) DSM-IV and 3) ICD-10 classification systems and 4) clinical diagnoses made by the treating psychiatrists collected retrospectively from the electronic medical records of the patients.

This study was granted ethical approval by the South London and Maudsley and Institute of Psychiatry Local Research Ethics Committee. All individuals included gave informed written consent to be assessed at baseline and to be contacted again at follow-up; they gave us permission to access their clinical records and to publish data originating from the study.

For replication of the utility of PRS to discriminate between schizophrenia and other psychoses, a second sample recruited from the same geographical area, the IMPACT (Improving Physical Health and Reducing Substance Use in Psychosis) study (15), was used. This comprises 280 patients with chronic psychosis (mean illness duration, 16 years) who participated in a randomized controlled trial of a psychosocial health promotion intervention. Diagnoses were extracted from the documented ICD-10 diagnosis in the clinical notes at the time of recruitment (16). Because this was a case-only sample, PRSs from the IMPACT study were compared with control subjects from the GAP sample. Both samples were genotyped on the same array, and genotypic data were processed and analyzed together.

Because the GAP sample included only 70 African European control subjects, we obtained a second sample of sub-Sahara African ancestry control subjects collected from the same geographical area for the South London Ethnicity and Stroke Study (SLESS) (17). The control subjects were recruited by random sampling of general practitioner lists from South London, and these data are available as part of a collaboration between the BRC for Mental Health and the Guy's and St Thomas' National Health Service Foundation Trust BRC (see detailed description of this cohort in the Supplement).

Genotyping Procedures

DNA was extracted from blood or cheek swabs (80% and 20% of the GAP sample, respectively). When several extractions for the same individual were performed, we used DNA from blood. The samples were genotyped at the South London and Maudsley NHS Trust/King's College London BRC Genomics Laboratory on the Illumina HumanCore Exome BeadChip. This array provides genetic data for identified genome-wide significant single nucleotide polymorphisms (SNPs), a cost-effective genome-wide coverage of 265,000 highly informative tag SNPs, plus 245,000 rare, predicted deleterious variants. The latter were excluded from our analysis. Genotypes were processed using the GenomeStudio Analysis software version 2011.1 (Illumina Inc., San Diego, CA).

Quality control (QC) included exclusion of SNPs with minor allele frequency (MAF) $<$1%, SNPs and individuals with genotypic failure $>$1%, SNPs with Hardy Weinberg equilibrium $p < 10^{-5}$ in control subjects, mismatch between recorded and genotypic sex, and related individuals. Cryptic relatedness and duplicated samples were identified with pairwise identity by descent method (pi-hat $>$ .1875). Imputation was performed with IMPUTE2 (18) based on the 1000 Genomes phase 3 reference panel (19), using haplotypes from all the ancestral populations (20). The imputed markers underwent a second stage of QC to exclude SNPs that were missing in $>$5% of individuals or had imputation information score (INFO) $<$0.8. QC was performed with PLINK 1.9 (https://www.cog-genomics.org/plink2) (21).

The SLESS sample was genotyped using the Illumina Multi-Ethnic Genotyping Array, a multi-ethnic platform with $>$1.7 million markers (http://www.illumina.com/products/infinium-multi-ethnic-global-array.html). After repeating the above QC procedures, we merged the two samples using only the markers that had been genotyped in both arrays. We excluded any related individuals between the two datasets. We excluded any markers that differed between the two African control groups (detailed QC methods can be found in the Supplement).

Calculation of PRSs

We used the latest Psychiatric Genomics Consortium (PGC2) schizophrenia meta-analysis (2) as discovery sample to calculate