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

PRSs for each individual in our sample as the weighted sum of the risk alleles they carried. Sample relatedness between the GAP data and PGC2 was tested using the GWAS data. We identified that 80 individuals were already included in PGC2 as part of the Wellcome Trust Case Control Consortium 2 (22); hence, we used the PGC2 leave-one-out discovery dataset, excluding the Wellcome Trust Case Control Consortium 2 sample. To use informative and independent markers, we selected SNPs with MAF >10% and INFO >0.9, according to the PGC2 protocol, and we included only one SNP from the major histocompatibility complex linkage disequilibrium (LD) region on chromosome 6 (hg19; chr6:27-33Mb).

Population stratification was corrected with principal component analysis (PCA) using EIGENSTRAT (23). Two subsamples were selected for further analysis based on the loadings on the first two principal components (PCs), one with European-only ancestry and one with African ancestry (combining African and African Caribbean origin). To further correct for stratification, we repeated the PCA in each of the two subsamples separately, and we retained 10 PCs for each, which were used as covariates in the PRS analyses.

We performed clumping separately in the European and African subsamples by retaining the SNP with the smallest p value from each LD block (excluding SNPs with $r^2 > .1$ in 250-kb windows). Each allele was weighted by the logarithm of the odds ratio (OR) as estimated in the PGC2 study. Ten different PRSs, using subsets of the total SNPs based on different p value thresholds for the association with schizophrenia in the discovery PGC sample, were computed and were compared for the best discrimination between case and control subjects. To construct PRS, we used the PRS software (PRSice; http://prsice.info/) (24).

Statistical Analysis

Association of PRSs with case-control status was performed with logistic regression, and Nagelkerke's pseudo-R^2 was calculated to measure the proportion of variance explained. To estimate heritability (i.e., variance explained at the liability scale) assuming a liability-threshold model, a lifetime risk of 1% for psychosis and 0.72% for schizophrenia (25), and adjusting for case-control ascertainment, we used the GEnetic Analysis Repository software (http://sourceforge.net/p/gbchen/wiki/GEAR/; Supplemental Table S1). For each analysis we estimated and analyzed PRSs at 10 different levels of significance at the discovery sample. To correct for multiple hypothesis testing in each sample (European FEP, African FEP, European chronic psychosis), we estimated the equivalent number of effective tests using the correlation matrix of PRS at the 10 p value thresholds (http://gump.qimr.edu.au/general/daleN/matSpD/) and performed Bonferroni correction on the sum of effective tests across the three samples. The effective number of independent variables were 5, 6.4, and 5.3, respectively, and the significance threshold required to keep type I error rate at 5% was .003.

To evaluate the specificity of PRS to schizophrenia, we divided case subjects according to each diagnostic approach (consensus, OPCRIT/DSM, OPCRIT/ICD, clinical) into two diagnostic categories, schizophrenia and other psychoses, including delusional disorder, acute and transient psychotic disorder, schizoaffective disorder, other nonorganic psychotic disorder, bipolar disorder, and psychotic depression. We explored standardized mean differences (SMDs) of the PRSs between the case subjects stratified in the two diagnostic categories and control subjects (each diagnostic category was compared with all the control subjects), and we compared with logistic regression PRSs between case subjects who had met criteria for schizophrenia with at least one of the four diagnostic approaches and case subjects with any other psychosis.

To better visualize the effect of PRS on the risk of psychosis, we estimated case-control OR and 95% confidence intervals dividing our sample in quintiles by PRS. To be able to compare our estimates with the outcomes of previous studies (2,26) and to adjust for oversampling of control subjects approximating a prevalence of psychosis of 0.72% (25), we used a simulation method to extract deciles from our observed data. Detailed statistical methods are presented in the Supplement.

RESULTS

Genotyping and Quality Control

From quality control procedures, 710 individuals (445 case subjects with FEP and 265 control subjects) and 290,871 genotyped markers with MAF >1% were available for analysis from the GAP study and 248 case subjects from the IMPACT study. After imputation 5.2×10^6 variants with INFO score >0.8 were identified and reduced with clumping to 56,059 independent variants for PRS analysis. Sex, DNA origin, age, and ethnic distribution of case and control subjects are presented in Table 1. The GAP sample sex, ethnicity, and age distribution well represent a FEP adult population from the study geographical area as indicated by previously published studies looking at this patient group (27). After identifying the two main ethnic groups in the GAP study (Supplemental Figure S2) and performing separate PCA in each, we excluded 11 additional outliers in the European subsample, leaving 328 (161 case and 167 control subjects) individuals for the final analysis from GAP and 131 case subjects from IMPACT. For the African subsample, 189 case and 69 control subjects from GAP as well as 828 control subjects from SLESS were included in the analysis.

Polygenic Prediction of Case-Control Status

In the overall GAP sample the best discrimination of case-control status by the PRS was achieved with the inclusion of all the independent markers after clumping (n = 65,479), which explained 3.6% of the variance (Nagelkerke's pseudo-R^2; $p = 10^{-5}$). When we stratified our analysis by the two main ethnic groups in the GAP study, we observed that PRS explained 9.4% of the variance in individuals of European ancestry (5.2% on the liability scale, $p = 9.5 \times 10^{-7}$), but it was not predictive of case-control status in individuals of African ancestry (p = .08). Optimal discrimination was achieved with the inclusion of markers with a p value significance threshold (p_T) <.1. After the expansion of the African control subjects with the inclusion of SLESS, PRS explained 1.1% of the variance (p = .004) in individuals of African descent (Supplemental Table S2). The bar plots of variance explained by PRS in Europeans and Africans at 10 p_T values in the discovery sample are presented in Figure 1.

Table 1. Sample Characteristics of Case and Control Subjects

Characteristic	FEP Cases (GAP) (n = 445)	Chronic Cases (IMPACT) (n = 248)	Controls (GAP) (n = 265)	African Controls (SLESS) (n = 828)
Age, Years	27.7 ± 8.5	44 ± 10.1	30.2 ± 9.3	58.7 ± 12.0
Sex, Female	145 (32)	102 (39)	113 (43)	397 (48)
Buccal DNA	99 (22)	0	44 (17)	0
European Ancestry	171 (38)	137 (54)	168 (63)	0
African Ancestry	205 (46)	92 (36)	70 (27)	828 (100)
Other/Mixed Ethnicity	69 (16)	27 (10)	27 (10)	0

Values are mean ± SD or n (%).

FEP, first-episode psychosis; GAP, Genetics and Psychosis study; IMPACT, Improving Physical Health and Reducing Substance Use in Psychosis; SLESS, South London Ethnicity and Stroke Study.

In the GAP European subsample, case subjects had on average higher adjusted PRSs than control subjects, with SMD of 0.54, after correction for population stratification. The density distributions of PRSs were partially overlapping (Figure 2), demonstrating low discriminative power for FEP (area under the curve = 0.65). In European chronic case subjects from the IMPACT sample the optimal case-control discrimination by PRS was also achieved at $p_T < .1$, explaining 6% of the variance ($p = 10^{-4}$). We observed that case subjects with chronic psychosis had lower PRS than FEP case subjects, but the difference was not significant.

Discrimination Between Schizophrenia and Other Psychoses

In Figure 3 we present the SMDs of PRSs between case subjects divided into two diagnostic categories (those who had met criteria for schizophrenia with at least one of the four diagnostic approaches described previously and those who had not) and control subjects. Diagnosis with at least one of the four diagnostic approaches was available in 151 European and 177 African FEP case subjects and 132 European case subjects with chronic psychoses. We observed that both schizophrenia and other psychoses groups in all three samples had higher PRSs than their corresponding control subjects and that, in European samples, individuals with a diagnosis of schizophrenia had higher PRSs than other psychoses.

In the GAP FEP European subsample we repeated the case-control analysis stratifying case subjects by diagnosis and observed that PRS was more predictive for case subjects with schizophrenia ($R^2 = 16.3\%$, $p = 3.7 \times 10^{-7}$) than for case subjects with other psychoses ($R^2 = 2.7\%$, $p = .03$). In a case-only analysis, comparing those with at least one diagnosis of

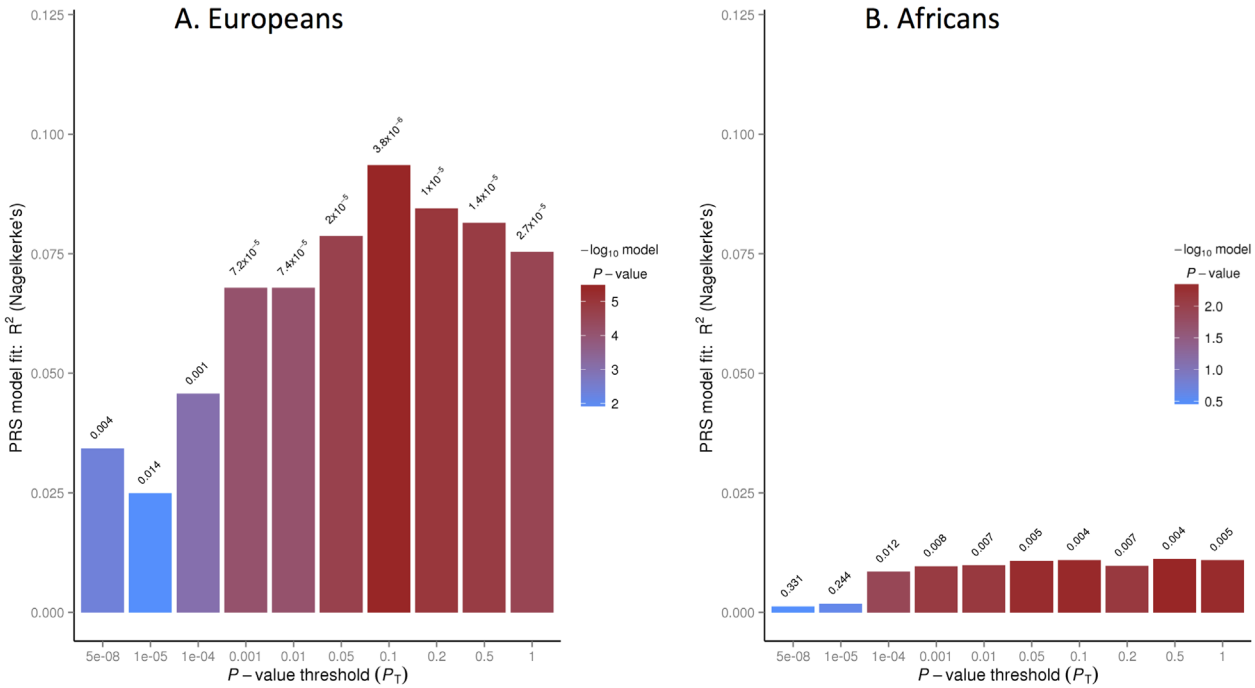

Figure 1. Proportion of variance of psychosis case-control status explained by polygenic risk score (PRS) in the European and African first-episode psychosis samples. The bars represent PRS calculated for 10 subsets of markers at different p value thresholds in the latest Psychiatric Genomics Consortium meta-analysis. The African sample includes case subjects from the Genetics and Psychosis study and combined control subjects from the Genetics and Psychosis study and South London Ethnicity and Stroke Study.

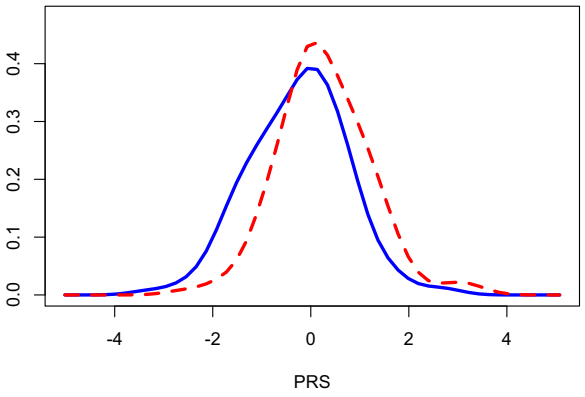

Figure 2. Density distribution of polygenic risk score (PRS) in European first-episode psychosis case and control subjects. PRS represents the standardized residuals of PRS after adjustment for the 10 principal components. Blue line indicates control subjects; red line, case subjects.

schizophrenia in any of the four diagnostic approaches (n = 86) versus those FEP case subjects who never met diagnostic criteria for schizophrenia (n = 65), we found that schizophrenia case subjects had higher PRS than case subjects with other psychoses (Nagelkerke's R^2 = 9.2%, p = .002).

In the IMPACT chronic psychosis sample, stratification by the recorded ICD-10 diagnosis to schizophrenia and other psychoses resulted, similar to the findings in FEP, in higher PRS for case subjects with a diagnosis of schizophrenia than case subjects with any other psychoses (R^2 = 6%, p = .008). In a case-control analysis, PRS discriminated control subjects from chronic case subjects with schizophrenia (R^2 = 10%, $p = 1.2 \times 10^{-5}$) but not from psychotic patients with other diagnoses (p = .2).

In the FEP African sample, we observed a different PRS prediction profile. For schizophrenia case subjects, the optimal discrimination from control subjects was achieved with the inclusion of fewer markers in the PRS ($p_T < .001$, R^2 = 2.4%, $p = 2 \times 10^{-4}$), while in other psychoses versus control subjects no p_T reached significance (lowest p value .13 at $p_T < .1$). A case-only analysis of Africans showed the best discrimination between schizophrenia and other psychosis in PRS constructed with fewer markers (optimal $p_T < 10^{-4}$, R^2 = 7.5%, p = .002), and the difference disappeared with the inclusion of more markers ($p_T >$ .05). Detailed presentation of stratified analyses by diagnosis in the three samples is shown in Supplemental Figures S5–S8

OR of FEP in Relation to PRS in Europeans

To represent visually the effect of PRS on the risk of psychosis, we ranked our European FEP subsample by PRS adjusted for the 10 PCs, we divided it in quintiles, and we measured case-control OR with the middle quintile (median PRS) as the baseline. With the increase of PRS we notice a gradient increase in the risk of psychosis, with the fifth quintile having higher OR and the first quintile having lower OR in comparison with the median. A similar profile was observed in the case-only comparison of schizophrenia with other psychoses (Figure 4). After simulation to a larger sample with the same characteristics and using the lowest PRS group as baseline, this translates to an OR of 7.7 comparing the top and bottom deciles (Supplemental Figure S9).

DISCUSSION

Polygenic Discrimination of Case-Control Status and Risk Prediction in FEP

The dual goals of schizophrenia genetics are to improve our understanding of the biology of the disorder with the hope of thereby developing novel therapeutic agents and to identify individuals at elevated risk before disease onset, aiming to expedite early diagnosis (28). A polygenic theory of schizophrenia

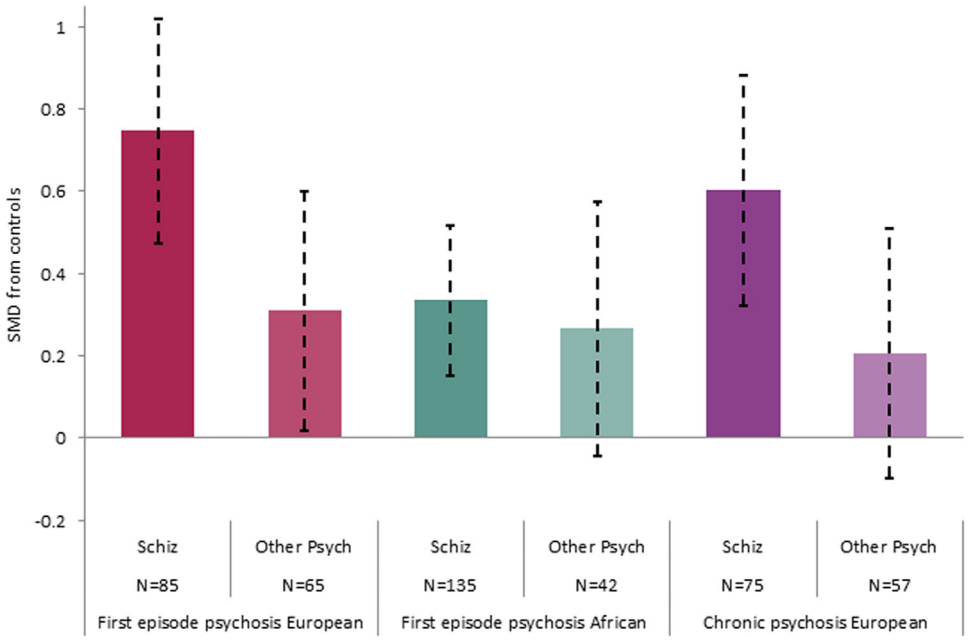

Figure 3. Standardized mean difference (SMD) of polygenic risk score (PRS) between first-episode psychosis European, first-episode psychosis African, and chronic psychosis European case subjects stratified according to diagnosis and corresponding control subjects. The vertical dashed error bars represent 95% confidence intervals. Psych, psychosis; Schiz, schizophrenia.

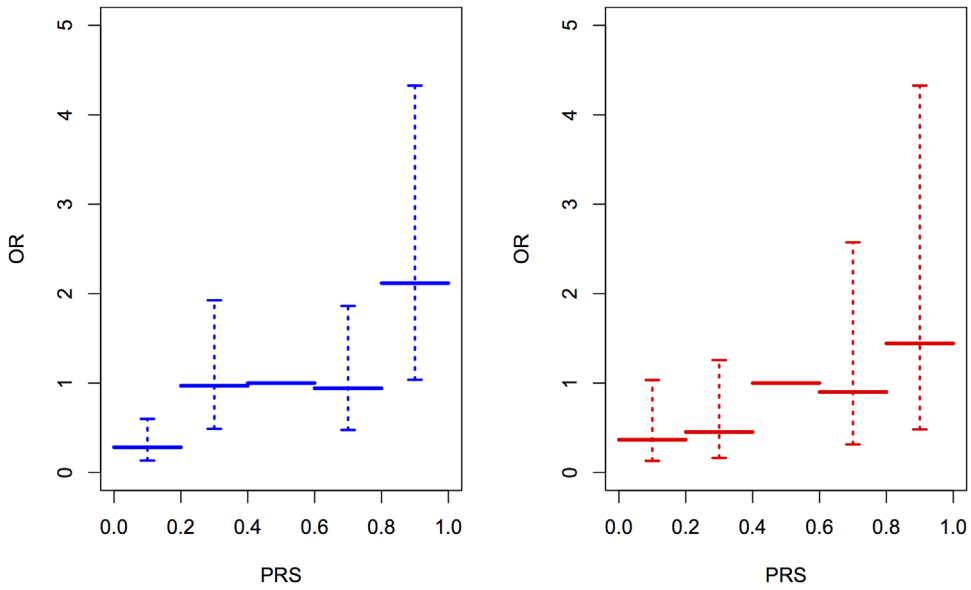

Figure 4. Odds ratios (ORs) of psychosis at different levels of polygenic risk score (PRS) in European-ancestry first-episode psychosis (FEP) case and control subjects. Subjects were ranked according to the PRS (adjusted for the 10 principal components) from lower to higher and divided into quintiles. The width of each bar on the x axis represents the proportion of individuals exposed at each level of risk, the y axis corresponds to the observed OR, and the vertical dotted lines represent 95% confidence intervals. Each quintile is compared with the median (baseline) group. On the left we represent the OR of PRS in an analysis of case subjects versus control subjects and on the right the OR of each group of case subjects only, in the comparison of schizophrenia with other psychoses.

was first hypothesized in 1967 (29) based on epidemiologic evidence of family studies that illustrated familiar clustering of the disease and concordance rates consistent with a multifactorial model rather than single gene causation, agreeing with the findings from PGC GWAs. Our study successfully replicates PRS prediction of case-control status in a typical clinical sample of European patients admitted with FEP to psychiatric services in South London but also examines the potential clinical utility of PRS within FEP.

At present, the discriminative accuracy of PRS for schizophrenia is not sufficient for use in population screening to identify individuals at high risk of the disease, and, indeed, PRS may never be powerful enough for screening (2,30). However, PRS explains a substantial amount of the variance of schizophrenia in Europeans, probably more than any traditional risk factor. In our samples, the OR for psychosis of the high- versus the low-exposure group is higher for PRS (OR = 7.7) than those reported for bullying victimization (OR = 2.36), vitamin D deficiency (OR = 3), and daily high-potency cannabis use (OR = 5.4) and is similar to social disadvantage 1 year before disease onset (OR = 9.85) (31–34). Given its strength but the lack of power for screening, the question arises as to what is the potential clinical utility PRS?

FEP provides an opportunity to test PRS against clinical outcomes beyond case status. In our sample, only 57% of FEP case subjects obtained a diagnosis of schizophrenia, while the remainder had bipolar disorder, psychotic depression, or acute or unspecified psychotic disorder. We found significant discrimination between schizophrenia and other psychoses within our FEP case subjects of European descent (9.2% of variance explained). Previous studies of the schizophrenia PGC1 PRS found it to be higher in patients with a broad spectrum of psychotic disorders versus control subjects (35,36), but discrimination between schizophrenia and other psychotic disorders was not or marginally significant. The improvement of discriminative accuracy seen in our sample is explained by the increased predictive power of PRS constructed from the newer PGC2 dataset. This supports the utility of further increasing the global schizophrenia sample size with genome-wide genotyping.

Note that the discriminative accuracy of PRS was higher for schizophrenia case subjects in our FEP cohort than for our group of (older) case subjects with chronic schizophrenia, identified from a clinical trial cohort. We cannot draw definite conclusions because the sample sizes are small; however, this finding is not unexpected. Patients with an average of 16 years of psychotic illness who maintain capacity to consent and participate in an intervention study are most probably not at the extreme end of the continuum in terms of illness severity. Furthermore, this observation supports the utility of PRS in early stages of the disease, because this design may be better at including patients with potentially more severe illness progression. Moreover, when we stratified chronic case subjects by diagnosis, we observed that, in the chronic group, PRS did not discriminate other (nonschizophrenia) psychoses from control subjects. It may be that if someone has maintained a nonschizophrenic psychotic disorder for 16 years, he or she is highly unlikely to develop schizophrenia later. However, these observations of PRS prediction in chronic versus FEP case subjects will require replication in independent samples.

Ethnic Differences

One of our striking findings was the substantial difference in PRS discrimination of case-control status between individuals from European and African ancestry. It is remarkable that PRS explains as much as 9.4% of the variance in Europeans, while it has low discriminative ability in Africans (1.1% of the variance explained in the joint analysis of GAP sample with SLESS control subjects). The PGC2 discovery data included 49 European, 3 Asian, and no African ancestry samples (2), while it is well established that Africans have higher genetic diversity and consequently shorter LD blocks (37), which is particularly relevant because most associated

SNPs are genetic markers in LD with the unrecognized causal variants (38).

Other causes for this difference should be considered. It is already suggested that schizophrenia is a constellation of different disorders (39,40). There is always the possibility that different genetic and biological pathways are involved. However, previous work shows that many schizophrenia risk alleles are shared across ethnic groups, predating African-European divergence (41,42). Another possibility is that there is a different environmental burden to the disease, especially because most African individuals are first- or second-generation migrants, a well-established environmental risk factor (43); hence, this group may have a lower genetic risk threshold for the development of the illness. One more important consideration is a different level of misdiagnosis between different ethnic groups. Strong evidence that individuals of African and African Caribbean origin have fourfold to sevenfold increase in the risk of psychosis in South London (44) may suggest that mild forms of psychosis are more readily identified in this group (45). All of these factors may confer lower discriminative power.

Limitations

The predictive ability of PRS depends, among other factors, on the number of markers in the genotyping panel (7), which affects the accuracy of the simulated SNP calls. The array we used provided an efficient genome-wide coverage for our European sample but not for the Africans as discussed previously, and genotyping with denser arrays may be necessary to measure the predictive power of PRS in non-European populations. Although our study was sufficient to analyze European and Africans, we did not have enough samples for separate analysis of the other ethnic groups, especially because we expected lower prediction of PRS, and insufficient evidence exists at present for combined analyses across ethnicities (46). In our total sample, PRS explained only 3.6% of the variance in case-control status prediction, capturing most probably the genetic signal mainly from the Europeans.

Conclusions

The continuing effort to increase the global sample size has resulted in the discriminative ability of the PRS improving and, as we have shown, the predictive power is sufficient to detect genetic signal in moderate-to-small samples. Genotyping of large samples from different ethnic groups would be important to be able to generalize the findings to non-European populations. Ideally, this should include samples collected at the country of origin, to avoid potential confounders, such as migration and other socioeconomic differences that may affect case-control status (43).

To explore the clinical importance of the association of PRS with psychosis and to better understand the biology of the genetic predisposition to psychosis by common variants, further research in the field should go beyond the association with case-control status. A nonexclusive list of important research questions include the association of PRS with 1) severity or course of illness, 2) conversion of individuals with prodromal symptoms to psychosis, 3) response to treatment, 4) endophenotypes and symptom profiles, and 5) physical health and quality of life.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by Guy's and St. Thomas Charity Grant No. R080529 (to EV and CML), The European Union FP7-People-2011-IAPP PsychDPC Grant No. GA 286213 (to CML), the Psychiatry Research Trust (to RMM), and the National Institute for Health Research Biomedical Research Centre at South London and Maudsley National Health Service Foundation Trust and King's College London (to GB).

The views expressed are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research, or the Department of Health.

The authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Medical Research Council, Social, Genetic & Developmental Psychiatry Centre (EV, MDF, JC, JE, PO, CC, HP, SN, KJA, CML, GB), Department of Psychosis Studies (MDF, CI, OA, TRM, PG-S, ZA, SS, FG, PD, ASD, RMM), Psychological Medicine (AK, VM, KI, CP), Basic and Clinical Neuroscience (JP), and Centre for Neuroimaging Sciences (DP), Institute of Psychiatry, Psychology & Neuroscience, King's College London; National Institute for Health Research, Mental Health Biomedical Research Centre (EV, CC, HP, SN, PD, GB), South London and Maudsley National Health Service Foundation Trust and King's College London; Department of Medical and Molecular Genetics (MT), King's College London, London, United Kingdom; School of Psychology (KEG), University of Sussex, Brighton and Sussex Partnership National Health Service Foundation Trust, West Sussex; South London and Maudsley National Health Service Foundation Trust (SS), London; and Department of Clinical Neurosciences (HSM), Neurology Unit, University of Cambridge, Cambridge, United Kingdom; Instituto de Medicina Molecular (DP), Faculdade de Medicina da Universidade de Lisboa, Lisboa, Portugal; and Department of Psychiatry (KJA), University of Alberta, Edmonton, Alberta, Canada.

Address correspondence to Evangelos Vassos, M.D., Institute of Psychiatry, Psychology & Neuroscience, MRC SGDP Centre, De Crespigny Park, London SE5 8AF, United Kingdom; E-mail: evangelos.vassos@kcl.ac.uk.

Received Oct 29, 2015; revised Jan 20, 2016; accepted Jan 22, 2016.

Supplementary material cited in this article is available online at http://dx.doi.org/10.1016/j.biopsych.2016.06.028.

REFERENCES

1. Gibson G (2012): Rare and common variants: Twenty arguments. Nat Rev Genet 13:135–145.
2. Psychiatric Genomics Consortium Schizophrenia Working Group (2014): Biological insights from 108 schizophrenia-associated genetic loci. Nature 511:421–427.
3. So HC, Kwan JS, Cherny SS, Sham PC (2011): Risk prediction of complex diseases from family history and known susceptibility loci, with applications for cancer screening. Am J Hum Genet 88:548–565.
4. Maier R, Moser G, Chen GB, Ripke S, Cross-Disorder Working Group of the Psychiatric Genomics ConsortiumCoryell W, *et al*. (2015): Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. Am J Hum Genet 96:283–294.
5. International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, *et al*. (2009): Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460:748–752.
6. Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park JH (2013): Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nat Genet 45:400–405, 405e1–3.
7. Dudbridge F (2013): Power and predictive accuracy of polygenic risk scores. PLoS Genet 9:e1003348.
8. Janssens AC, Ioannidis JP, Bedrosian S, Boffetta P, Dolan SM, Dowling N, *et al*. (2011): Strengthening the reporting of genetic risk

prediction studies (GRIPS): Explanation and elaboration. Eur J Hum Genet 19:18 p preceding 494.
9. Di Forti M, Iyegbe C, Sallis H, Kolliakou A, Falcone MA, Paparelli A, *et al*. (2012): Confirmation that the AKT1 (rs2494732) genotype influences the risk of psychosis in cannabis users. Biol Psychiatry 72: 811–816.
10. World Health Organization (1992): The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. Geneva: WHO.
11. World Health Organization, Division of Mental Health (1999): Schedules for Clinical Assessment in Neuropsychiatry (Version 2.1), 2.1. ed. Geneva: WHO Assessment, Classification and Epidemiology.
12. Bebbington P, Nayani T (1995): The Psychosis Screening Questionnaire. Int J Method Psych 5:11–19.
13. Sheitman BB, Lee H, Strous R, Lieberman JA (1997): The evaluation and treatment of first-episode psychosis. Schizophr Bull 23:653–661.
14. McGuffin P, Farmer A, Harvey I (1991): A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. Arch Gen Psychiatry 48: 764–770.
15. Gaughran F, Stahl D, Ismail K, Atakan Z, Lally J, Gardner-Sood P, *et al*. (2013): Improving physical health and reducing substance use in psychosis–Randomised control trial (IMPACT RCT): Study protocol for a cluster randomised controlled trial. BMC Psychiatry 13:263.
16. Gardner-Sood P, Lally J, Smith S, Atakan Z, Ismail K, Greenwood KE, *et al*. (2015): Cardiovascular risk factors and metabolic syndrome in people with established psychotic illnesses: Baseline data from the IMPaCT randomized controlled trial. Psychol Med 45:2619–2629.
17. Khan U, Crossley C, Kalra L, Rudd A, Wolfe CD, Collinson P, *et al*. (2008): Homocysteine and its relationship to stroke subtypes in a UK black population: The south London ethnicity and stroke study. Stroke 39:2943–2949.
18. Howie BN, Donnelly P, Marchini J (2009): A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. Plos Genet 5:e1000529.
19. 1000 Genomes Project Corsortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, *et al*. (2012): An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65.
20. Howie B, Marchini J, Stephens M (2011): Genotype imputation with thousands of genomes. G3 (Bethesda) 1:457–470.
21. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015): Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience 4:7.
22. Psychosis Endophenotypes International Consortium, Wellcome Trust Case-Control Consortium 2, Bramon E, Pirinen M, Strange A, Lin K, *et al*. (2014): A genome-wide association analysis of a broad psychosis phenotype identifies three loci for further investigation. Biol Psychiatry 75:386–397.
23. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006): Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909.
24. Euesden J, Lewis CM, O'Reilly PF (2015): PRSice: Polygenic Risk Score software. Bioinformatics 31:1466–1468.
25. Saha S, Chant D, Welham J, McGrath J (2005): A systematic review of the prevalence of schizophrenia. PLoS Med 2:e141.
26. Agerbo E, Sullivan PF, Vilhjalmsson BJ, Pedersen CB, Mors O, Borglum AD, *et al*. (2015): Polygenic risk score, parental socioeconomic status, family history of psychiatric disorders, and the risk for schizophrenia: A Danish population-based study and meta-analysis. JAMA Psychiatry 72:635–641.
27. Morgan C, Dazzan P, Morgan K, Jones P, Harrison G, Leff J, *et al*. (2006): First episode psychosis and ethnicity: Initial findings from the AESOP study. World Psychiatry 5:40–46.
28. Mowry BJ, Gratten J (2013): The emerging spectrum of allelic variation in schizophrenia: Current evidence and strategies for the identification and functional characterization of common and rare variants. Mol Psychiatry 18:38–52.
29. Gottesman II, Shields J (1967): A polygenic theory of schizophrenia. Proc Natl Acad Sci U S A 58:199–205.
30. Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM (2006): Predictive testing for complex diseases using multiple genes: Fact or fiction? Genet Med 8:395–400.
31. Trotta A, Di Forti M, Mondelli V, Dazzan P, Pariante C, David A, *et al*. (2013): Prevalence of bullying victimisation amongst first-episode psychosis patients and unaffected controls. Schizophr Res 150: 169–175.
32. Crews M, Lally J, Gardner-Sood P, Howes O, Bonaccorso S, Smith S, *et al*. (2013): Vitamin D deficiency in first episode psychosis: A case-control study. Schizophr Res 150:533–537.
33. Di Forti M, Marconi A, Carra E, Fraietta S, Trotta A, Bonomo M, *et al*. (2015): Proportion of patients in south London with first-episode psychosis attributable to use of high potency cannabis: A case-control study. Lancet Psychiat 2:233–238.
34. Stilo SA, Di Forti M, Mondelli V, Falcone AM, Russo M, O'Connor J, *et al*. (2013): Social disadvantage: Cause or consequence of impending psychosis? Schizophrenia Bull 39:1288–1295.
35. Tesli M, Espeseth T, Bettella F, Mattingsdal M, Aas M, Melle I, *et al*. (2014): Polygenic risk score and the psychosis continuum model. Acta Psychiatr Scand 130:311–317.
36. Bigdeli TB, Bacanu SA, Webb BT, Walsh D, O'Neill FA, Fanous AH, *et al*. (2014): Molecular validation of the schizophrenia spectrum. Schizophr Bull 40:60–65.
37. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, *et al*. (2009): The genetic structure and history of Africans and African Americans. Science 324:1035–1044.
38. Balding DJ (2006): A tutorial on statistical methods for population association studies. Nat Rev Genet 7:781–791.
39. Kendler KS, Karkowski-Shuman L, O'Neill FA, Straub RE, MacLean CJ, Walsh D (1997): Resemblance of psychotic symptoms and syndromes in affected sibling pairs from the Irish Study of High-Density Schizophrenia Families: Evidence for possible etiologic heterogeneity. Am J Psychiatry 154:191–198.
40. Tandon R, Keshavan MS, Nasrallah HA (2008): Schizophrenia, "just the facts" what we know in 2008. 2. Epidemiology and etiology. Schizophr Res 102:1–18.
41. Ioannidis JP, Ntzani EE, Trikalinos TA (2004): 'Racial' differences in genetic effects for complex diseases. Nat Genet 36: 1312–1318.
42. de Candia TR, Lee SH, Yang J, Browning BL, Gejman PV, Levinson DF, *et al*. (2013): Additive genetic variation in schizophrenia risk is shared by populations of African and European descent. Am J Hum Genet 93:463–470.
43. Cooper B (2005): Immigration and schizophrenia: The social causation hypothesis revisited. Br J Psychiatry 186:361–363.
44. Fearon P, Kirkbride JB, Morgan C, Dazzan P, Morgan K, Lloyd T, *et al*. (2006): Incidence of schizophrenia and other psychoses in ethnic minority groups: Results from the MRC AESOP Study. Psychol Med 36:1541–1550.
45. Trierweiler SJ, Neighbors HW, Munday C, Thompson EE, Binion VJ, Gomez JP (2000): Clinician attributions associated with the diagnosis of schizophrenia in African American and non-African American patients. J Consult Clin Psychol 68:171–175.
46. Medina-Gomez C, Felix JF, Estrada K, Peters MJ, Herrera L, Kruithof CJ, *et al*. (2015): Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: The Generation R Study. Eur J Epidemiol 30:317–330.