

Genome-wide Association Study of Clinical Features in the Schizophrenia

Psychiatric Genomics Consortium: Confirmation of Polygenic Effect on Negative Symptoms

Tim B. Bigdeli, Ph.D.^{1,2,*}, Roseann E. Peterson, PhD², Stephan Ripke, M.D., Ph.D.^{3,4}, Silviu-Alin Bacanu, Ph.D.², Richard L. Amdur, Ph.D.³⁸, Pablo V. Gejman, M.D.^{5,6}, Douglas F. Levinson, M.D.⁷, Brien P. Riley, Ph.D.^{2,8}, David St. Clair, M.D., Ph.D.⁹, Marcella Rietschel, M.D., Ph.D.¹⁰, James T.R. Walters, M.D., Ph.D.¹¹, Roel A. Ophoff, Ph.D.¹²⁻¹⁴, Andrew McQuillin, Ph.D.¹⁵, Hugh Gurling, M.D, FRCPsych¹⁵, Dan Rujescu, M.D., Ph.D.^{16,17}, Patrick F. Sullivan, M.D., FRANZCP¹⁸⁻²⁰, George Kirov, M.D., Ph.D.¹¹, Michele T. Pato, M.D.^{1,21}, Carlos N. Pato, M.D., Ph.D.^{1,21}, Ole A. Andreassen, M.D., Ph.D.^{22,23}, Michael J. Owen, M.D., Ph.D.^{11,24}, Michael C. O'Donovan, M.D., Ph.D.^{11,24}, Aiden Corvin, M.D., Ph.D.²⁵, Anil K Malhotra, M.D.²⁶⁻²⁸, Bryan J. Mowry, M.D., FRANZCP^{29,30}, Tõnu Esko, Ph.D.³¹⁻³⁴, Thomas Werge, Ph.D.³⁵⁻³⁷, Kenneth S. Kendler, M.D.^{2,8}, Schizophrenia Working Group of the Psychiatric Genomics Consortium^a, Ayman H. Fanous, M.D.^{1,2,38-40,*}

¹Department of Psychiatry and Behavioral Sciences, State University of New York Downstate Medical Center, Brooklyn, New York, USA

²Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University School of Medicine, Richmond, VA, USA

³Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

⁴Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

⁵Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, IL, USA

⁶Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, USA

⁷Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA

⁸Department Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA, USA

⁹University of Aberdeen, Institute of Medical Sciences, Aberdeen, Scotland, UK

¹⁰Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Heidelberg, Germany

¹¹MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK

¹²Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA

¹³Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

¹⁴Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, The Netherlands

¹⁵Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, UK

- ¹⁶Department of Psychiatry, University of Halle, Halle, Germany
- ¹⁷Department of Psychiatry, University of Munich, Munich, Germany
- ¹⁸Department of Genetics, University of North Carolina, Chapel Hill, NC, USA
- ¹⁹Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA
- ²⁰Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
- ²¹Department of Psychiatry and Zilkha Neurogenetics Institute, Keck School of Medicine at University of Southern California, Los Angeles, CA, USA
- ²²NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway
- ²³Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway
- ²⁴National Centre for Mental Health, Cardiff University, Cardiff, Wales
- ²⁵Neuropsychiatric Genetics Research Group, Department of Psychiatry, Trinity College Dublin, Ireland
- ²⁶The Zucker Hillside Hospital, Glen Oaks, NY, USA
- ²⁷The Feinstein Institute for Medical Research, Manhasset, NY, USA
- ²⁸The Hofstra NS-LIJ School of Medicine, Hempstead, NY, USA
- ²⁹Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia
- ³⁰Queensland Centre for Mental Health Research, University of Queensland, Brisbane, Queensland, Australia
- ³¹Estonian Genome Center, University of Tartu, Tartu, Estonia
- ³²Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA
- ³³Department of Genetics, Harvard Medical School, Boston, MA, USA
- ³⁴Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA
- ³⁵Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Denmark
- ³⁶Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
- ³⁷The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark
- ³⁸Mental Health Service Line, Washington VA Medical Center, Washington DC, USA
- ³⁹Department of Psychiatry, Georgetown University School of Medicine, Washington DC, USA
- ⁴⁰Department of Psychiatry, Keck School of Medicine at University of Southern California, Los Angeles, CA, USA

^aFull list of consortium members provided in the supplemental information.

*Corresponding author

SUNY Downstate Medical Center

450 Clarkson Avenue

Psychiatry - MSC 1203

Brooklyn, NY 11203

Phone No: (718) 270-2077

Fax No: (718) 270-4617

Email: ayman.fanous@downstate.edu; tim.bigdeli@downstate.edu

Abstract

Schizophrenia is a clinically heterogeneous disorder. Proposed revisions in DSM-5 included dimensional measurement of different symptom domains. We sought to identify common genetic variants influencing these dimensions, and confirm a previous association between polygenic risk of schizophrenia and the severity of negative symptoms. The Psychiatric Genomics Consortium study of schizophrenia comprised 8,432 cases of European ancestry with available clinical phenotype data. Symptoms averaged over the course of illness were assessed using the OPCRIT, PANSS, LDPS, SCAN, SCID, and CASH. Factor analyses of each constituent PGC study identified positive, negative, manic, and depressive symptom dimensions. We examined the relationship between the resultant symptom dimensions and aggregate polygenic risk scores indexing risk of schizophrenia. We performed genome-wide association study (GWAS) of each quantitative traits using linear regression and adjusting for significant effects of sex and ancestry. The negative symptom factor was significantly associated with polygene risk scores for schizophrenia, confirming a previous, suggestive finding by our group in a smaller sample, though explaining only a small fraction of the variance. In subsequent GWAS, we observed the strongest evidence of association for the positive and negative symptom factors, with SNPs in *RFX8* on 2q11.2 ($P=6.27\times 10^{-8}$) and upstream of *WDR72/UNC13C* on 15q21.3 ($P=7.59\times 10^{-8}$), respectively. We report evidence of association of novel modifier loci for schizophrenia, though no single locus attained established genome-wide significance criteria. As this may have been due to insufficient statistical power, follow-up in additional samples is warranted. Importantly, we replicated our previous finding that polygenic risk explains at least some of the variance in negative symptoms, a core illness dimension.

Key Words: genetics, GWAS, symptom dimension, factor analysis, modifier

Introduction

The symptomatic presentation, course, outcome, and severity of schizophrenia (SZ) have long been known to vary widely from patient to patient. DSM and ICD subtypes have their historical antecedents in the historical rubrics paranoia, hebephrenia, and catatonia, all of which predate the modern concept of SZ¹. Whether such putative, categorical subtypes best describe variation in clinical features is unclear, as many patients meet criteria for the undifferentiated subtype². Dimensional constructs such as positive, negative, disorganization, and mood symptoms have been proposed using both theoretical and empirical criteria³, and evidence from family and twin studies suggests that these are heritable⁴. Of major interest to researchers and clinicians alike is the potential of symptomatic dimensions for predicting outcome, severity, and treatment response.

We have previously suggested that genes could affect clinical variation in at least two ways⁴. Modifier genes might influence the clinical character of the illness without increasing disease susceptibility. The most elegant examples of modifier genes *per se* are found in studies of cystic fibrosis. Though caused by mutations in the *CFTR* gene, the associated clinical features of cystic fibrosis, some of which influence mortality, are influenced by genes having no contribution to the risk of illness⁵. Genetic modifiers of disease severity have been identified for a range of diseases, from facioscapulohumeral muscular dystrophy to Alzheimer's disease^{6,7}. Second, allelic variation in disease susceptibility genes themselves might influence clinical characteristics, or predispose to a particular illness subtype. Such subtype-specific loci, or "susceptibility-modifier genes", have been identified in several disorders, including Huntington's disease, breast cancer and frontotemporal lobar degeneration⁸⁻¹¹.

Identification of both classes of genes in SZ could be beneficial in several ways. First, this might yield clues to underlying pathophysiological processes more or less specific to a clinical subtype or symptom dimension. This has the potential to inform questions of diagnosis and classification such as the use of dimensional approaches, as was debated in the development of DSM-5¹² but ultimately not adopted. Intriguingly, the demonstration of pleiotropic effects—both of specific loci and in the form of shared polygenetic liability—may hold additional clues. A recent study found that a polygenic risk score for bipolar disorder significantly predicted the severity of manic symptoms in SZ¹³. Also noteworthy are reported copy-number variant (CNV) associations shared between SZ, autism and intellectual disability¹⁴,

¹⁵, which suggests a link between negative symptom-like psychopathology and the potentially debilitating effects of gene-disrupting CNVs ^{16, 17}. Importantly, such insights could provide leads in the development of novel treatments, since it is symptom dimensions, rather than the illness itself, that are ameliorated by medications.

We previously published a large genome-wide association study of positive, negative, and mood symptoms in a sample of 2,454 cases, which provided suggestive evidence of novel modifier genes influencing these dimensions ¹⁸. Furthermore, the polygenic risk of SZ predicted negative/disorganized symptom severity. In this report, we have used a similar approach in over 8,000 cases from the Psychiatric Genomics Consortium (PGC) study of SZ ¹⁹. Firstly, we sought to replicate our previous observation that polygenic risk of SZ influences the severity of negative symptoms. We also hypothesized that, with enhanced statistical power, we might detect additional evidence of modifier loci influencing classic clinical dimensions of psychotic illness.

Methods

Ascertainment and Assessment

The subsamples included in this study comprise 19 constituent sites from Stage 2 of the Psychiatric Genomics Consortium study of SZ. Ascertainment, diagnostic assessment, genotyping, and genotype quality control have been previously described ^{19, 20}. Briefly, 52 samples from the US, Europe, and Australia comprising 34,241 cases, 45,604 controls and 1,235 parent affected-offspring trios were genotyped using a number of commercial single nucleotide polymorphism (SNP) genotyping platforms. These data were processed using the stringent PGC quality control procedures, followed by imputation of SNPs and insertion-deletions using the 1000 Genomes Project reference panel (UCSC hg19/NCBI 37)²¹ using IMPUTE2 ²², resulting in nearly 9.5M markers for GWAS analysis.

In the current study, we focus on clinical dimensions of illness as the phenotypes of interest, and have performed a case-only GWAS of these quantitative traits. Clinical dimensions were derived directly from the individual items of symptom checklists or rating scales assessed in the individual PGC sites if they were available. These instruments included the Comprehensive Assessment of Symptoms and History

(CASH)²³, Lifetime Dimensions of SZ Scale (LDPS)²⁴, Operational Criteria Checklist for Psychotic Illness (OPCRIT)²⁵, Positive and Negative Syndrome Scale (PANSS)²⁶. For some subsamples, we used the individual items of the diagnostic interviews, including the Schedule for Clinical Assessment in Neuropsychiatry (SCAN)²⁷ and Structured Clinical Interview for DSM-III-R (SCID)²⁸. The individual sites, their sample sizes, and the instruments used for developing quantitative phenotypes are presented in Supplemental Table 1.

Quantitative phenotypes

Quantitative traits for the Stage 1 sites were derived from the individual items of assessment instruments as previously described¹³. In brief, exploratory factor analysis of each individual site's data, followed by confirmatory factor analysis, identified four factors that were most prominent across sites: Positive, Negative, Manic, and Depressive symptoms. We then harmonized these factors across sites and instruments, then Z-transformed factor scores within each site. For the current expansion into the Stage 2 sites, we opted to retain the four-factor model. We selected the relevant items comprising these factors from the instruments used in the Stage 2 sites, and used them to create factors scores, which were subsequently Z-transformed. A more detailed description of these methods is given in the accompanying Supplemental Note.

Polygenic scoring analyses

To test for a polygenic effect on symptom dimensions, we performed risk score profiling as previously described^{19, 29}. Briefly, SZ risk scores were generated for each study site in the PGC2 study of SZ¹⁹, using every other study as the training set in an iterative, "leave-one-out" procedure. This approach resolved any issues related to significance testing of a score among subjects included in the training set, while offering improved power and less sample attrition compared to subdividing the full cohort into approximate halves. We computed several scores based on varying *P*-value threshold signifying the proportion of SNPs with smaller *P*-values in the training set; *P*-value thresholds ranged between 0.0001 and 1.0. We tested for association between symptom factors and SZ polygenic scores by linear regression, adjusting for study-site and all covariates used in the primary association analyses. To investigate possible

heterogeneity among study sites, we also analyzed each site separately and combined site-specific results by fixed-effects meta-analysis.

Genome-wide association and meta-analysis

We tested for association between SNPs and each trait by linear regression, as implemented in PLINK v1.07 (<http://pngu.mgh.harvard.edu/~purcell/plink/>)³⁰, using allelic dosages and adjusting for significant covariates including sex and ancestry principal components (where applicable). For each trait, relevant covariates were identified for the full cohort by backwards-stepwise regression, after adjusting for study site. We performed GWAS of each trait separately for individual study sites, combining summary statistics in subsequent random-effects meta-analyses using METAL³¹. We excluded all SNPs with minor allele frequency (MAF) less than 0.01, average statistical imputation information (INFO) less than 0.5, or absent from more than half of total number of sub-studies.

For X chromosome SNPs, tests of association were carried out separately for male and female subjects. For each study site, we meta-analyzed these male- and female-specific results, subsequently combining the resultant summary statistics across study sites in a second round of meta-analysis.

Gene-set enrichment analysis

We applied a new method, Data-Driven Expression Prioritized Integration for Complex Traits (DEPICT)³², to identify significantly enriched gene-sets/pathways and tissues/cell-types. Briefly, genes harboring or in the vicinity of associated SNPs are tested for enrichment for “reconstituted” gene sets, comprised of curated sets expanded to include coregulated loci. Tissue and cell type enrichment analysis is conducted by testing whether genes were highly expressed in any of 209 MeSH annotations based on microarray data for the Affymetrix U133 Plus 2.0 Array platform. As input, we used SNPs significant at the recommended *P*-value inclusion thresholds of 5×10^{-8} and 10^{-5} ; we relaxed the latter threshold to 5×10^{-5} to allow for the inclusion of a greater number of associated loci.

Results

Exploratory factor analysis

As previously described, we used a combination of scree plots and clinical judgment to select a four-factor model consisting of positive, negative, manic, and depressive symptoms¹³. We observed that these four symptom dimensions were the most salient in the EFAs of the seven OPCRIT sites. EFA of the PANSS in three sites (Oslo, Munich, and CATIE) revealed separate negative and disorganized factors, as did a published EFA of the CASH in the one site using that instrument in this study. In our previous EFA of the LDPS in the MGS sites, there was a single negative symptom factor which also included disorganization symptoms. We therefore used a single negative factor, and combined negative and disorganization symptoms into one factor in the PANSS and CASH sites.

We observed separate manic and depressive factors in the OPCRIT, CASH and PANSS sites, but not the LDPS sites. We therefore opted to divide the LDPS mood factor into separate manic and depressive factors. As different instruments have differences in emphasis and granularity, we sought to test whether they tapped into underlying constructs that were more or less the same. In the Dublin sample, the Pearson product-moment correlations for the negative, manic, and depressive factors across the PANSS and OPCRIT were, respectively, 0.85, 0.70, and 0.48 (Supplementary table S7).

Polygenic scoring analyses

Following replicated evidence for a polygenic contribution to SZ²⁹, we assessed whether weighted polygenic risk scores based on a training-set of cases and controls¹⁹ could predict each symptom factor in an independent target dataset, namely each constituent dataset included herein. Given the requirement that training and target datasets be independent, scores were generated via a “leave-one-out” procedure (see Methods). Results for scores based on varying *P*-value thresholds are displayed in Figure 1a. Overall, we observe the strongest relationship with polygenic risk for the negative symptom factor. We observed the most significant such enrichment for a score constructed of SNPs with *P*-values less than 0.1 in the PGC2 GWAS of SZ ($R^2 = 0.00386$; $P = 6.39 \times 10^{-9}$). We observed little or no evidence of association with polygenic risk for the manic ($R^2 = 6.39 \times 10^{-4}$; $P = 0.0204$), depressive ($R^2 = 3.84 \times 10^{-4}$; $P = 0.0546$), and positive symptom factors ($R^2 = 6.46 \times 10^{-5}$; $P = 0.214$).

Given a previously observed association between SZ risk and negative symptom severity in the MGS study¹⁸, we sought to assess to what extent this particular study might be driving the observed pattern of

findings. In Table 1, we report the significance and predictive value of the SZ polygenic score as applied to each symptom dimension in the full PGC sample, the MGS study alone, and excluding MGS from the combined PGC sample. For negative symptoms, exclusion of the MGS study yielded nearly identical results with respect to variance explained ($R^2 = 0.0037$), though with an attendant diminution of statistical significance resulting from the smaller sample size ($P = 1.83 \times 10^{-6}$). Examination of the other symptom dimensions reveals a lesser degree of consistency across cohorts.

We followed-up the possibility of heterogeneity across study sites by individually analyzing each based on the overall best P -value threshold and subsequently combining study-specific results by fixed-effects meta-analysis. For negative symptoms, a forest plot displaying estimated effects for each site is given in Figure 1b; we did not observe significant heterogeneity in effects sizes across sites (Cochran's $P = 0.0955$). Corresponding forest plots for depressive, manic, and positive symptoms are included in the Supplemental material; of these, only positive symptoms demonstrated significant heterogeneity across study sites (Cochran's $P = 0.0003$).

Genome-wide association and meta-analysis

Genomic inflation factors (λ) were 0.993, 1.014, 1.0, 1.003 for the positive, negative, manic, and depressive factors, respectively. We observed the strongest evidence of association overall for the positive symptom factor on 2q11.2 at SNPs in *RFX8* ($P = 6.27 \times 10^{-8}$), followed by the negative symptom factor on 15q21.3, with SNPs upstream of *WDR72* ($P = 7.59 \times 10^{-8}$). Regional association and forest plots for these two SNPs are provided in the accompanying supplemental information. Although no single SNP surpassed established genome-wide significance criteria (i.e., 5×10^{-8}), we observed several moderate associations ($P < 10^{-6}$) at distinct chromosomal loci (Figure 2); 5, 4, and 3 for the positive, negative, and depressive symptom factors, respectively. These are summarized, with respect to the two strongest single SNP associations, in Table 2.

Gene-set enrichment analyses

We investigated whether particular gene-sets were enriched for associations with negative symptoms using DEPICT. For SNPs significant at $P < 10^{-5}$, representing 29 independent loci, a single

pathway, *response to drug* (GO:0042493), was significantly enriched (FDR < 0.2), or contained more significant genes than expected by chance, after correction for multiple testing. This gene-ontology (GO) category is defined as any process that results in a change of state or activity of a cell or an organism in response to a drug stimulus, and includes both *response to antipsychotic drug* (GO:0097332) and *response to antidepressant* (GO:0036276) as child terms (14).

We also considered a more inclusive threshold of $P < 5 \times 10^{-5}$, which yielded a total of 109 independent loci. A total of seven (7) gene-sets were found to be significantly enriched following correction for multiple testing. Of particular note is the observed enrichment of protein-protein-interaction (PPI) networks of low-density lipoprotein-related proteins 1 and 2 (FDR < 0.01), and ankyrin-2 (FDR < 0.05).

For each SNP P -value threshold considered, Table 3 gives the gene-sets that remained significant after correction for multiple testing (FDR < 0.20).

Discussion

To our knowledge, this is the largest GWAS of SZ-related quantitative traits performed to date. Of particular interest is our finding of a significant correlation between polygenic risk of SZ and severity of negative symptoms. This replicates, in a much larger multi-national sample, a previous observation of this relationship in the MGS sample¹⁸, and to a greater degree of statistical significance. More broadly, this finding provides additional support for genetic factors underlying the observed clinical heterogeneity in SZ⁴, and that at least some risk genes for SZ are in fact susceptibility-modifier genes. Some of these genes and their gene products might therefore provide clues to developing more effective treatments for this often-disabling symptom dimension.

Experiment-wide, the strongest evidence of single-marker association was observed for positive symptoms, at SNPs in *RFX Family Member 8, Lacking RFX DNA Binding Domain (RFX8)*, which encodes a potential transcription factor but which has not previously been implicated in SZ. An association between intergenic SNPs at 1p13.3 and negative symptoms also approached, albeit without attaining, genome-wide significance ($P < 10^{-7}$). These SNPs fall between, and upstream of *WD repeat domain 72 (WDR72)* and *unc-13 homolog C (UNC13C)*. In particular, *UNC13C* encodes a protein essential to presynaptic transmission

and synaptic plasticity^{33, 34}, and has been implicated in several neuropsychiatric traits, including post-traumatic stress disorder^{35, 36}.

At more modest significance thresholds ($P < 10^{-6}$), we observed several associations between the positive, negative, and depressive symptoms factors and SNPs mapping to both genes and intergenic regions, which vary with respect to the degrees of support for a role in SZ etiology. Notably, these SNPs were not significantly associated with SZ risk ($P > 0.05$) in the full PGC study¹⁹. While our single SNP results could be due to a paucity of true allelic effects on these clinical dimensions of illness, in our view, this is more likely due to insufficient power to detect small allelic effects in even a large sample such as this. This is suggested by the limited success in identifying replicable genome-wide significant effects in sample sizes comparable to ours or larger, in studies of heritable quantitative traits such as personality^{37, 38}. The GWAS literature in general suggests that in some complex traits, very large studies and considerable phenotypic refinement are required to detect small effect genes^{39, 40}.

That exploratory gene-set enrichment analyses highlighted several etiologically relevant molecular networks would also seem to support this interpretation. The observed enrichment of LRP1 and ANK2 protein-protein interaction networks for associations with negative symptoms is particularly salient, given the implication of the former in Alzheimer's disease⁴¹ and, more recently, of the *ANK2* gene in autism^{42, 43}, which have features of the cognitive and social aspects of negative symptoms.

Perhaps most convincing, however, is the highly significant association between negative symptom severity and polygenic risk of SZ, suggesting that many SZ risk loci also influence its clinical presentation. That SZ risk was only found to predict negative symptom severity may reflect the temporal stability of negative symptoms and their more accurate assessment by interviewers, thus contributing to a higher signal-to-noise ratio compared to the other dimensions^{1, 44, 45}. However, the finding that polygenic risk of bipolar disorder significantly predicts the less reliably assessed manic symptom factor in SZ¹³ suggests that this not the case. It is possible that negative symptoms are a marker for cases of SZ that arise more directly from common variants or the rare variants that are in LD with them, while phenocopies are more likely to be related to environmental etiologies, such as substance misuse. This is supported by the observation that cocaine- and amphetamine-induced psychotic states tend to present with more positive than negative symptoms⁴⁶. The differential impact of the polygenic risks of bipolar disorder (Ruderfer et

al., 2014) versus SZ (as reported here) suggests potential biological underpinnings of both the commonalities and differences between these two disorders, whose relationship has been the subject of considerable debate⁴⁷⁻⁴⁹. For example, bipolar disorder polygenes might explain some of their observed phenomenological overlap, in the form of the manic-like symptoms seen in both, while SZ polygenes might explain phenomena more specific to SZ, such as negative symptoms.

The heterogeneity of the assessment instruments could have increased the false negative rate by decreasing the signal-to-noise ratio in GWAS. The instruments used in this study differ in purpose, content, and granularity. The SCID²⁸ and SCAN²⁷ are structured interviews designed to make diagnoses rather than comprehensively inventory signs and symptoms. The OPCRIT²⁵, while it is a comprehensive checklist of classic signs and symptoms of SZ and affective disorders, is also meant to aid in diagnosis using multiple criteria. The PANSS²⁶, on the other hand, is not intended for diagnosis, but rather, for a fine-grained assessment of positive and negative symptoms of SZ, for purposes such as determining treatment efficacy. Unlike diagnostic interviews, it is therefore lacking in items reflecting classic manic symptoms, and instead, more sensitively assesses psychotic excitement and agitation. Similarly, PANSS depressive items do not comprehensively include classic neurovegetative symptoms, perhaps resulting in the lower PANSS-OPCRIT correlation we observed for the negative factor compared to others.

Because the PANSS is designed to assess common positive and negative symptoms of SZ, PANSS factors tended to be more normally distributed. The OPCRIT's depressive and especially, manic factors demonstrated considerable positive skew, in part because of the relatively low prevalence of classic manic symptoms, as compared to excitement and hostility, in SZ. Although the "manic" constructs in the OPCRIT and PANSS are therefore different, it was reassuring that they were strongly correlated in the Dublin sample, suggesting that they index the same underlying construct. Nevertheless, their divergent distributions and item content could have resulted in false negative results when analyzed together.

Related to the heterogeneity of the instruments used is the heterogeneity in factor solutions across sites. In order to maximize statistical power, we opted to use the most prominent and compelling set of clinical dimensions across sites, in all of the sites. This resulted in our using the less than best-fitting models in some of the sites, which could have reduced power in these sites, and concomitantly, in the study as a whole. Our combining of two factors into one in some of the sites (PANSS, SCAN, SCID)

might have diluted a putative genetic signal affecting one of these factors alone, but not the other one. These limitations highlight the need for standardized ratings in future quantitative measures of symptoms.

An additional limitation of this study was the lack of inclusion of potentially important covariates, including environmental factors such as cannabis use and urbanicity, as well as cognitive impairment. The latter in particular might be related to negative and disorganization symptoms. However, environmental factors occurring prior to the onset of illness are difficult to reliably determine retrospectively, and were therefore largely unavailable in these samples.

An important next step will be to attempt replication of these findings in additional independent datasets, including newer additions to the Psychiatric Genomics Consortium. This could include samples of patients with bipolar disorder and/or major depressive disorder, who sometimes manifest the some of the same symptoms seen in SZ. Furthermore, as large numbers of efficacy studies have used the PANSS, it might be possible to expand future studies such as ours with more routine biobanking and genotyping in such studies.

Acknowledgements

The work specific to this report was funded by the United States Department of Veterans Affairs Merit Review Program (5I01CX000278) to A.H.F. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 279227 (M.R.).

Statistical analyses were carried out on the Genetic Cluster Computer (<http://www.geneticcluster.org>) which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation and the VU University Amsterdam.

Conflict Of Interest

The authors declare no conflict of interest.

References

1. Andreasen NC, Carpenter WT, Jr. Diagnosis and classification of schizophrenia. *Schizophr Bull* 1993;19(2):199-214.
2. McGlashan TH, Fenton WS. Classical subtypes for schizophrenia: literature review for DSM-IV. *Schizophr Bull* 1991;17(4):609-632.
3. Peralta V, Cuesta MJ. How many and which are the psychopathological dimensions in schizophrenia? Issues influencing their ascertainment. *Schizophr Res* Apr 30 2001;49(3):269-285.
4. Fanous AH, Kendler KS. Genetic heterogeneity, modifier genes, and quantitative phenotypes in psychiatric illness: searching for a framework. *Mol Psychiatry* Jan 2005;10(1):6-13.
5. Wright FA, Strug LJ, Doshi VK, et al. Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. *Nat Genet* Jun 2011;43(6):539-546.
6. Leduc V, De Beaumont L, Theroux L, et al. HMGCR is a genetic modifier for risk, age of onset and MCI conversion to Alzheimer's disease in a three cohorts study. *Mol Psychiatry* Jul 2015;20(7):867-873.
7. Sacconi S, Lemmers RJ, Balog J, et al. The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. *Am J Hum Genet* Oct 3 2013;93(4):744-751.
8. Brinkman RR, Mezei MM, Theilmann J, Almquist E, Hayden MR. The likelihood of being affected with Huntington disease by a particular age, for a specific CAG size. *Am J Hum Genet* May 1997;60(5):1202-1210.
9. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* Jun 2008;40(6):703-706.
10. Meyer KB, O'Reilly M, Michailidou K, et al. Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. *Am J Hum Genet* Dec 5 2013;93(6):1046-1060.
11. Gallagher MD, Suh E, Grossman M, et al. TMEM106B is a genetic modifier of frontotemporal lobar degeneration with C9orf72 hexanucleotide repeat expansions. *Acta Neuropathol* Mar 2014;127(3):407-418.
12. Regier DA, Narrow WE, Kuhl EA, Kupfer DJ. The conceptual development of DSM-V. *Am J Psychiatry* Jun 2009;166(6):645-650.
13. Ruderfer DM, Fanous AH, Ripke S, et al. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol Psychiatry* Sep 2014;19(9):1017-1024.
14. Stefansson H, Meyer-Lindenberg A, Steinberg S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* Jan 16 2014;505(7483):361-366.
15. Guilmatre A, Dubourg C, Mosca AL, et al. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Arch Gen Psychiatry* Sep 2009;66(9):947-956.
16. Grayton HM, Fernandes C, Rujescu D, Collier DA. Copy number variations in neurodevelopmental disorders. *Prog Neurobiol* Oct 2012;99(1):81-91.
17. Rosenfeld JA, Ballif BC, Torchia BS, et al. Copy number variations associated with autism spectrum disorders contribute to a spectrum of neurodevelopmental disorders. *Genet Med* Nov 2010;12(11):694-702.
18. Fanous AH, Zhou B, Aggen SH, et al. Genome-wide association study of clinical dimensions of schizophrenia: polygenic effect on disorganized symptoms. *Am J Psychiatry* Dec 2012;169(12):1309-1317.
19. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* Jul 24 2014;511(7510):421-427.
20. Schizophrenia Psychiatric Genome-Wide Association Study C. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* Oct 2011;43(10):969-976.
21. Genomes Project C, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. *Nature* Oct 28 2010;467(7319):1061-1073.
22. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* Nov 2011;1(6):457-470.
23. Andreasen NC, Flaum M, Arndt S. The Comprehensive Assessment of Symptoms and History (CASH). An instrument for assessing diagnosis and psychopathology. *Arch Gen Psychiatry* Aug 1992;49(8):615-623.

24. Levinson DF, Mowry BJ, Escamilla MA, Faraone SV. The Lifetime Dimensions of Psychosis Scale (LDPS): description and interrater reliability. *Schizophr Bull* 2002;28(4):683-695.
25. McGuffin P, Farmer A, Harvey I. A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. *Arch Gen Psychiatry* Aug 1991;48(8):764-770.
26. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull* 1987;13(2):261-276.
27. Wing JK, Babor T, Brugha T, et al. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Arch Gen Psychiatry* Jun 1990;47(6):589-593.
28. Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. *Arch Gen Psychiatry* Aug 1992;49(8):624-629.
29. International Schizophrenia C, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* Aug 6 2009;460(7256):748-752.
30. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* Sep 2007;81(3):559-575.
31. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* Sep 1 2010;26(17):2190-2191.
32. Pers TH, Karjalainen JM, Chan Y, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* 2015;6:5890.
33. Augustin I, Korte S, Rickmann M, Kretschmar HA, Sudhof TC, Herms JW, Brose N. The cerebellum-specific Munc13 isoform Munc13-3 regulates cerebellar synaptic transmission and motor learning in mice. *J Neurosci* Jan 1 2001;21(1):10-17.
34. Chen Z, Cooper B, Kalla S, Varoquaux F, Young SM, Jr. The Munc13 proteins differentially regulate readily releasable pool dynamics and calcium-dependent recovery at a central synapse. *J Neurosci* May 8 2013;33(19):8336-8351.
35. Ashley-Koch AE, Garrett ME, Gibson J, et al. Genome-wide association study of posttraumatic stress disorder in a cohort of Iraq-Afghanistan era veterans. *J Affect Disord* Sep 15 2015;184:225-234.
36. Miller JA, Woltjer RL, Goodenbour JM, Horvath S, Geschwind DH. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome Med* 2013;5(5):48.
37. Genetics of Personality C, de Moor MH, van den Berg SM, et al. Meta-analysis of Genome-wide Association Studies for Neuroticism, and the Polygenic Association With Major Depressive Disorder. *JAMA Psychiatry* Jul 1 2015;72(7):642-650.
38. Luciano M, Huffman JE, Arias-Vasquez A, et al. Genome-wide association uncovers shared genetic effects among personality traits and mood states. *Am J Med Genet B Neuropsychiatr Genet* Sep 2012;159B(6):684-695.
39. Major Depressive Disorder Working Group of the Psychiatric GC, Ripke S, Wray NR, et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* Apr 2013;18(4):497-511.
40. consortium C. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* Jul 30 2015;523(7562):588-591.
41. Liu Q, Zerbinatti CV, Zhang J, et al. Amyloid precursor protein regulates brain apolipoprotein E and cholesterol metabolism through lipoprotein receptor LRP1. *Neuron* Oct 4 2007;56(1):66-78.
42. Iossifov I, O'Roak BJ, Sanders SJ, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* Nov 13 2014;515(7526):216-221.
43. De Rubeis S, He X, Goldberg AP, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* Nov 13 2014;515(7526):209-215.
44. Fanous AH, Amdur RL, O'Neill FA, Walsh D, Kendler KS. Concordance between chart review and structured interview assessments of schizophrenic symptoms. *Compr Psychiatry* Apr 2012;53(3):275-279.
45. Ho BC, Flaum M, Hubbard W, Arndt S, Andreasen NC. Validity of symptom assessment in psychotic disorders: information variance across different sources of history. *Schizophr Res* Jun 1 2004;68(2-3):299-307.

46. Harris D, Batki SL. Stimulant psychosis: symptom profile and acute clinical course. *Am J Addict* Winter 2000;9(1):28-37.
47. Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* Jan 17 2009;373(9659):234-239.
48. Kendler KS, McGuire M, Gruenberg AM, Walsh D. Examining the validity of DSM-III-R schizoaffective disorder and its putative subtypes in the Roscommon Family Study. *Am J Psychiatry* May 1995;152(5):755-764.
49. Kraepelin E, Diefendorf AR. *Clinical psychiatry*. New York, London,: The Macmillan company; 1907.

Figure legends

Figure 1. (a) Predictive value of SCZ polygene scores. Results based on varying SNP P -value inclusion thresholds are grouped by symptom dimension along the x -axis. Proportion of variance explained (adjusted R^2) is shown on the y -axis. Displayed P -values correspond to the inclusion thresholds yielding the largest proportion of variance explained. (b) Forest plot displaying study-specific results for the association of SCZ polygene scores with negative symptoms. p_T is the selected P -value threshold for the polygenic score tested; for each site, β , se , p , and n are the estimated effect and its standard error from linear regression, the corresponding P -value, and the number of cases with available clinical data, respectively.

Figure 2. Manhattan plots for GWAS of four clinical dimensions. Red and blue horizontal lines show thresholds for genome-wide (5×10^{-8}) and suggestive (10^{-6}) significance, respectively. For suggestively associated SNPs, the most significant “independent” marker within a 500kb region is represented as a blue diamond; neighboring SNPs in linkage disequilibrium ($r^2 > 0.1$) with this most significant marker are highlighted.

Table 1. Association of PGC2-trained polygenic scores with symptom dimensions of SZ. For each symptom dimension, the total N of cases, P -value threshold (P_T) yielding the greatest predictive value, variance explained by this score (R^2), and its significance (P -value) are given for the full sample (i.e., “all PGC2”), the MGS study alone, and all non-MGS PGC2 studies.

Symptom	Cohort	N	P_T	R^2	P -value
Positive	MGS only	2 386	0.1	1.53×10^{-3}	0.0558
	non-MGS PGC2	5 944	10^{-4}	1.07×10^{-4}	0.422
	all PGC2	8 330	0.1	6.46×10^{-5}	0.214
Negative	MGS only	2 386	0.1	0.00452	9.95×10^{-4}
	non-MGS PGC2	6 041	0.1	0.00373	1.83×10^{-6}
	all PGC2	8 427	0.1	0.00386	6.39×10^{-9}
Manic	MGS only	2 384	10^{-4}	3.50×10^{-3}	0.00375
	non-MGS PGC2	4 566	0.2	9.49×10^{-5}	0.511
	all PGC2	6 950	10^{-4}	6.39×10^{-4}	0.0204
Depressive	MGS only	2 386	0.1	1.27×10^{-3}	0.0783
	non-MGS PGC2	4 578	0.001	4.43×10^{-4}	0.153
	all PGC2	6 964	0.1	3.84×10^{-4}	0.0546

Table 2. Association results for SNPs attaining $P < 10^{-6}$ in primary GWAS of symptom dimensions. For each SNP, *Chr* and *Mb* give its genomic coordinates (hg19); *INFO* is the statistical imputation information; *Freq* is the frequency of the reference (first listed) allele, and *Z-score* is its estimated effect; *P* and *P_{SZ}* are the *P*-values in the current and full PGC case-control analyses, respectively. The nearest gene within 1Mb is shown; its position relative to a gene is given parenthetically (negative and positive kb values indicate up- and downstream positions).

Symptom	Chr	Mb	SNP	Alleles	INFO	Freq	Z	P	P _{SZ}	Nearest Gene (kb)
Positive	1p13.3	107.5	rs7522504	C/G	0.981	0.327	4.96	7.24×10^{-7}	0.802	<i>PRMT6</i> (-102.2kb)
	2p24.2	19.08	rs72778999	A/T	0.726	0.919	4.93	8.35×10^{-7}	0.0874	<i>NT5C1B</i> (+306kb)
	2q11.2	102.08	rs74263251	A/C	0.913	0.934	5.41	6.27×10^{-8}	0.973	<i>RFX8</i> (0)
	8q21.12	80.02	rs78234477	C/T	0.951	0.014	-5.12	3.06×10^{-7}	0.878	<i>IL7</i> (+304.4kb)
	22q12.1	27.61	rs474758	C/T	0.977	0.781	4.91	8.94×10^{-7}	0.195	<i>LOC101929539</i> (+294.4kb)
Negative	6q16.2	99.99	rs56238691	C/T	0.805	0.014	5.19	2.15×10^{-7}	0.904	<i>CCNC</i> (0)
	14q22.3	57.93	rs2484490	C/T	0.917	0.522	4.9	9.44×10^{-7}	0.0636	<i>C14orf105</i> (-5.61kb)
	15q21.3	54.12	rs144529410	C/G	0.717	0.026	5.38	7.59×10^{-8}	0.531	<i>UNC13C</i> (-188.3kb)
	16q23.1	77.85	rs3909915	C/G	0.748	0.985	-5.13	2.85×10^{-7}	0.890	<i>VAT1L</i> (0)
Depressive	2q33.3	206.1	rs145452099	C/T	0.694	0.097	4.96	7.09×10^{-7}	0.149	<i>PARD3B</i> (0)
	3p26.1	8.31	rs59441182	A/G	0.954	0.64	5.05	4.33×10^{-7}	0.391	<i>LMCD1-AS1</i> (0)
	3q25.2	153.95	rs112126680	G/T	0.688	0.963	4.96	6.96×10^{-7}	0.224	<i>ARHGEF26</i> (0)

Table 3. Gene-set enrichment analysis results for negative symptoms. Nominal *P*-values and estimated false-discovery rate (FDR) for suggestive gene-sets (FDR < 0.20), based on analyses of SNPs significant at a specified *P*-value threshold. For each SNP *P*-value threshold considered, the total number of independent loci considered (pairwise $r^2 < 0.1$ within 500kb) is given parenthetically.

<i>P</i> -value threshold (No. loci)	Gene-set ID	Description	<i>P</i> -value	FDR
1×10^{-5} (29)	GO:0042493	response to drug	3.85×10^{-7}	< 0.05
5×10^{-5} (109)	ENSG00000081479	LRP2 PPI subnetwork	1.81×10^{-6}	< 0.01
	ENSG00000123384	LRP1 PPI subnetwork	5.47×10^{-6}	< 0.01
	MP:0004056	abnormal myocardium compact layer morphology	2.40×10^{-5}	< 0.05
	ENSG00000145362	ANK2 PPI subnetwork	3.24×10^{-5}	< 0.05
	ENSG00000188612	SUMO2 PPI subnetwork	7.34×10^{-5}	< 0.20
	GO:0042417	dopamine metabolic process	1.27×10^{-4}	< 0.20
	GO:0010657	muscle cell apoptotic process	1.72×10^{-4}	< 0.20



