Article

Genome-Wide Association Study of Clinical Dimensions of Schizophrenia: Polygenic Effect on Disorganized Symptoms

Ayman H. Fanous, M.D.
Baiyu Zhou, Ph.D.
Steven H. Aggen, Ph.D.
Sarah E. Bergen, Ph.D.
Richard L. Amdur, Ph.D.
Jubao Duan, Ph.D.
Alan R. Sanders, M.D.
Jianxin Shi, Ph.D.
Bryan J. Mowry, M.D., F.R.A.N.Z.C.P.
Ann Olincy, M.D.
Farooq Amin, M.D.
C. Robert Cloninger, M.D.
Jeremy M. Silverman, Ph.D.
Nancy G. Buccola, M.S.N., A.P.R.N.
William F. Byerley, M.D.
Donald W. Black, M.D.
Robert Freedman, M.D.
Frank Dudbridge, Ph.D.
Peter A. Holmans, Ph.D.
Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium
Stephan Ripke, M.D.
Pablo V. Gejman, M.D.
Kenneth S. Kendler, M.D.
Douglas F. Levinson, M.D.

Objective: Multiple sources of evidence suggest that genetic factors influence variation in clinical features of schizophrenia. The authors present the first genome-wide association study (GWAS) of dimensional symptom scores among individuals with schizophrenia.

Method: Based on the Lifetime Dimensions of Psychosis Scale ratings of 2,454 case subjects of European ancestry from the Molecular Genetics of Schizophrenia (MGS) sample, three symptom factors (positive, negative/disorganized, and mood) were identified with exploratory factor analysis. Quantitative scores for each factor from a confirmatory factor analysis were analyzed for association with 696,491 single-nucleotide polymorphisms (SNPs) using linear regression, with correction for age, sex, clinical site, and ancestry. Polygenic score analysis was carried out to determine whether case and comparison subjects in 16 Psychiatric GWAS Consortium (PGC) schizophrenia samples (excluding MGS samples) differed in scores computed by weighting their genotypes by MGS association test results for each symptom factor.

Results: No genome-wide significant associations were observed between SNPs and factor scores. Most of the SNPs producing the strongest evidence for association were in or near genes involved in neurodevelopment, neuroprotection, or neurotransmission, including genes playing a role in Mendelian CNS diseases, but no statistically significant effect was observed for any defined gene pathway. Finally, polygenic scores based on MGS GWAS results for the negative/disorganized factor were significantly different between case and comparison subjects in the PGC data set; for MGS subjects, negative/disorganized factor scores were correlated with polygenic scores generated using case-control GWAS results from the other PGC samples.

Conclusions: The polygenic signal that has been observed in cross-sample analyses of schizophrenia GWAS data sets could be in part related to genetic effects on negative and disorganized symptoms (i.e., core features of chronic schizophrenia).

(Am J Psychiatry 2012; 169:1309–1317)

Schizophrenia patients differ greatly in observed levels of hallucinations, delusions, and negative, disorganized, manic, and depressive symptoms, as well as in age at onset, course of illness, and comorbidities. Historically, categorical subtypes of schizophrenia were identified, such as the paranoid, catatonic, and hebephrenic subtypes that were combined by Kraepelin into dementia praecox, and the positive and negative subtypes (1). Twin studies of schizophrenia as a category have provided the best evidence to date for strong heritability (70%–80%) (2). Thus, it is possible that most clinical diversity is due to the kind of individual variation in the underlying pathology that is seen in, for example, Huntington's disease, which has one common genetic basis. An alternative approach is to define distinct dimensional features (e.g., a definition based on factor analyses of symptom data) (3).

A number of studies suggest that there is a genetic basis for clinical heterogeneity (4–6). Associations with dimensional components of schizophrenia could provide insights into targets for pharmacological therapy, factors

influencing specific functional impairments, or the clinical subgroups most likely to be relevant to associations with the categorical diagnosis. We previously hypothesized two classes of genetic effects (5): clinical modifier genes that influence features of illness without altering the risk of illness itself and susceptibility modifier genes that influence the risk of illness in a way that affects its clinical features (akin to subtypes of illness).

Genome-wide association study (GWAS) data provide an opportunity to consider dimensional approaches in new ways. Schizophrenia GWAS analyses have detected three types of highly significant effects: 1) common single-nucleotide polymorphisms (SNPs) in at least seven genes or nongenic regions are strongly associated with schizophrenia, although with small individual effects (7); 2) a set of rare chromosomal deletions and duplications (copy number variants) have large effects on risk but only in a small proportion of cases (8); and 3) a robust polygenic effect can be observed by predicting case-control status in a schizophrenia data set by computing scores for each subject that depend on association test results for large numbers of SNPs from a different schizophrenia data set (7, 9). The success of the GWAS approach suggests that it might also be used to explore the genetic basis of clinical heterogeneity.

To our knowledge, this study is the first GWAS analysis of clinical symptom dimensions in schizophrenia. We used data from one of the largest single GWAS (Molecular Genetics of Schizophrenia [MGS] study) (10), in which the assessment protocol included completion of a dimensional rating scale by an expert diagnostician after reviewing multiple sources of current and historical data. We used factor analysis to derive positive, negative/disorganized, and mood factors from these data and tested the association of each factor score with the SNPs from the MGS GWAS (10). We then used data from 16 other data sets in the Psychiatric GWAS Consortium (PGC) schizophrenia analysis (7) to generate polygenic scores for the MGS participants and carried out analyses to determine whether the strong polygenic effect observed across schizophrenia data sets is more strongly associated with any of the clinical dimensions.

Method

Clinical Sample and Assessments

The clinical methods of the study have been described elsewhere (10). Briefly, we examined 2,454 individuals of European ancestry for whom both GWAS data and valid dimensional rating data (2,436 individuals for chromosome X because of additional quality-control exclusions) were available. Participants were recruited through 10 university-based sites in the United States and Australia under a common protocol. They received consensus diagnoses of either DSM-IV schizophrenia (90%) or schizoaffective disorder (with criterion A schizophrenia symptoms for at least 6 months) based on available information from the Diagnostic Interview for Genetic Studies, version 2.0, informant reports, and psychiatric treatment records. At the same time that diagnoses were assigned (i.e., when all sources had been reviewed), a diagnostician also rated clinical features using the Lifetime Dimensions of Psychosis Scale (http://depressiongenetics.stanford.edu/ldps.html), which was designed to quantify the schizophrenia symptom dimensions identified by previous factor-analytic studies (11). The 14 scale items we used are listed in Table 1. For each item, separate ratings were made on 4-point subscales for typical severity and total duration, and these ratings were summed to produce a score for the item for this analysis. An additional four items were partially redundant or had insufficient variance to be useful in this study. Interrater reliability was measured for 41 participants (drawn from all sites) for whom complete scale ratings were obtained from pairs of raters at different sites, with acceptable intraclass correlation coefficients for the positive (0.74), negative/disorganized (0.66), and mood (0.67) factor scores described below.

DNA Extraction and Genotyping

DNA specimens were extracted from lymphocytes or from Epstein-Barr virus-transformed lymphoblastic cell lines and were assayed at the Broad Institute (Cambridge, Mass.) using Affymetrix 6.0 genotyping arrays (Affymetrix, Santa Clara, Calif.). Part of the MGS GWAS sample was genotyped under the auspices of the Genetic Association Information Network, and the remaining samples were genotyped (at the same laboratory several months later) under grant funding, but they constitute a single MGS sample. After quality control, 671,422 autosomal and 25,069 X-chromosome SNPs were selected for analysis (10).

Factor Analysis of the Lifetime Dimensions of Psychosis Scale

Exploratory factor analysis using all MGS GWAS participants with Lifetime Dimensions of Psychosis Scale ratings was performed in Mplus (http://www.statmodel.com/index.shtml) using an oblique geomin rotation (12). Prior to exploratory factor analysis, missing data points were imputed using the Proc MI statistical procedure in SAS (SAS Institute, Cary, N.C.) after excluding participants for whom data were missing for ≥50% of the items. The exploratory factor analysis included 2,454 participants of European ancestry and 1,137 African American participants, but we report only on the larger European ancestry data set rather than combining both data sets, since the genetic architecture for the two groups looks different (10). A three-factor solution was selected as providing the most parsimonious and interpretable factors. Based on the results from the exploratory factor analysis, a variable with a loading of at least 0.4 on a factor was selected as an indicator for that factor in the confirmatory factor analysis if its loadings on each of the other factors was at least 0.2 units less. Confirmatory factor analysis was performed (12) following the exploratory factor analysis structure, specifying a simple model with no cross loadings of items on factors. Goodness-of-fit was assessed using the comparative fit index, Tucker-Lewis index, and root mean square error of approximation from the confirmatory factor analysis.

Association Analysis

We implemented a case-only association test of allelic effects on three quantitative traits: positive, negative/disorganized, and mood factors. We used linear regression as implemented in PLINK (13) to test for allelic effects on scores for these three factors, with covariates including study site (categorical), age, sex, and principal components scores reflecting ancestry effects (five for autosomal SNPs and three for chromosome-X SNPs) (10). Because three different dimensions were tested, the threshold for genome-wide significance was set at 1.67×10^{-8}.

Pathway Analysis

We tested whether any known gene pathways (sets of functionally related genes) were overrepresented in the locations of the best association findings for each dimension using the ALIGATOR (Association LIst Go AnnoTatOR) method, which corrects for confounding factors and sources of bias, such as linkage disequilibrium between SNPs, variable gene size, overlapping genes, and multiple nonindependent gene ontology categories (14). We included pathways from the Gene Ontology, KEGG (Kyoto Encyclopedia of Genes and Genomes), Mouse Genome Informatics, PANTHER (Protein Analysis THrough Evolutionary Relationships), BioCarta, and Reactome databases.

Polygenic Score Analysis

It has been well established that GWAS results from one schizophrenia data set can be used to predict case-control status in a second data set (7, 9). A large genome-wide set of independent autosomal SNPs (which have been genotyped or imputed in each data set) is selected (after pruning to restrict linkage disequilibrium between SNPs); then the effect size beta for the test of association of each tested allele in the first data set is used as a weighting factor to create a polygenic score for each subject in the second data set as the sum across all SNPs for the number of test alleles carried by the subject, times the weight for each allele. The proportion of variance explained is small but increases with the sizes of the two data sets.

We assessed whether the polygenic signal was more closely related to any one symptom dimension. We used MGS dimensional GWAS data for each factor score as a training data set and the remaining 16 PGC data sets (case subjects, N=6,715; comparison subjects, N=9,978) (7) as the test data set. From all HapMap 3 SNPs that were either genotyped or imputed (with information content >0.9 using the Beagle genetic analysis software package [15] for the PGC samples), 110,942 autosomal SNPs were selected, with a linkage disequilibrium (r^2) <0.25 in 500 SNP windows. For each symptom factor separately, analyses were carried out for each of 10 bins of SNPs (Table 2); each bin included SNPs with p values in the MGS GWAS for that dimension that were below the specified values listed in Table 2. In each analysis, effect size beta values from the MGS dimensional analysis were used as weighting factors to compute polygenic scores for each participant in the 16 PGC data sets. PGC case-control status was then predicted by logistic regression analysis of polygenic scores plus covariates (PGC study site and nine principal component scores reflecting ancestry). Each analysis yielded a p value for the overall significance of the prediction of PGC case-control status, while correcting for covariates, and an estimate of the variance in case-control status that was explained (Nagelkerke's R^2 for the full model using the polygenic score plus the covariates, minus R^2 for the covariates alone).

We also examined the same effect in the opposite direction (i.e., not an independent analysis). Polygenic scores for the MGS GWAS case subjects were computed using association test results for the 16 PGC data sets combined, using the subset of the same SNPs that produced the most significant polygenic analysis for the categorical schizophrenia diagnosis (the best 20% of p values in the 16 PGC data sets, predicting MGS case-control status with $p=2.45\times10^{-54}$, 6.35% of variance explained). We then used linear regression to determine whether polygenic scores for the MGS case subjects were predicted by each factor score plus MGS ancestry and site covariates, and we report the p value for the effect of each factor score.

TABLE 1. Factor Loadings in the Exploratory Factor Analysis of the Lifetime Dimensions of Psychosis Scale Ratings in the Molecular Genetics of Schizophrenia Sample[a]

	Factor Loadings		
Sign/Symptom	Positive	Negative/ Disorganized	Mood
Delusions	**0.836**	0.288	0.233
Paranoia	**0.773**	0.243	0.188
Hallucinations	**0.779**	0.088	−0.108
Control delusions	**0.546**	0.156	0.148
Conversing/commenting/ continuous hallucinations	**0.746**	0.012	−0.066
Abnormal perception of thought	**0.504**	0.100	0.118
Blunted affect	0.145	**0.668**	−0.157
Poverty of speech	0.076	**0.707**	−0.162
Formal thought disorder	0.175	**0.597**	0.084
Bizarre behavior	0.188	**0.565**	0.114
Depression	0.185	−0.283	**0.450**
Mania	−0.033	0.063	**0.897**
Depression with psychosis	0.263	−0.172	**0.465**
Mania with psychosis	0.015	0.140	**0.934**

[a] Items in bold were specified in the confirmatory factor analysis as defining that factor, without cross loadings, and orthogonal factor scores were defined by the confirmatory factor analysis; these scores were analyzed for association with single-nucleotide polymorphism genotypes and corrected for age, sex, site, and ancestry covariates.

Results

Factor Analysis

Eigenvalues, exploratory factor analysis model fit indices, and clinical judgment were used to select a three-factor model as the most adequate and parsimonious representation of the item associations. Exploratory factor analysis factors and their item loadings are listed in Table 1. The three factors (clinical dimensions) were labeled as positive, negative/disorganized, and affective. The confirmatory factor analysis model fit indices for this three-factor model were the comparative fit index (0.91), Tucker-Lewis index (0.90), and root mean square error of approximation (0.12). Additional factors could have been extracted to improve the statistical fit, but such factors were more poorly marked and less likely to be replicable and meaningfully interpreted. The three-factor solution is clinically intuitive and consistent with previous studies. It is possible that ascertainment or rater differences across sites may have also contributed to the lower fit index values. However, as noted above, we accounted for the site mean differences as well as age and sex effects on the factor scores in the association regression models.

GWAS of Symptom Dimensions

Genomic inflation factors (λ) for analyses of positive, negative/disorganized, and affective factors were 0.98, 1.0, and 1.01, respectively, indicating no significant inflation of results by technical factors or population stratification. No

TABLE 2. Polygenic Score Analyses of Prediction of Psychiatric GWAS Consortium (PGC) Case-Control Status by Results of Each Molecular Genetics of Schizophrenia (MGS) Dimensional GWAS[a]

Symptom Factor	Single-Nucleotide Polymorphisms (SNPs)	Symptom Factor[b]					
		Positive		Negative/ Disorganized		Mood	
p-Value Threshold (Dimensional GWAS) to Select SNPs	N	p	Variance Explained[c]	p	Variance Explained[c]	p	Variance Explained[c]
0.0001	138	0.67	−1.29E-05	0.28	8.09E-05	0.69	−1.14E-05
0.001	654	0.46	−3.80E-05	0.25	9.32E-05	0.04	−0.00028
0.01	3,759	0.66	−1.33E-05	0.31	7.18E-05	0.27	−8.56E-05
0.05	13,289	0.98	−4.25E-08	0.24	9.60E-05	0.50	−3.19E-05
0.1	22,736	0.93	−5.84E-07	**0.01**	**0.0004**	0.59	−2.00E-05
0.2	38,939	0.78	5.45E-06	**0.01**	**0.0005**	~1.00	3.34E-10
0.3	52,843	0.62	1.73E-05	**0.02**	**0.0004**	0.97	−1.09E-07
0.4	64,993	0.75	7.07E-06	**0.01**	**0.0004**	0.75	−6.56E-06
0.5	76,114	0.81	4.02E-06	**0.006**	**0.0005**	0.81	−4.00E-06
1.0	110,942	0.77	6.07E-06	**0.007**	**0.0005**	0.73	−8.38E-06

[a] The data shown are results of 30 analyses, 10 for each symptom factor, based on a set of 110,942 SNPs without linkage disequilibrium (marker-marker correlation) >0.25. For each analysis, SNPs were selected that had p values below the threshold (listed in the first column) in the GWAS of that symptom factor in the MGS data set, and the effect size beta values for each SNP in that analysis were used as weights to compute polygenic scores in the 16 other PGC data sets. In each analysis, separately for each symptom factor, SNPs were selected according to their p value for association. An overall p value is shown for prediction of PGC case-control status (see the article text).

[b] Only the negative/disorganized factor MGS GWAS results predicted PGC case-control status with a typical pattern of gradually increasing significance with larger subsets of SNPs (although explaining a small proportion of variance), suggesting that the polygenic effect observed in schizophrenia GWAS analyses could be more closely related to negative/disorganized symptoms than to other symptom dimensions. Further analyses suggested that in the present data set, the effect was primarily due to disorganized symptoms (see the article text).

[c] Variance explained is Nagelkerke's R^2, minus R^2 for regression of covariates alone (negative values indicate an effect in the nonpredicted direction).

genome-wide significant associations were observed for any clinical dimension. Data for SNPs with a p value $<10^{-5}$ are summarized in Table 3, including gene symbols and a brief summary of functions. Only one region (chromosome 20q13.31) produced moderate evidence for association with two different factors (positive and negative/disorganized).

Only one region produced evidence for genome-wide significant association in the PGC two-stage analysis (full GWAS data for 9,394 case subjects and 12,462 comparison subjects and the addition of data for the most significant SNPs from 8,442 case subjects and 21,397 comparison subjects) (10). The PGC observed significant association for multiple SNPs across the major histocompatibility complex region, spanning the HLA (human leukocyte antigen) genes, and we observed moderate evidence for association of negative/disorganized factor scores with SNPs downstream of *HLA-DQA1*.

Pathway Analyses

Pathway analyses were performed separately for SNPs within genes (267,899 SNPs, 15,998 genes) and then for SNPs within 20 kb of genes (360,811 SNPs, 22,604 genes). The threshold for selecting significant SNPs in this context was set such that 5% of genes included one such SNP (p=0.007 and 799 genes for SNPs within genes; p=0.005 and 1,130 genes for SNPs within 20 kb of genes). In both analyses, the number of pathways that were enriched (i.e., pathways that contained more significant genes than expected by chance) did not reach overall significance after correction for multiple testing.

Polygenic Score Analyses

Results for the prediction of PGC case-control status with polygenic scores based on each MGS dimensional GWAS analysis are summarized in Table 2. For the negative/disorganized factor, p values became nominally significant when polygenic scores for participants in the PGC study were computed based on results of the best 10% of SNPs in the dimensional GWAS analysis, with the lowest p value (0.007) obtained using all SNPs, although only 0.05% of the variance in PGC case-control status was predicted. There was no evidence that polygenic scores based on the positive or mood factor GWAS results could predict PGC case-control status.

In a related analysis of the MGS case subjects, polygenic scores were computed based on log odds ratio values from the other 16 PGC data sets and were used to predict (by linear regression) each factor score, with site, sex, age at interview, and MGS ancestry principal components as covariates; p values for negative/disorganized, positive, and mood factors were 0.03, 0.5, and 0.7, respectively. There was no significant interaction between sex and polygenic scores in predicting negative/disorganized factor scores. To further explore the relationship between negative/disorganized factor and polygenic scores, we carried out separate linear regression analyses of the raw sums of severity plus the duration ratings for Lifetime Dimensions of Psychosis Scale items for negative (blunted affect and poverty of speech) and disorganized (formal thought disorder and disorganized behavior) symptoms as

TABLE 3. Single-Nucleotide Polymorphisms (SNPs) of Moderate Association to Each Symptom Dimension[a]

SNP	Chromosome/ Band	Location (Base Pair)	Beta	p	Closest Gene (Symbol, Distance [base pair], Gene Name)	Function/Relevance
Positive symptoms						
rs7233060	18q23	75,493,367	0.1225	2.53×10^{-07}	*CTDP1*, −47421, CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) phosphatase, subunit 1	Makes *POLR2A* available for initiation of gene expression; mutations cause Charcot-Marie-Tooth (demyelinating) disease
rs17206232	5q12.3	64,469,156	0.1350	1.45×10^{-06}	*ADAMTS6*, 11162, ADAM metallopeptidase with thrombospondin type 1 motif, 6	*ADAMTS4*/*ADAMTS5* induce neurite extension in cultured neurons (25)
rs2323266	13q21.2	60,863,304	−0.1032	3.13×10^{-06}	*PCDH20*, 18515, protocadherin 20	Neuronal survival, synaptogenesis (26). Hippocampal circuitry formation, synaptic plasticity (27). Variants in *PCDH19* associated with epilepsy (28, 29).
rs10900020	10q11.21	44,147,203	−0.1536	3.46×10^{-06}	*CXCL12*, 38407, chemokine (C-X-C motif) ligand 12	Diverse roles in neuronal migration, growth factor signaling, neuroprotection (30). Increased GABA, glutamate, dopamine release (31).
rs10052004	5q11.2	56,702,257	−0.09205	3.62×10^{-06}	Intergenic	
rs959770	4p16.3	2,365,095	−0.1198	9.40×10^{-06}	*ZFYVE28*, within, zinc finger, FYVE domain containing 28	Regulation of epidermal growth factor receptor activity
rs11699237	20q13.31	55,117,403	0.1644	9.96×10^{-06}	Intergenic	
Negative/disorganized symptoms						
rs11699237	20q13.31	55,117,403	0.1221	3.13×10^{-06}	Intergenic	
rs1455244	18p11.21	11,484,199	−0.06046	3.22×10^{-06}	Intergenic (195 kb upstream of closest gene, *GNAL*, guanine nucleotide binding protein [G protein], alpha activating activity polypeptide, olfactory type)	Coupled to mesolimbic and mesocortical dopamine-1 receptors (32)
rs7172342	15q22.2	59,123,734	−0.0795	3.83×10^{-06}	*RORA*, within, RAR-related orphan receptor A	Transcription factor involved in cerebellar dendritic development and synapse formation (33). Decreased expression in autism (34).
rs4530903	6p21.32	32,689,867	0.09191	4.83×10^{-06}	Between *HLA-DRB1* and *HLA-DQA1*, 35343, −23293, major histocompatibility complex, class II genes	Immunity. Common SNPs in this region are strongly associated with schizophrenia (7, 9, 10, 22).
rs13278432	8q13.2	68,884,401	0.07197	9.65×10^{-06}	Intergenic	
Mood symptoms						
rs1920592	12q24.21	113,189,969	−0.1125	1.05×10^{-06}	Intergenic	
rs4798896	18q23	74,013,910	0.09174	3.81×10^{-06}	Intergenic	
rs489332	9q21.13	77,218,166	−0.1097	5.57×10^{-06}	Intergenic	
rs1351267	3q25.1	153,246,391	0.08821	6.83×10^{-06}	Intergenic	
rs10924245	1q44	243,800,231	−0.1661	6.93×10^{-06}	*KIF26B*,within, kinesin family member 26B	Regulation of cell-cell adhesion
rs17290922	16q13	55,581,818	−0.1338	7.78×10^{-06}	*NLRC5*,within, nucleotide-binding oligomerization domains 27	Induce major histocompatibility class-I genes (35)
rs4702765	5p15.2	10,980,604	0.2237	1.06×10^{-05}	*CTNND2*, 44347, catenin (cadherin-associated protein), delta 2	Binds presenilin-1. Maintenance of dendrites and dendritic spines (36). Mutations cause cri du chat syndrome (37). Rare copy number variant observed in schizophrenia (38).

[a] The most significant SNP from any cluster of nearby SNPs with p values $<10^{-5}$ are listed.

predictors of polygenic scores, with site and ancestry component covariates. A significant effect was observed for disorganized symptoms (p=0.004) but not negative symptoms (p=0.37); analyzed separately, both disorganized symptom items contributed to the prediction of polygenic scores (formal thought disorder, p=0.01; bizarre behavior, p=0.03).

Discussion

To our knowledge, this is the first GWAS of clinical dimensions of schizophrenia. There have been several previous reports of relationships between putative schizophrenia candidate genes and clinical measures (16–20). SNPs in *DTNBP1* were reported to be more strongly associated with negative symptoms and SNPs in *COMT* with manic symptoms in two independent samples (16–19). SNPs in *ZNF804A* were reported to be more strongly associated with manic-like symptoms in one sample (20). Another study presented association results for a small case-control sample in regions with previously demonstrated evidence for linkage to schizophrenia symptom factors and reported SNPs with moderate levels of association with positive and disorganized symptom scores (21).

In the present study, we did not detect any association for clinical factor scores at a genome-wide significant threshold of significance, which is not surprising given that much larger samples have been required to detect significant associations of schizophrenia with common SNPs (7, 9, 10, 22). With one exception, there was no overlap between the best MGS dimensional GWAS association signals and the significant associations detected by the PGC. This suggests either that differential genetic effects on symptoms (if they exist) are largely distinct from those on risk of illness or that much larger samples are needed to detect individual SNPs that influence both symptom dimensions and illness risk. The exception was the moderate association that we observed between negative/disorganized symptoms and SNPs between *HLA-DRB-1* and *HLA-DQA1*, part of the broad major histocompatibility complex region (spanning all of the HLA genes) in which many SNPs are significantly associated with schizophrenia (7, 9, 10, 22). It is not yet known how sequence variation in HLA genes predisposes to schizophrenia or whether and why this might be more related to negative/disorganized symptoms.

The most intriguing result is that case-control status of participants in the PGC analysis was predicted by polygenic scores that were computed on the basis of MGS association test results for negative/disorganized scores for thousands of SNPs, with the signal here apparently generated primarily by ratings of disorganized symptoms (formal thought disorder and bizarre behavior). This suggests that the well-replicated polygenic effect seen in cross-data-set analyses of schizophrenia (7, 9) might be most closely related to these aspects of the disorder, which in turn suggests that treatments might be able to target these features. Note that within-subject analyses of MGS factor scores are unlikely to be related to case-control analyses: when case subjects have a higher frequency of specific SNP alleles than comparison subjects, the polygenic effect observed in our study would not be detected if factor scores were randomly distributed among case subjects. The effect is modest and is difficult to correct for multiple testing because 10 partially correlated analyses were carried out for each factor score. However, the pattern of results is typical of other schizophrenia polygenic analyses, becoming gradually more significant as larger proportions of SNPs are included. This is believed to be the case because many SNPs influence risk, many of them with very small effect sizes that produce completely nonsignificant individual p values in most GWAS data sets such that their effects can only be detected in aggregate (9, 23). However, when we used MGS case-control GWAS results as weights for polygenic scores in the remaining 16 PGC data sets, 2.2% of the variance in case-control status in those data sets could be predicted, much larger than the 0.05% of variance that can be predicted with polygenic scores based on MGS negative/dimensional GWAS results.

The size of the polygenic effect that can be detected for symptom scores may be restricted by what we view as an inherent noisiness of clinical ratings in schizophrenia, such that it is noteworthy to detect any genetic association signal using factor scores. Clinical ratings rely on the self-report of patients who may fail to recognize or may deny their symptoms, as well as on records (often cursory) from brief hospital stays and clinic visits. We also observed site differences in factor score means, and we cannot determine whether these were due to true differences in sampling or subtle differences in rater styles. Nevertheless, our three-factor solution is clinically intuitive and consistent with previous work. Factor analytic studies of schizophrenia have been reviewed by Peralta and Cuesta (3). Selected models have typically included three to five factors, including various combinations of positive, "bizarre" positive (Schneiderian), negative, disorganized, manic, and depressive factors. It has not been unusual to see (as in our study) negative symptoms combined in one factor with disorganized symptoms, positive with bizarre positive symptoms, and manic with depressive features.

Larger sample sizes are needed to determine whether significant associations with symptom dimensions can be detected for individual SNPs, genes, and pathways. We note that most of the best-supported genes in our study have functions (as summarized in Table 3) that could plausibly be related to schizophrenia, including involvement in known CNS diseases and roles in neurodevelopment, neuroprotection, and neurotransmission.

A number of methodological limitations of this study should be considered. We cannot rule out the possibility that other factor solutions (e.g., with disorganization,

bizarre psychosis, mania, or depression symptoms in separate factors) or other rating scales or procedures might produce stronger genetic associations. We also lacked sufficient systematic information to study environmental variables, such as lifetime cannabis abuse, immigration, and urbanicity, which tend to exert their putative effects through early exposures that are difficult to capture retrospectively (24). Additionally, we lacked formal cognitive testing of subjects, which might shed light on whether the clinical ratings of disorganized symptoms were related to specific neuropsychological impairments. The most critical limitations are those that constrain the power of the analyses (as discussed above): sample size, which was insufficient to produce genome-wide significant association results for individual SNPs, and the imprecision with which clinical symptoms can be measured.

Regarding sample size, this is the largest schizophrenia genetics study with a single assessment protocol that included detailed lifetime symptom ratings by expert raters, and thus our results deserve to be considered separately as well as in combination with other samples that were rated by other methods. The PGC is undertaking such a cross-data-set analysis (in which we are taking part), which could shed additional light on whether significant associations can be observed between individual SNPs and symptom dimensions and whether the polygenic effect on negative/disorganized symptoms can be replicated and strengthened despite the need to combine different types of rating systems from different studies.

In conclusion, we carried out GWAS analyses of positive, negative/disorganized, and mood factor scores in 2,454 individuals with schizophrenia. No single SNP produced significant evidence for association at a genome-wide threshold, and thus larger samples will be required to search for these associations. However, a polygenic score analysis produced evidence that there is a relationship between negative/disorganized factor scores and the polygenic signal that is observed in cross-sample analyses of schizophrenia GWAS data sets, with further analyses suggesting that this effect was primarily due to disorganized symptoms (duration and severity of formal thought disorder and bizarre behavior). This suggests that at least part of the effect of multiple common SNPs is on the deteriorative course of illness that has generally been considered the hallmark of the syndrome.

Received Feb. 14, 2012; revision received May 23, 2012; accepted June 20, 2012 (doi: 10.1176/appi.ajp.2012.12020218). From the Mental Health Service Line, Washington, D.C.; the Veterans Affairs Medical Center, Washington, D.C.; Department of Psychiatry, Georgetown University School of Medicine, Washington, D.C.; Departments of Psychiatry and Human Genetics, Virginia Commonwealth University, Richmond, Va.; Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, Calif.; the Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetics Research, Massachusetts General Hospital, Boston; the Center for Psychiatric Genetics, NorthShore University HealthSystem Research Institute, Evanston, Ill.; Queensland Brain Institute, University of Queensland, Brisbane and Queensland Centre for Mental Health Research, Brisbane, Australia; Department of Psychiatry, University of Colorado Denver, Aurora, Colo.; Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, and Emory University, Atlanta; Department of Psychiatry, Washington University, St. Louis, Mo.; Department of Psychiatry, Mount Sinai School of Medicine, N.Y.; School of Nursing, Louisiana State University Health Sciences Center, New Orleans; Department of Psychiatry, University of California at San Francisco, San Francisco; the Mental Health Clinical Research Center, and Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City; the Medical Research Council-Biostatistics Unit, Institute of Public Health, Cambridge, United Kingdom; the MRC Centre for Neuropsychiatric Genetics and Genomics, Department of Psychological Medicine and Neurology, School of Medicine, Heath Park, Cardiff, United Kingdom; and the Center for Human Genetic Research, Massachusetts General Hospital, Boston, and Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Mass.

Dr. Black has received research support from AstraZeneca. All other authors report no financial relationships with commercial interests.

Drs. Fanous and Zhou contributed equally to this study.

Members of the Genome-Wide Association Study Consortium are listed in the data supplement accompanying the online edition of this article.

Data collection, genotyping, and analysis were supported by NIH R01 grants (MH-67257 to Dr. Buccola, MH-59588 to Dr. Mowry, MH-59571 to Dr. Gejman, MH-59565 to Dr. Freedman, MH-59587 to Dr. Amin, MH-60870 to Dr. Byerley, MH-59566 to Dr. Black, MH-59586 to Dr. Silverman, MH-61675 to Dr. Levinson, MH-60879 to Dr. Cloninger, and MH-81800 to Dr. Gejman), NIH U01 grants (MH-46276 to Dr. Cloninger, MH-46289 to Dr. Charles Kaufmann, MH-46318 to Dr. Ming T. Tsuang, MH-79469 to Dr. Gejman, and MH-79470 to Dr. Levinson), VA grant 1I01CX000278 (to Dr. Fanous), NARSAD Young Investigator Awards (to Drs. Duan and Saunders), the Genetic Association Information Network (GAIN), and the Paul Michael Donovan Charitable Foundation. Genotyping was carried out by the Broad Institute Center for Genotyping and Analysis at MIT and Harvard (by S. Gabriel and D.B. Mirel), with support from grant U54 RR020278 from the National Center for Research Resources. Genotyping of half of the European ancestry sample and almost all the African American sample was carried out with support from GAIN.

The GAIN quality control team (G.R. Abecasis and J. Paschall) made important contributions to the Molecular Genetics of Schizophrenia genome-wide association study project.

The authors thank the study participants and research staff at the study sites.

References

1. Andreasen NC, Carpenter WT Jr: Diagnosis and classification of schizophrenia. Schizophr Bull 1993; 19:199–214
2. Farmer AE, McGuffin P, Gottesman II: Twin concordance for DSM-III schizophrenia. Scrutinizing the validity of the definition. Arch Gen Psychiatry 1987; 44:634–641
3. Peralta V, Cuesta MJ: How many and which are the psychopathological dimensions in schizophrenia? Issues influencing their ascertainment. Schizophr Res 2001; 49:269–285
4. Cardno AG, Sham PC, Murray RM, McGuffin P: Twin study of symptom dimensions in psychoses. Br J Psychiatry 2001; 179: 39–45
5. Fanous AH, Kendler KS: Genetic heterogeneity, modifier genes, and quantitative phenotypes in psychiatric illness: searching for a framework. Mol Psychiatry 2005; 10:6–13
6. Fanous AH, Kendler KS: Genetics of clinical features and subtypes of schizophrenia: a review of the recent literature. Curr Psychiatry Rep 2008; 10:164–170
7. Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, Lin DY, Duan J, Ophoff RA, Andreassen OA, Scolnick E,

Cichon S, St Clair D, Corvin A, Gurling H, Werge T, Rujescu D, Blackwood DH, Pato CN, Malhotra AK, Purcell S, Dudbridge F, Neale BM, Rossin L, Visscher PM, Posthuma D, Ruderfer DM, Fanous A, Stefansson H, Steinberg S, Mowry BJ, Golimbet V, De Hert M, Jönsson EG, Bitter I, Pietiläinen OP, Collier DA, Tosato S, Agartz I, Albus M, Alexander M, Amdur RL, Amin F, Bass N, Bergen SE, Black DW, Børglum AD, Brown MA, Bruggeman R, Buccola NG, Byerley WF, Cahn W, Cantor RM, Carr VJ, Catts SV, Choudhury K, Cloninger CR, Cormican P, Craddock N, Danoy PA, Datta S, de Haan L, Demontis D, Dikeos D, Djurovic S, Donnelly P, Donohoe G, Duong L, Dwyer S, Fink-Jensen A, Freedman R, Freimer NB, Friedl M, Georgieva L, Giegling I, Gill M, Glenthøj B, Godard S, Hamshere M, Hansen M, Hansen T, Hartmann AM, Henskens FA, Hougaard DM, Hultman CM, Ingason A, Jablensky AV, Jakobsen KD, Jay M, Jürgens G, Kahn RS, Keller MC, Kenis G, Kenny E, Kim Y, Kirov GK, Konnerth H, Konte B, Krabbendam L, Krasucki R, Lasseter VK, Laurent C, Lawrence J, Lencz T, Lerer FB, Liang KY, Lichtenstein P, Lieberman JA, Linszen DH, Lönnqvist J, Loughland CM, Maclean AW, Maher BS, Maier W, Mallet J, Malloy P, Mattheisen M, Mattingsdal M, McGhee KA, McGrath JJ, McIntosh A, McLean DE, McQuillin A, Melle I, Michie PT, Milanova V, Morris DW, Mors O, Mortensen PB, Moskvina V, Muglia P, Myin-Germeys I, Nertney DA, Nestadt G, Nielsen J, Nikolov I, Nordentoft M, Norton N, Nöthen MM, O'Dushlaine CT, Olincy A, Olsen L, O'Neill FA, Orntoft TF, Owen MJ, Pantelis C, Papadimitriou G, Pato MT, Peltonen L, Petursson H, Pickard B, Pimm J, Pulver AE, Puri V, Quested D, Quinn EM, Rasmussen HB, Réthelyi JM, Ribble R, Rietschel M, Riley BP, Ruggeri M, Schall U, Schulze TG, Schwab SG, Scott RJ, Shi J, Sigurdsson E, Silverman JM, Spencer CC, Stefansson K, Strange A, Strengman E, Stroup TS, Suvisaari J, Terenius L, Thirumalai S, Thygesen JH, Timm S, Toncheva D, van den Oord E, van Os J, van Winkel R, Veldink J, Walsh D, Wang AG, Wiersma D, Wildenauer DB, Williams HJ, Williams NM, Wormley B, Zammit S, Sullivan PF, O'Donovan MC, Daly MJ, Gejman PV; Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium: Genome-wide association study identifies five new schizophrenia loci. Nat Genet 2011; 43:969–976

8. Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J, Zhang N, Mowry BJ, Olincy A, Amin F, Cloninger CR, Silverman JM, Buccola NG, Byerley WF, Black DW, Kendler KS, Freedman R, Dudbridge F, Pe'er I, Hakonarson H, Bergen SE, Fanous AH, Holmans PA, Gejman PV: Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am J Psychiatry 2011; 168:302–316
9. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P; International Schizophrenia Consortium: Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009; 460:748–752
10. Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, Dudbridge F, Holmans PA, Whittemore AS, Mowry BJ, Olincy A, Amin F, Cloninger CR, Silverman JM, Buccola NG, Byerley WF, Black DW, Crowe RR, Oksenberg JR, Mirel DB, Kendler KS, Freedman R, Gejman PV: Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature 2009; 460:753–757
11. Levinson DF, Mowry BJ, Escamilla MA, Faraone SV: The Lifetime Dimensions of Psychosis Scale (LDPS): description and interrater reliability. Schizophr Bull 2002; 28:683–695
12. Muthen LK, Muthen BO: Mplus: statistical analysis with latent variables: user's guide. Los Angeles, Muthen and Muthen, 2001
13. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559–575
14. Holmans P, Green EK, Pahwa JS, Ferreira MA, Purcell SM, Sklar P, Owen MJ, O'Donovan MC, Craddock N; Wellcome Trust Case-Control Consortium: Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. Am J Hum Genet 2009; 85:13–24
15. Browning SR, Browning BL: Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet 2007; 81:1084–1097
16. DeRosse P, Funke B, Burdick KE, Lencz T, Ekholm JM, Kane JM, Kucherlapati R, Malhotra AK: Dysbindin genotype and negative symptoms in schizophrenia. Am J Psychiatry 2006; 163: 532–534
17. DeRosse P, Funke B, Burdick KE, Lencz T, Goldberg TE, Kane JM, Kucherlapati R, Malhotra AK: COMT genotype and manic symptoms in schizophrenia. Schizophr Res 2006; 87:28–31
18. Fanous AH, van den Oord EJ, Riley BP, Aggen SH, Neale MC, O'Neill FA, Walsh D, Kendler KS: Relationship between a high-risk haplotype in the DTNBP1 (dysbindin) gene and clinical features of schizophrenia. Am J Psychiatry 2005; 162: 1824–1832
19. McClay JL, Fanous A, van den Oord EJ, Webb BT, Walsh D, O'Neill FA, Kendler KS, Chen X: Catechol-O-methyltransferase and the clinical features of psychosis. Am J Med Genet B Neuropsychiatr Genet 2006; 141B:935–938
20. Cummings E, Donohoe G, McDonald C, Dinan TG, O'Neill FA, O'Callaghan E, Waddington JL, Murphy KC, Gill M, Morris DW, Corvin A: Clinical symptomatology and the psychosis risk gene ZNF804A. Schizophr Res 2010; 122:273–275
21. DeRosse P, Lencz T, Burdick KE, Siris SG, Kane JM, Malhotra AK: The genetics of symptom-based phenotypes: toward a molecular classification of schizophrenia. Schizophr Bull 2008; 34: 1047–1053
22. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, Werge T, Pietiläinen OP, Mors O, Mortensen PB, Sigurdsson E, Gustafsson O, Nyegaard M, Tuulio-Henriksson A, Ingason A, Hansen T, Suvisaari J, Lonnqvist J, Paunio T, Børglum AD, Hartmann A, Fink-Jensen A, Nordentoft M, Hougaard D, Norgaard-Pedersen B, Böttcher Y, Olesen J, Breuer R, Möller HJ, Giegling I, Rasmussen HB, Timm S, Mattheisen M, Bitter I, Réthelyi JM, Magnusdottir BB, Sigmundsson T, Olason P, Masson G, Gulcher JR, Haraldsson M, Fossdal R, Thorgeirsson TE, Thorsteinsdottir U, Ruggeri M, Tosato S, Franke B, Strengman E, Kiemeney LA, Melle I, Djurovic S, Abramova L, Kaleda V, Sanjuan J, de Frutos R, Bramon E, Vassos E, Fraser G, Ettinger U, Picchioni M, Walker N, Toulopoulou T, Need AC, Ge D, Yoon JL, Shianna KV, Freimer NB, Cantor RM, Murray R, Kong A, Golimbet V, Carracedo A, Arango C, Costas J, Jönsson EG, Terenius L, Agartz I, Petursson H, Nöthen MM, Rietschel M, Matthews PM, Muglia P, Peltonen L, St Clair D, Goldstein DB, Stefansson K, Collier DA; Genetic Risk and Outcome in Psychosis (GROUP): Common variants conferring risk of schizophrenia. Nature 2009; 460:744–747
23. Visscher PM, Goddard ME, Derks EM, Wray NR: Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. Mol Psychiatry 2011; 17: 474–485
24. van Os J, Kenis G, Rutten BP: The environment and schizophrenia. Nature 2010; 468:203–212
25. Hamel MG, Ajmo JM, Leonardo CC, Zuo F, Sandy JD, Gottschall PE: Multimodal signaling by the ADAMTSs (a disintegrin and metalloproteinase with thrombospondin motifs) promotes neurite extension. Exp Neurol 2008; 210:428–440
26. Morishita H, Yagi T: Protocadherin family: diversity, structure, and function. Curr Opin Cell Biol 2007; 19:584–592
27. Kim SY, Mo JW, Han S, Choi SY, Han SB, Moon BH, Rhyu IJ, Sun W, Kim H: The expression of non-clustered protocadherins in adult rat hippocampal formation and the connecting brain regions. Neuroscience 2010; 170:189–199

28. Jamal SM, Basran RK, Newton S, Wang Z, Milunsky JM: Novel de novo PCDH19 mutations in three unrelated females with epilepsy female restricted mental retardation syndrome. Am J Med Genet A 2010; 152A:2475–2481
29. Marini C, Mei D, Parmeggiani L, Norci V, Calado E, Ferrari A, Moreira A, Pisano T, Specchio N, Vigevano F, Battaglia D, Guerrini R: Protocadherin 19 mutations in girls with infantile-onset epilepsy. Neurology 2010; 75:646–653
30. Li M, Chang CJ, Lathia JD, Wang L, Pacenta HL, Cotleur A, Ransohoff RM: Chemokine receptor CXCR4 signaling modulates the growth factor-induced cell cycle of self-renewing and multipotent neural progenitor cells. Glia 2011; 59:108–118
31. Guyon A, Nahon JL: Multiple actions of the chemokine stromal cell-derived factor-1alpha on neuronal activity. J Mol Endocrinol 2007; 38:365–376
32. Zhuang X, Belluscio L, Hen R: G(olf)alpha mediates dopamine D1 receptor signaling. J Neurosci 2000; 20:RC91
33. Gold DA, Gent PM, Hamilton BA: ROR alpha in genetic control of cerebellum development: 50 staggering years. Brain Res 2007; 1140:19–25
34. Nguyen A, Rauch TA, Pfeifer GP, Hu VW: Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. FASEB J 2010; 24:3036–3051
35. Meissner TB, Li A, Biswas A, Lee KH, Liu YJ, Bayir E, Iliopoulos D, van den Elsen PJ, Kobayashi KS: NLR family member NLRC5 is a transcriptional regulator of MHC class I genes. Proc Natl Acad Sci USA 2010; 107:13794–13799
36. Matter C, Pribadi M, Liu X, Trachtenberg JT: Delta-catenin is required for the maintenance of neural structure and function in mature cortex in vivo. Neuron 2009; 64:320–327
37. Medina M, Marinescu RC, Overhauser J, Kosik KS: Hemizygosity of delta-catenin (CTNND2) is associated with severe mental retardation in cri-du-chat syndrome. Genomics 2000; 63:157–164
38. Vrijenhoek T, Buizer-Voskamp JE, van der Stelt I, Strengman E, Sabatti C, Geurts van Kessel A, Brunner HG, Ophoff RA, Veltman JA; Genetic Risk and Outcome in Psychosis (GROUP) Consortium: Recurrent CNVs disrupt three candidate genes in schizophrenia patients. Am J Hum Genet 2008; 83:504–510