Uncovering the Hidden Risk Architecture of the Schizophrenias: Confirmation in Three Independent Genome-Wide Association Studies

Javier Arnedo, M.S., Dragan M. Svrakic, M.D., Ph.D., Coral del Val, Ph.D., Rocío Romero-Zaliz, Ph.D., Helena Hernández-Cuervo, M.D., Molecular Genetics of Schizophrenia Consortium, Ayman H. Fanous, M.D., Michele T. Pato, M.D., Carlos N. Pato, M.D., Ph.D., Gabriel A. de Erausquin, M.D., Ph.D., C. Robert Cloninger, M.D., Ph.D., Igor Zwir, Ph.D.

Objective: The authors sought to demonstrate that schizophrenia is a heterogeneous group of heritable disorders caused by different genotypic networks that cause distinct clinical syndromes.

Method: In a large genome-wide association study of cases with schizophrenia and controls, the authors first identified sets of interacting single-nucleotide polymorphisms (SNPs) that cluster within particular individuals (SNP sets) regardless of clinical status. Second, they examined the risk of schizophrenia for each SNP set and tested replicability in two independent samples. Third, they identified genotypic networks composed of SNP sets sharing SNPs or subjects. Fourth, they identified sets of distinct clinical features that cluster in particular cases (phenotypic sets or clinical syndromes) without regard for their genetic background. Fifth, they tested whether SNP sets were associated with distinct phenotypic sets in a replicable manner across the three studies.

Results: The authors identified 42 SNP sets associated with a 70% or greater risk of schizophrenia, and confirmed 34 (81%) or more with similar high risk of schizophrenia in two independent samples. Seventeen networks of SNP sets did not share any SNP or subject. These disjoint genotypic networks were associated with distinct gene products and clinical syndromes (i.e., the schizophrenias) varying in symptoms and severity. Associations between genotypic networks and clinical syndromes were complex, showing multifinality and equifinality. The interactive networks explained the risk of schizophrenia more than the average effects of all SNPs (24%).

Conclusions: Schizophrenia is a group of heritable disorders caused by a moderate number of separate genotypic networks associated with several distinct clinical syndromes.

Am J Psychiatry 2015; 172:139–153; doi: 10.1176/appi.ajp.2014.14040435

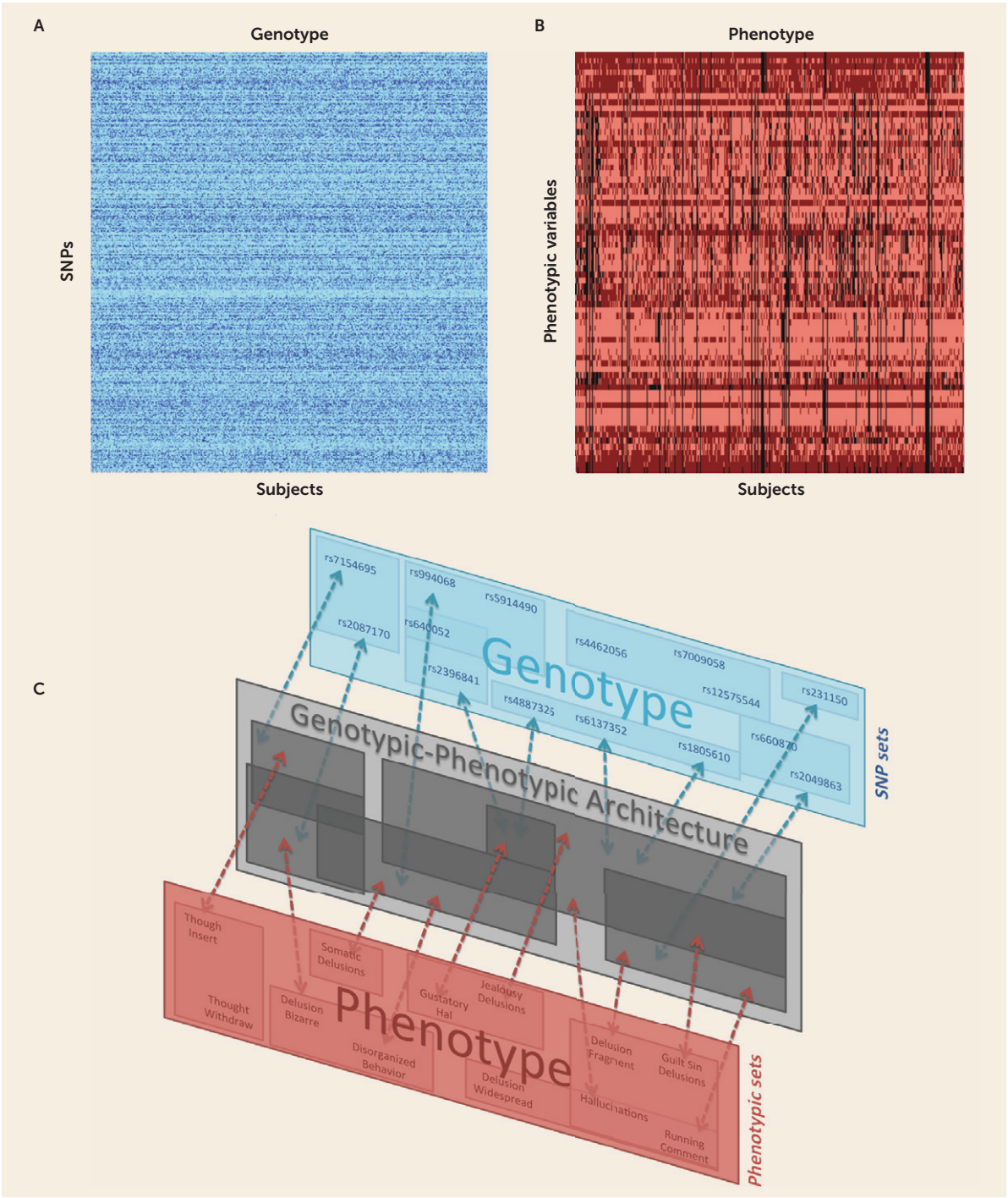
Complex diseases, such as schizophrenia, may be influenced by hundreds or thousands of genetic variants that interact with one another in complex ways, and consequently display a multifaceted genetic architecture (1). The genetic architecture of heritable diseases refers to the number, frequency, and effect sizes of genetic risk alleles and the way they are organized into genotypic networks (2). In complex disorders, the same genotypic networks may lead to different clinical outcomes (a concept known as multifinality, which is called pleiotropy in genetics), and different genotypic networks may lead to the same clinical outcome (equifinality, which is also described as heterogeneity) (1, 3). In general, geneticists must expect the likelihood that many genes affect each trait and each gene affects many traits (4). Consequently, research on complex heritable disorders like schizophrenia is likely to yield weak and inconsistent results unless the complexity of their genetic and phenotypic architecture is taken into account (5).

For example, twin and family studies of schizophrenia consistently indicate that the variability in risk of disease is highly heritable (81%) (6, 7), but only 25% of the variability has been explained by specific genetic variants identified in genome-wide association studies (GWAS) (8). This is not surprising for complex disorders like schizophrenia because current GWAS methods have been unable to characterize the gene-gene interactions (Figure 1A) that influence the developing clinical profiles (Figure 1B) in complex ways (10). The frequent failure to account for most of the heritability of complex disorders has been called the "missing" (11) or "hidden" (12) heritability problem.

In past studies of schizophrenia, the missing heritability problem has been approached by analyzing the explained variance in large individual samples or by using meta-analysis to combine data sets (9, 13, 14). Efforts have also been made to consider the impact of variation related to ethnicity, sex,

This article is featured in this month's **AJP Audio** *and is discussed in an* **Editorial** *by Dr. Jablensky (p. 105)*

FIGURE 1. Perception and Visualization of a Genome-Wide Association Study (GWAS)[a]



[a] Panel A is a matrix corresponding to the genome-wide association data set utilized in this work: Genetic Association Information Network (GAIN) and non-GAIN schizophrenia samples of the Molecular Genetics of Schizophrenia study (9). Allele values are indicated as BB (dark blue), AB (intermediate blue), AA (light blue), and missing (black). Panel B is a matrix corresponding to the distinct phenotypic consequences using data at the symptom level from the Diagnostic Interview for Genetic Studies corresponding to the GWAS in panel A (see Appendix I, catalog of phenotypic features, and Figures S1 and S2 in the online data supplement). Values are indicated as present (garnet), absent (salmon), and missing (black). Panel C presents schematics of the "divide and conquer" approach, in which natural partitions of GWAS data (identified as sets of interacting single-nucleotide polymorphisms [SNPs] or SNP sets) were cross-matched with decomposed schizophrenia phenotype (identified as clusters of naturally occurring schizophrenia symptoms or phenotypic sets), revealing a specific and distributed genotypic-phenotypic architecture (networks of SNPs associated with sets of schizophrenia symptoms). This complex architecture is "invisible" to traditional GWAS.

chromosomes, functional observations, or allele frequency (8). Nevertheless, most of the heritability of schizophrenia remains unexplained (8).

We have chosen to measure and characterize the complexity of both the genotypic and the phenotypic architecture of schizophrenia (Figure 1C). Past studies have generally ignored variation in clinical features, categorizing people as either having or not having schizophrenia, and they have looked only at the average effects of genetic variants, ignoring their organization into interactive genotypic networks. We postulate that schizophrenia heritability is not missing but is distributed into different networks of interacting genes that influence different people (15–17). Unlike previous studies that neglected clinical heterogeneity among subjects with schizophrenia (14, 18, 19), we characterized the clinical phenotype in detail. We also allowed for possible developmental complexity, including equifinality (or heterogeneity) and multifinality (or pleiotropy).

We investigated the architecture of schizophrenia in the Molecular Genetics of Schizophrenia (MGS) study, in which all subjects had consistent and detailed genotypic and phenotypic assessments (9). We then replicated the results in two other independent samples in which comparable genotypic and phenotypic features were available: the Clinical Antipsychotic Trial of Intervention Effectiveness (CATIE) and the Portuguese Island studies from the Psychiatric Genomics Consortium (PGC) (19–23).

METHOD

We first identified sets of interacting single-nucleotide polymorphisms (SNPs) that cluster within subgroups of individuals (SNP sets) regardless of clinical status in the MGS Consortium study, employing our generalized factorization method (24–27) combined with non-negative matrix factorization to identify candidates for functional clusters (17) (see Figures S1 and S2 in the data supplement that accompanies the online edition of this article). This approach performs an unsupervised co-clustering of subjects together with distinguishing genotypic/phenotypic features based on the empirical data alone. We combined the Genetic Association Information Network (GAIN) and non-GAIN samples of the MGS study, which constitute one GWAS (9). The 4,196 cases and 3,827 controls in the MGS study were combined to identify SNP sets. We had data of good quality on 696,788 SNPs on these cases and controls, and from these we preselected 2,891 SNPs that had at least a loose association (p values $<1.0\times10^{-2}$) with a global phenotype of schizophrenia (see the data supplement). SNP sets were labeled by a pair of numbers based on the order in which they were chosen by the algorithm (see the data supplement). Each SNP set was composed of a particular group of subjects described by a particular set of homozygotic and/or heterozygotic alleles; subjects and/or SNPs may be present in more than one set (17, 24, 25). The SNP sets identified by our generalized factorization method are optimal clusters of SNPs in particular subjects that encode AND/OR interactions between SNPs and subjects (Figure 2A–F, Table 1; see also Figure S3 and the Method section in the data supplement). These SNP sets and their relations with one another characterize the genetic architecture of schizophrenia-associated SNPs in all subjects, including cases and controls (Figure 1A).

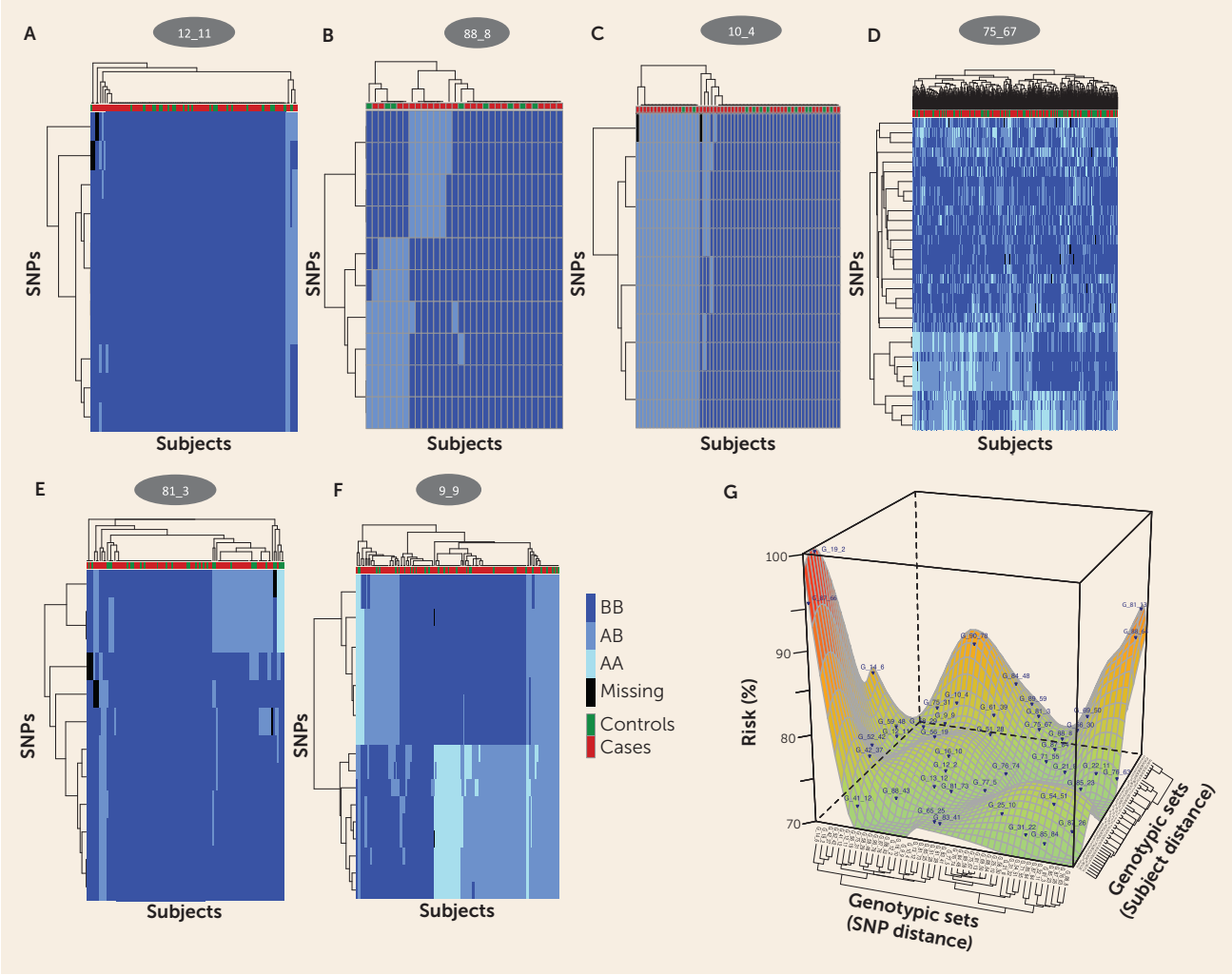
Second, we examined the risk of schizophrenia for each SNP set and identified those with high risk. The statistical significance of the association of SNP sets with schizophrenia was calculated using the SNP-Set Kernel Association Test (SKAT) program, which properly accounts for multiple comparisons (15–17).

Third, we checked for significant overlap among SNP sets in terms of subjects and/or SNPs using hypergeometric statistics (24, 25, 28) (see Figures S1 and S2 in the online data supplement). This allowed us to characterize the relations among SNP sets and to identify SNP sets that were connected to each other by having certain SNPs or subjects in common, thereby composing genotypic networks. Disjoint networks shared neither SNPs nor subjects, as expected if schizophrenia is a heterogeneous group of diseases.

Fourth, we identified sets of distinct clinical features that cluster in particular cases with schizophrenia (i.e., phenotypic sets or clinical syndromes) without regard for their genetic background (29), again using non-negative matrix factorization (17). Ninety-three clinical features of schizophrenia from interviews based on the Diagnostic Interview for Genetic Studies (30), as well as the Best Estimate Diagnosis Code Sheet submitted by GAIN/non-GAIN to dbGaP, were initially considered with the MGS sample (see references 31, 32; see also Appendix I in the online data supplement). The Diagnostic Interview for Genetic Studies was utilized for the Portuguese Island samples. Corresponding features were extracted in CATIE from the Positive and Negative Syndrome Scale, the Quality of Life Questionnaire, and the Structured Clinical Interview for DSM-IV (23). These phenotypic sets and their relations with one another characterize the phenotypic architecture of schizophrenia (Figure 1B).

Fifth, we tested whether SNP sets were associated with distinct phenotypic sets in the MGS sample, and we tested the replicability of these relations in the two other independent studies. Replication was evaluated in terms of replication of the SNP sets and their corresponding risk, as well as the relationships between SNP sets and phenotypic sets. In the samples that used the Diagnostic Interview for Genetic Studies (the MGS and Portuguese Island samples), the specific phenotypic features can be compared. Since the CATIE study did not use the Diagnostic Interview for Genetic Studies, we estimated the corresponding symptoms from available phenotypic data (based on the Positive and Negative Syndrome Scale, the Quality of Life Questionnaire, and the Structured Clinical Interview for DSM-IV). Genotypic and phenotypic data were available for 738

FIGURE 2. Examples of Identified Single-Nucleotide Polymorphism (SNP) Sets Represented as Heat Map Submatrices and their Corresponding Risk[a]



[a] Allele values are indicated as BB (dark blue), AB (intermediate blue), AA (light blue), and missing (black). Subject status (i.e., cases and controls) was superimposed after SNP set identification: cases in red and controls in green. Genotypic SNP sets are labeled by a pair of numbers representing the maximum number of clusters and the order in which they were selected by the method. All SNP sets are calculated with the generalized factorization method based on the non-negative matrix factorization method (see the Method section in the online data supplement). Dendrograms were artificially superimposed for visualization purposes. (See Figure S3 in the data supplement for all SNP sets at more than 70% of risk.) Panels A–F illustrate SNP sets, representing submatrices of the original genome-wide association study matrix and composed of shared SNPs and/or subjects. Panel A presents a SNP set exhibiting a homogeneous configuration in which all subjects in that group share the same interaction among a specific set of homozygotic alleles (i.e., SNP × … × SNP interactions). Panel B presents a SNP set encoding subjects exhibiting a particular heterozygotic genotype with respect to the A allele in a subset of SNPs and another heterozygote genotype with respect to the B allele in a different subset of SNPs (i.e., AND-type of interactions). Panel C presents a SNP set composed of subjects who share a particular genotype value for a subset of SNPs, and another subset of subjects sharing a different genotype value for the same subset of SNPs (i.e., OR-type of interactions). Inclusion-type relations are exemplified by a SNP set (panel A) subsumed under a more general SNP set (panel C), and both sets provide different descriptions of target subjects. Panels D–F present SNP sets that combine all previous interactions into more complex structures. Panel G presents a surface representing the risk function of the uncovered SNP sets. The risk (z-axis; red=high, blue=low) was calculated based on the distribution subject status (i.e., cases and controls) within each SNP set, and the surface was plotted interpolating the relation domains. Dendrograms reflect the order adopted for plotting SNP sets. SNP sets were clustered by shared SNP (x-axis) and by shared subjects (y-axis) using hypergeometric statistics (see the Method section in the data supplement). (Close-located SNP sets in an edge share more SNPs and/or subjects than those located far away.)

cases in CATIE and 346 cases in the Portuguese Island study (see the online data supplement). The significance of cohesive relations among SNP sets and clinical syndromes was tested using hypergeometric statistics (17, 24, 25, 28). The relations between the genotypic and phenotypic clusters characterize the genotypic-phenotypic architecture (Figure 1C).

Methodological details and references for the "divide and conquer" algorithm that we developed and used are available in the online data supplement (24–27). Our web server application PGMRA (17), for identifying genotype-phenotype relations in GWAS, is online at http://phop.ugr.es/fenogeno. Statistical analysis was performed by SKAT (15, 16), also accessible through PGMRA.

TABLE 1. Single-Nucleotide Polymorphism (SNP) Sets Reported With ≥70% Risk of Schizophrenia, Statistical Comparison With Individual SNPs, and Composition[a]

	SKAT p Values						
SNP set	Group	Average SNP	Best SNP	Worst SNP	Subjects (N)	SNPs (N)	Risk (%)
19_2	2.88E−05	3.43E−02	4.60E−04	1.38E−02	9	9	100
88_64	1.43E−11	2.06E−03	2.15E−07	1.79E−02	176	6	96
81_13	1.46E−10	5.44E−03	2.15E−07	3.70E−02	234	10	95
87_76	7.11E−07	1.05E−02	1.37E−05	3.13E−02	74	3	95
58_29	5.41E−04	6.52E−03	2.07E−04	2.83E−02	125	6	94
83_41	3.87E−05	1.56E−04	1.01E−04	2.68E−04	61	4	93
9_9	1.51E−06	2.52E−03	1.23E−04	1.18E−02	144	19	92
10_4	3.83E−05	1.72E−02	2.11E−04	1.05E−02	58	11	91
14_6	2.38E−06	1.85E−03	1.23E−04	5.87E−03	22	11	90
56_30	1.91E−10	4.33E−03	2.15E−07	2.10E−02	382	11	88
42_37	4.15E−06	2.35E−02	6.59E−05	1.38E−02	70	24	86
65_25	3.95E−05	1.99E−02	2.53E−04	8.83E−02	62	5	86
71_55	1.90E−05	3.99E−04	2.63E−05	1.08E−03	63	6	86
12_11	6.53E−04	2.28E−02	7.34E−03	1.05E−01	94	11	84
90_78	7.87E−04	2.99E−02	3.58E−02	9.53E−02	200	4	83
77_5	4.86E−05	5.01E−04	2.08E−05	1.49E−03	297	5	82
88_8	2.88E−04	2.95E−02	3.58E−02	8.36E−02	32	10	82
51_28	2.07E−04	2.25E−02	1.75E−02	3.13E−02	258	3	81
59_48	2.32E−09	9.48E−03	2.38E−05	2.96E−02	174	7	80
41_12	1.36E−03	1.62E−02	1.12E−01	2.17E−02	78	3	76
22_11	6.24E−05	4.29E−04	1.33E−04	1.08E−03	97	12	75
13_12	4.52E−05	3.61E−04	5.88E−05	1.45E−03	148	10	75
31_22	1.01E−04	2.37E−04	1.11E−04	4.03E−04	92	7	74
85_84	1.53E−05	1.01E−04	1.37E−05	1.81E−04	39	4	74
87_84	1.19E−04	1.40E−02	1.37E−05	1.30E−02	22	13	74
16_10	1.81E−03	1.59E−02	2.92E−03	5.92E−02	141	12	73
56_19	2.02E−04	6.69E−04	1.02E−04	1.76E−03	90	5	73
75_31	2.61E−05	1.37E−02	1.02E−04	9.53E−02	197	8	73
81_73	1.13E−05	2.99E−02	2.57E−04	1.29E−02	213	10	73
85_23	6.20E−03	9.46E−03	5.58E−03	1.16E−02	53	4	73
21_8	6.24E−05	4.29E−04	1.33E−04	1.08E−03	188	12	71
76_74	1.58E−17	1.33E−02	1.12E−05	1.17E−02	284	14	71
61_39	1.04E−03	2.43E−02	1.90E−03	5.45E−02	51	3	71
75_67	3.76E−18	7.16E−02	2.15E−07	1.00E−03	877	32	71
76_63	2.07E−02	2.25E−02	1.75E−02	3.13E−02	34	3	71
81_3	6.24E−05	4.29E−04	1.33E−04	1.08E−03	107	12	71
87_26	2.49E−03	6.03E−03	4.14E−03	1.12E−02	28	5	71
88_43	1.37E−04	1.85E−03	6.03E−04	4.82E−03	70	7	71
25_10	3.49E−06	1.67E−03	1.11E−04	1.53E−02	124	9	70
12_2	1.81E−03	1.59E−02	2.92E−04	5.92E−02	194	12	70
52_42	5.70E−05	5.06E−03	6.59E−05	3.60E−02	87	16	70
54_51	1.49E−05	5.01E−04	2.08E−04	1.49E−03	132	5	70

[a] SKAT=SNP-Set Kernel Association Test.

RESULTS

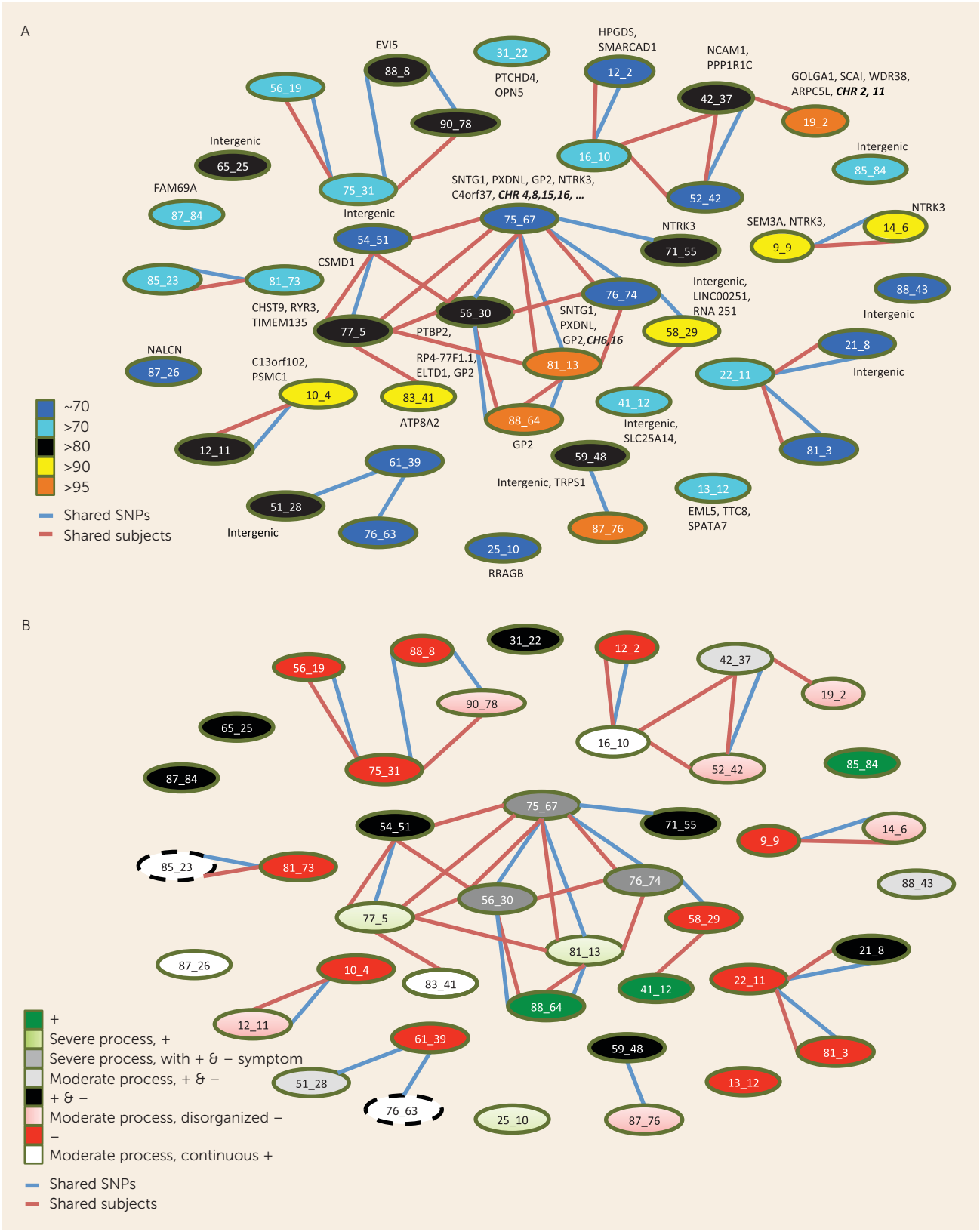
Identifying Many SNP Sets as Candidates for Schizophrenia Risk

We first investigated the genotypic architecture of schizophrenia in the MGS study to identify SNP sets without knowledge of the subject's clinical status (i.e., case or control) (9). Our exhaustive search uncovered 723 nonidentical and possibly overlapping SNP sets in the MGS samples. The SNP sets varied in terms of numbers of both subjects and SNPs. For example, one group contains 70 subjects and 24 SNPs, as expected because few subjects can share a large number of SNPs. Conversely, another group contains 258 subjects and three SNPs, as expected because a large number of subjects are likely to share only a few SNPs. Initially, we retained a large number of SNP sets merely to identify the genotypic clusters in all subjects whether they had schizophrenia or not.

SNP Sets Vary Greatly in Risk for Schizophrenia

Second, we computed the risk for schizophrenia in carriers of each SNP set (33) (Figure 2A–F; see also Figure S3 in the online data supplement). The risk of schizophrenia was normally distributed, as expected when capturing the full range of variability. Ninety-eight of the 723 SNP sets had a risk of schizophrenia greater than 66% and accounted for 90% of schizophrenia cases in the MGS study. Forty-two SNP sets had a risk of schizophrenia ≥70% (Table 1; see also Figure S4 in the data supplement). For example, SNP set 19_2 had a risk of

FIGURE 3. Dissection of a Genome-Wide Association Study (GWAS) and Identification of the Genotypic and Phenotypic Architecture of Schizophrenia[a]



100%, meaning that all carriers were schizophrenia cases. The ability of SNP sets to predict schizophrenia risk is illustrated in Figure 2G. SKAT showed that the association of schizophrenia with particular SNP sets was stronger than with the average effects of their constituent SNPs (Table 1). For example, the SNP set 81_13 has a p value of 1.46×10^{-10}, whereas the best and average SNPs within this set have p values of 2.15×10^{-7} and 5.44×10^{-3}, respectively. SKAT and PLINK (34) methods estimated similar p values for the individual SNPs (R^2=0.99; p values for F statistics, $<3.8\times10^{-46}$), showing that SKAT does not inflate results.

The global variance in liability to schizophrenia explained by the average effects of all SNPs simultaneously (8, 35) in our sample was 24%. While individual SNPs were mostly low penetrant, many high-risk SNP sets were highly penetrant (e.g., 100% to 70%; see Table 1) and much more informative in predicting schizophrenia risk.

Relations Among SNP Sets to One Another and to Gene Products

We hypothesized that schizophrenia may be an etiologically heterogeneous group of illnesses in which some genotypic networks are disjoint, that is, share neither SNPs nor subjects. To test this, we first checked for overlap in constituent SNPs and/or subjects among all the SNP sets at high risk for schizophrenia (see Figure S5 in the online data supplement). We found that 17 genotypic networks were disjoint, sharing neither SNPs nor subjects (Figure 3A), suggesting that these are distinct antecedents of schizophrenia. These networks vary in size and complexity: one highly connected network associates 11 SNP sets, whereas eight networks are composed of only a single isolated SNP set.

We also determined that some SNP sets share SNPs but not subjects (e.g., 59_48 and 87_76; Figure 3A), as expected because they involve the same SNPs but with different allele values (both alleles of a SNP can act as risk alleles in different genetic contexts). In contrast, we found that the 58_29 and 41_12 SNP sets do not share SNPs, but independently specify almost the same individuals (Figure 3A), as expected when, for example, distinct subsets of genotypic features influence a common developmental pathway. Finally, some SNP sets overlap in both SNPs and subjects, suggesting that one is a subset within the other (e.g., 88_64 and 81_13; see Figure S3A,C in the online data supplement). Therefore, the genotypic networks display distinct topologies differing in the way constituent SNPs and subjects are related.

When evaluating whether different genotypic networks operate through distinct mechanisms, we found that high-risk SNP sets mapped to various classes of genes (e.g., protein coding, ncRNA genes, and pseudogenes) related to known functions and causing different effects on their products (Figure 3A; see also Tables S1–S3 and Figure S6 in the online data supplement). We identified distinct pathways as exemplified in Table 2. Notably, all of these pathways are interconnected by the overlapping gene products that include genes previously associated with schizophrenia by GWAS, as well as genes known to be abnormally expressed in the brains of schizophrenia patients (see Table S4, Figure S7, and the Pathways section in the data supplement). The emerging picture is suggestive of a possible pathophysiology in which abnormal brain development interacts with environmental events triggering abnormal or exaggerated immune and oxidative processes that increase risk of schizophrenia.

Complex Genotypic-Phenotypic Relationships in Schizophrenia

Next we examined whether the complex genetic architecture of schizophrenia leads to phenotypic heterogeneity. Using data from the Diagnostic Interview for Genetic Studies (30), as well as from the Best Estimate Diagnosis Code Sheet submitted by GAIN/non-GAIN to dbGaP (see Appendix I, Figures S1 and S2, and the Method section in the online data supplement), we originally identified 342 nonidentical and possibly overlapping phenotypic sets of distinct clinical features that cluster in particular cases with schizophrenia (i.e., phenotypic sets or clinical syndromes) without regard for their genetic background. Different SNP sets were significantly associated with particular clinical syndromes (hypergeometric statistics, p values from 2×10^{-13} to 1×10^{-3}). However, the genotypic-phenotypic relations were complex (i.e., many-to-many [29]): the same genotypic network could be associated with multiple clinical outcomes (i.e., multifinality or pleiotropy) and different genotypic networks could lead to the same clinical outcome (i.e., equifinality or heterogeneity; Table 3; see also Table S5 in the data supplement). The genotypic-phenotypic relations were highly significant

[a] Panel A presents a genotypic network, in which nodes indicate SNP sets linked by shared SNPs (blue lines) and/or subjects (red lines). The risk value, which was incorporated after the SNP set identification, was color-coded. The 42 SNP sets harboring ≥70% of risk were topologically organized into 17 disjoint subnetworks. Subsets of implicated genes are indicated. Highly connected SNP sets based on shared SNPs (blue lines) and subjects (red lines) might share a phenotypic profile (e.g., 81_13 and 88_64; see Table 3). Yet a super-SNP set, such as 81_13, may have unique—in addition to common—descriptive phenotypic features (see Table 3). Disconnected SNP sets, such as 71_55 and 14_6, belong to disjoint networks that may include the same gene (i.e., NTKR3; see Table S1 and Figure S6B in the online data supplement) but carry SNPs that are located in the promoter and coding region, respectively. Both SNPs may produce distinct molecular consequences (see Table S3 and Figure S6B in the data supplement) and phenotypic profiles (see Table 3). Panel B shows the classes of schizophrenia mapped to the disease architecture (see Table 3). Eight classes of schizophrenia were identified by independently characterizing each phenotypic feature included in a genotypic-phenotypic relationship; classifying each item based on the symptoms as purely positive, purely negative, primarily positive, or primarily negative symptoms; and clustering these relationships based on their recoded phenotypic domain using non-negative matrix factorization. SNP sets harboring only positive symptoms are indicated in green, whereas those displaying negative symptoms are in red. Intermediate combinations including severe and/or moderate processes combined with positive and/or negative and/or disorganized symptoms were also color-coded. Dashed lines indicate nonsignificant matching.

TABLE 2. Examples of Products of Genes Uncovered by the SNP Sets in Interconnected Signaling Pathways[a]

Signaling Pathways/ Function	Genes	SNP Sets	Symptoms
Neural development	DKK4, STKY1, VANGL1	75_67	Severe process, + and −
	NCAM1	42_37	Moderate process, + and −
		52_42	Moderate process, −
	CHST9	81_73	−
	EML5	13_12	−
	SEM3A	9_9	Moderate process, −
Neurotrophin function	NTRK3	75_67	Severe process, + and −
	Upstream region	71_55	+ and −
	SNTG1	81_13	Severe process, +
	MAGEH1	25_10	Severe process, +
Neurotransmission	NETO2	76_74, 75_67	Severe process, with + and −
	OPN5	31_22	+
	NALCN	87_26	Moderate process, continuous +
Neuronal function and neurodegenerative disorders	SPATA7, ZC3H14	13_12	−
	SLC20A2	41_12	+

[a] The 42 SNP sets at high risk for schizophrenia involved at least 96 gene loci, including 54 protein-coding loci and 42 polymorphisms at regulatory sites, as well as 112 polymorphisms in either intergenic or unannotated regions (see full Tables S1 and S4 and Figure S7 in the online data supplement).

by a permutation test (empirical p value $<4.7\times10^{-3}$; Table 3; see also Table S5).

Specifically, we identified a phenotypic set indicating a general process of severe deterioration (i.e., continuous positive symptoms with marked and progressive impairment) that was associated with many SNP sets (e.g., SNP sets 75_67 and 56_30, with p values $<2.3\times10^{-13}$ and 2.55×10^{-5}, respectively; Table 3, Figure 3A). Other SNP sets were associated with a general process of moderate deterioration (moderate or fluctuating impairment despite a continuous mixture of symptoms), as in SNP sets 14_6, and 42_37 (p values $<5\times10^{-4}$; Table 3, Figure 3A).

We identified specific clinical syndromes that were unambiguously associated with particular genotypic networks. For example, specific phenotypic sets differentiate among SNP sets even within the same network, which illustrate similar but not identical forms of multifinality in schizophrenia (e.g., 76_74 and 58_29; Table 3, Figure 3A, blue lines). Particular phenotype sets can also distinguish SNP sets connected only by shared subjects (Figure 3A, red lines). For example, SNP set 76_74 shares subjects with 56_30 and with 81_13; however, the latter SNP sets are associated with a specific phenotypic set not present in 76_74 (Table 3).

Positive and Negative Symptoms Differentiate Classes of Schizophrenia

Genotypic and phenotypic relationships could be grouped into eight classes of schizophrenia, as shown in Figure 3B and Table 3 (31, 32, 36). First, we identified SNP sets involving subjects with predominantly positive symptoms (e.g., 41_12 and 88_64) and few residual symptoms. Second, we identified SNP sets represented by predominantly negative and disorganized symptoms (e.g., 10_4 and 61_39), decreased psychosocial function, and continuous residual symptoms. As discussed in the online data supplement (see the Replicability of the Phenotypic Features section), bizarre delusions and symptoms of cognitive and behavioral disorganization, such as thought insertion and disorganized speech among others, were accepted as fuzzy indicators of either positive or negative classes of schizophrenia but were considered to be more common in negative and disorganized classes (e.g., in Table 3, thought echo and commenting hallucinations in "negative schizophrenia" with phenotypic set 46_29 associated with SNP set 14_6).

Third, several SNP sets harbor mixed positive and negative symptoms (e.g., 59_48 and 54_51). These three classes were enriched by considering the general severe and moderate patterns, which were frequent in several networks (Figure 3B), as described above. Because the latter patterns appear in combination with a set of only positive symptoms (e.g., 81_13), both positive and negative symptoms (e.g., 75_67), and only negative symptoms (e.g., 19_2), we were able to classify schizophrenia into eight classes (Figure 3B). A principal-components analysis of the phenotypic features in the Diagnostic Interview for Genetic Studies confirmed this classification (see Table S6 and the Method section in the online data supplement).

Replication of Results in Two Independent Samples

We tested the replicability of our findings in the MGS study by carrying out the same analyses of the genotypic and phenotypic architecture of schizophrenia in the CATIE (19, 22, 23) and Portuguese Island (19, 21) samples. A total of 1,303 SNPs were shared between the selected SNPs in the MGS (see the Data Cleaning section in the online data supplement) and CATIE samples, and 1,234 SNPs between the MGS and Portuguese Island samples. Imputed variants were not considered, to avoid possible biases.

We found that 31 and 30 of the 42 SNP sets selected in the MGS sample were also identified in the CATIE and Portuguese Island samples, respectively (see Tables S7 and S8 in the online data supplement). Together, both samples reproduced at least 81% of the SNP sets at risk (see Table S9 in the data supplement). In addition, most of the SNP sets replicated in the two PGC samples achieved risk values as high as those of the MGS sample (>70%) (see Table S8): 70% of those identified exhibit >70% risk, and 90% show >60% risk. Some SNP sets exhibited slightly higher risk values than those in the MGS sample.

The genotypic-phenotypic relations in CATIE and the Portuguese Island studies closely matched those observed in

the MGS study (hypergeometric statistics, p values 1×10^{-7} to 1×10^{-2}; see Tables S7 and S8 and the Replicability section in the data supplement). The eight schizophrenia classes exhibited high reproducibility. For example, except for one relation ("–" in the MGS study and "+ and –" in CATIE; see Table S9 in the data supplement), all relations exhibited similar positive and negative symptoms in the MGS study and CATIE. Three relations showed less specific symptoms in CATIE than in the MGS study, as expected because CATIE did not use the Diagnostic Interview for Genetic Studies (see Table S10 and the Replicability section in the data supplement).

We found few differences when comparing the MGS and Portuguese Island studies (see Table S9 in the data supplement), except differences in severity that preserved the sign of the symptoms. Three relations with negative symptoms in the MGS study exhibited negative and positive symptoms in the Portuguese Island sample (see Table S9). Only two SNP sets in the Portuguese Island sample had no significant cross-match with the phenotypic features expected from the MGS study.

DISCUSSION

Our findings indicate that schizophrenia comprises several distinct clinical syndromes associated with many disjoint genotypic networks. Consequently, much of the heritability of schizophrenia has not been detected by approaches that classify subjects only according to whether or not they have schizophrenia. Our purely data-driven analysis shows that the elusive heritability of schizophrenia is not missing, but is encoded in a complex distribution of genotypic-phenotypic relationships.

We found that 42 interactive SNP sets had greater than 70% risk of schizophrenia. The interactive SNP sets explained the risk more fully than the average effects of all SNPs simultaneously and were more strongly related to their particular syndromes of schizophrenia than are their individual SNPs (Table 1). Consequently, identifying the organization of SNPs into interactive SNP sets enabled us to increase the power to detect associations: 98 SNP sets with greater than 66% risk accounted for 90% of cases. The constituent genes in these networks belong to signaling pathways highly associated with schizophrenia (see Figure S7 in the online data supplement). Our findings have broad implications, so we will consider their strengths and limitations carefully.

Strengths and Limitations

Two particular features of our methods merit consideration in terms of their strengths and limitations. First, we concurrently used detailed assessments of both the genotype and the phenotype to identify their associations, thereby combining genomic and phenomic information (29). Other approaches decrease the number of variables before analysis ("data reduction"), even if the biological importance of these variables is not known a priori. The evidence we have that schizophrenia is a heterogeneous group of disorders suggests that reducing clinical information about schizophrenia to a single categorical diagnosis is inadequate.

Despite the detailed phenotypic information we had available about subjects, there are still limitations to data obtained even from reliable structured interviews like the Diagnostic Interview for Genetic Studies. Interview data are based on self-reports that are interpreted and coded by interviewers. Subjects may not be willing or able to report their symptoms accurately. We had obtained information from treatment records and family history reports, but we chose not to use such additional information, except for the resulting best-estimate final DSM ratings of diagnosis, because its extent and quality varied in unmeasured ways between cases. The greatest limitation in the phenotypic assessments in available GWAS databases has been the overreliance on subjective symptoms with an absence of objective measurements, such as cognitive tests, brain electrophysiology, and neuroimaging (37). Subjective symptoms are fuzzy indicators of the underlying pathophysiology. Objective measures could complement the assessment of symptoms and could be applied to both cases and controls, thereby providing a more comprehensive and valid characterization of the phenotype of all subjects. The biggest challenge in GWAS is access to studies with rich phenotypic data about both subjective and objective measures obtained systematically from all subjects.

Our finding of robust replicability based on detailed symptom profiles from interview data alone do have important implications for the size of samples and the scope of phenotypic assessments in genomic studies of complex disorders. We obtained robust replication of results in moderate size samples, such as 738 cases in CATIE and 346 cases in the Portuguese Island study, which shows that it is incorrect to assume that extremely large samples are needed to obtain robust and replicable findings. Difficulty in replication in previous work can be better explained by the neglect of the complexity of genetic and phenotypic architecture rather than by moderate sample size. We identified more information by combining rich data about the complex architecture of genotypes and interview-based symptoms in such moderate size samples than has been obtained in analysis of much larger compilations of multiple samples that relied on additive gene effects on categorical diagnosis (8).

Nearly all previous genetic association studies have relied on patient interviews for clinical description, as detailed objective testing has been impractical in large samples because of cost and the difficult logistics of securing cooperation with time-consuming test batteries. Now that we have shown that replicable results can be obtained in moderate size samples, it is feasible to complement interview data with more objective and thorough assessments. Fundamental research into the causes and characteristics of the schizophrenias is likely to require phenotypic assessment beyond the clinical features needed for clinical diagnosis and treatment, which has given primacy to signs and symptoms assessable by interview alone (37).

Second, we have strived to extract the maximum information available in a single GWAS without making restrictive

TABLE 3. Subset of Genotypic-Phenotypic AND/OR Relationships (Hypergeometric Statistics)[a]

Schizophrenia Class, Symptoms[b], and DSM Ratings	Phenotypic Sets	SNP Sets	p
Severe process, with positive and negative symptom schizophrenia			
Positive symptoms; moderate severity of impairment; unable to function since onset	15_13	56_30	2.55E–05
Auditory hallucinations (2 or more voices; running commentaries)	12_11		1.79E–04
Auditory hallucinations (2 or more voices; running commentaries); thought echoing; withdrawal; insertion and broadcasting; delusions of mind reading	21_1		3.66E–04
Hallucinations (any); auditory hallucinations (ever; 2 or more voices); grossly disorganized behavior	50_46		5.70E–04
Hallucinations (mood incongruent); auditory hallucinations; somatic hallucinations (olfactory; gustatory; tactile); religious delusions; delusions of mind reading; delusions of control; thought echoing; withdrawal; insertion and broadcasting	9_6		4.45E–03
Hallucinations (mood incongruent); persecutory delusions; delusions of reference; jealousy delusions; bizarre delusions; disorganized odd behavior; disorganized odd speech; delusions, fragmented (unrelated themes); delusions, widespread (intrude into most aspects of life); thought insertion; flat affect; avolition and apathy	46_23		4.15E–03
Continuously positive symptoms; severe impairment; continuous course; no affective symptoms	15_13	75_67	2.31E–13
Grossly disorganized behavior; severe impairment; continuous course	54_11		4.90E–06
Delusions of persecution and reference; disorganized speech; severe impairment; unable to function since onset	30_17		2.56E–04
Auditory hallucinations (ever; 2 or more voices; running commentaries); jealousy delusions	18_13		3.50E–04
Thought insertion and withdrawal	27_6		3.62E–03
Hallucinations (any); auditory hallucinations (2 or more voices); grossly disorganized behavior	50_46		3.61E–03
Delusions, persecutory and reference; delusions, widespread (intrude into most aspects of life)	61_18		4.28E–03
Disorganized; odd speech	64_11		1.45E–03
Delusions, widespread (intrude into most aspects of life); continuous course	65_64		1.21E–03
Continuously positive symptoms; severe impairment; unable to function since onset; no affective symptoms	15_13	76_74	1.07E–07
Delusions, widespread (intrude into most aspects of life)	65_64		1.47E–03
Positive and negative schizophrenia			
Auditory hallucinations; delusions (any); bizarre delusions; disorganized speech and behavior; flat affect; alogia; avolition	12_4	59_48	1.88E–04
Auditory hallucinations (2 or more voices; running commentaries)	42_9	71_55	1.98E–03
Negative schizophrenia			
Thought insertion and withdrawal	52_28	58_29	1.44E–04
Disorganized speech; odd speech	7_3	9_9	1.97E–04
Flat affect; persecutory delusions	48_41		2.23E–03
Delusions of mind reading; guilt delusions; sin delusions; jealousy delusions	26_8		4.20E–03
Flat affect; apathy; avolition	69_41	22_11	5.52E–05
Flat affect; apathy; avolition; alogia; continuous mixture of positive and negative symptoms	10_5		4.62E–04
Disorganized and odd speech	17_2		1.01E–04
Positive schizophrenia			
Hallucinations (any); auditory hallucinations (ever; 2 or more voices); no affective symptoms	63_24	88_64	3.45E–04
Delusions of jealousy; auditory hallucinations (running commentaries)	69_66		4.49E–03
Severe process, positive schizophrenia			
Continuously positive symptoms; severe impairment; unable to function since onset; no affective symptoms	22_13	77_5	5.66E–05
Auditory hallucinations (2 or more voices; running commentaries)	18_13		3.25E–03
Hallucinations (any); auditory hallucinations (2 or more voices; running commentaries); continuous course	53_6		4.76E–03

continued

TABLE 3, *continued*

Schizophrenia Class, Symptoms[b], and DSM Ratings	Phenotypic Sets	SNP Sets	p
Severe process, positive schizophrenia			
Auditory hallucinations (ever; voices; noises; music)	59_41		1.22E−03
Continuously positive symptoms; severe impairment; unable to function since onset; no affective symptoms	20_19	81_13	2.83E−04
Hallucinations (any); auditory hallucinations (ever; 2 or more voices); bizarre delusions; delusions, fragmented (unrelated themes); delusions, widespread (intrude into most aspects of life)	55_7		8.57E−04
Delusions of reference; delusions of persecution	34_17		2.40E−03
Auditory hallucinations (running commentaries); jealousy delusions	69_66		1.30E−03
Severe impairment; unable to function since onset; no affective symptoms	27_7	25_10	4.76E−06
Auditory hallucinations (2 or more voices; running commentaries)	18_13		9.50E−05
Auditory hallucinations (ever; voices; noises; music); auditory hallucinations (2 or more voices; running commentaries); thought echoing	4_1		2.49E−03
Delusions of reference; delusions of persecution	66_54		2.10E−03
Bizarre delusions; delusions of mind reading; delusions, widespread (intrude into most aspects of life)	8_4		1.93E−03
Moderate process, disorganized negative schizophrenia			
Grossly disorganized or catatonic behavior; disorganized speech	51_38	19_2	4.03E−04
Moderate deterioration; unable to function since onset; no affective symptoms	42_7	14_6	4.96E−04
Grossly disorganized and inappropriate behavior	18_3		2.55E−03
Auditory hallucinations (running commentaries); thought echoing	46_29		3.78E−03
Moderate process, positive and negative schizophrenia			
Hallucinations (any); auditory hallucinations (ever; voices; noises; music); continuous mixture of positive and negative symptoms; continuous course; moderate impairment; unable to function since onset; no affective symptoms	5_2	42_37	1.32E−04
Bizarre delusions; delusions of reference	57_39		4.70E−03
Continuous mixture of positive and negative symptoms; continuous course; moderate impairment; unable to function since onset; no affective symptoms	11_5	88_43	6.88E−04
Auditory hallucinations (ever); bizarre delusions; delusions, fragmented (unrelated to theme)	24_4	51_28	9.58E−04
Moderate process, continuous positive schizophrenia			
No affective symptoms	48_7	16_10	1.44E−03
Continuously positive symptoms; severe impairment; unable to function since onset; no affective symptoms	28_23	83_41	3.48E−03
Continuously positive symptoms; no affective symptoms	25_20	87_26	4.22E−03

[a] See Appendix I and full Table S5 in the online data supplement.
[b] Symptoms were assessed with the Diagnostic Interview for Genetic Studies.

a priori assumptions. In other words, we proceeded in a data-driven, model-free manner. As a consequence, whatever information emerged from the data mining process (such as the different classes of schizophrenia) is inherent to the data and was not artificially imposed by either an a priori model or previous knowledge of the data (such as the "case or control" status of the subjects). Nevertheless, our initial pool of 2,891 SNPs, preselected for at least loose association with schizophrenia in the MGS study (9), might be missing additional risk SNPs that would eventually show up in an even more exhaustive genomic analysis.

Our findings about the heterogeneity and complexity of schizophrenia (31, 32, 36) require a careful reconsideration of the concept of "replicability." In order to be meaningful in complex disorders like schizophrenia, efforts to replicate findings must take into account the distributed heritability and developmental complexity of the disease.

Replication: A Lock and Key Combination of Genomics and Phenomics

Replication is always critical, but it is not usually sought within a single large study. Here, internal replicability was addressed by resampling techniques (94% support; see the online data supplement), where the same SNP sets are systematically identified despite the random alteration of the parameters of the method (17) and/or the sample (38). In addition, our biggest challenge was to identify studies with rich phenotypic data for independent external replication. In most GWAS, phenotypic data have been of "secondary" interest, using a variety of structured or even unstructured interviews (14, 18, 19) (see the Replicability section in the data supplement). So why not attempt to replicate the genotypic architecture alone? The same answer applies for any method for validation of associations: genetic variants associated with individuals may be, and in all likelihood often are, completely

PATIENT PERSPECTIVES

A Patient with Severe Process, Positive Schizophrenia (Associated with SNP Set 81_13 and Phenotypic Sets 20_19 and 34_17)

"Ms. A" was a 23-year-old woman with DSM-IV schizophrenia and no history of substance abuse, depression, or mania. She was born 2 months premature due to maternal preeclampsia. At age 5, she taped the mouths of her dolls to try to stop her hallucinations of their calling her name and whispering to her. At age 7, she developed delusions of persecution and reference (as in phenotypic set 34_17) and the voices became louder. At age 9, she was diagnosed with paranoid schizophrenia and began treatment with antipsychotics. Her delusions about her classmate's harmful intentions provoked fights, so she dropped out of high school. Her delusions became widespread but not bizarre. Her hallucinations and delusions never remitted, and she developed no negative symptoms, disorganized speech, or behavior. She had continuous and progressive deterioration without associated affective symptoms, so that she was unable to work or marry (i.e., severe process as in phenotypic set 20_19). On mental status, she had appropriate behavior, oddly vague speech without loose associations, well-modulated affect, average intelligence, and poor insight and judgment. She felt she was being watched and followed.

Ms. A's clinical profile was specifically associated with the SNP set 81_13, which had a 95% risk of schizophrenia. This SNP set is a marker of a functional complex of several genes that may possibly influence brain function by regulation of neurodevelopment and neuronal cell signaling. For example, the gamma-1-syntrophin (SNTG1) gene encodes a brain-specific protein with two functional domains: one regulates alpha-adrenergic receptor signaling, and the other mediates dystrophin binding. Dystrophin interacts in turn with glycoprotein complexes, and another gene associated with 81_13 is glycoprotein-2 (GP2). GP2 is associated with risk of neuropathies, basal ganglia disorders, and schizophrenia. PDXNL encodes a peroxidsin-like endonuclease that selectively degrades mRNAs, suggesting that the SNP set 81_13 may function normally to maintain healthy neurodevelopment, but is associated with schizophrenia when deficient.

A Patient with Moderate Process, Disorganized Negative Schizophrenia (Associated with SNP Set 19_2 and Phenotypic Set 51_38)

"Mr. B" was a 23-year-old man with DSM-IV schizophrenia. At age 10, he began to collect odd things from the garbage and to speak in a vague, emotionless manner. He became childishly negativistic, obstinate, and isolated. By age 13, his behavior became more inappropriate and disorganized, and his speech was fragmented by frequent derailment (as in phenotypic set 51_38). He never had hallucinations. Occasionally he thought others were against him or making fun of him, but his convictions never lasted more than a few days and were not systematized or bizarre. He was increasingly unmotivated to initiate or persist in goal-directed tasks; he completed high school (with parental supervision) and then enrolled in college, but he soon dropped out. At age 18, he was depressed and used illicit drugs briefly. At age 23, he had been continuously psychotic with moderate deterioration since onset. He was living with his parents, unmarried, unemployed, and considering trying college again. On mental status examinations from ages 16 to 23, he always had flat affect, average intelligence, and poor insight and judgment, which were accompanied by disorganized speech and behavior at times of perceived stress.

The phenotypic set of disorganized speech and behavior was specifically associated with the SNP set 19_2, which carried a 100% risk of schizophrenia. This SNP set is a marker of a functional complex of several genes that act in concert with the gene GOLGA1 in ways that may possibly regulate the development and orchestration of cortico-striatal circuits underlying motivated activity, including speech and emotional expression. GOLGA1 encodes a key protein in the signaling pathways that regulate glycosylation and the transport of proteins and lipids in the Golgi apparatus. GOLGA1 alters splicing and polyadenylation in the cerebral cortex in patients with schizophrenia compared to others. It acts in concert with many other genes related to 19_2. For example, the genes WDR38 and SCAI influence signaling pathways for cell migration and transcriptional regulation in the basal ganglia, which are critical for coordination of speech and emotional expression via the prefrontal-striatal-prefrontal loop. GOLGA1 variation has been associated with schizophrenia, Parkinson's disease, Sjögren's syndrome, and sleep disorders.

A Patient with Severe Process, Positive and Negative Schizophrenia (Associated with SNP set 75_67 and Phenotypic Sets 15_13, 30_17, 61_18, and 65_64)

"Ms. C" was a 35-year-old woman with DSM-IV schizophrenia and no history of substance abuse, depression, or mania. She required an individualized educational program for learning disabilities from age 6 on. At age 17, she began hearing voices that told her people were out to harm her. Her persecutory delusions about classmates led to conflict, so she dropped out of high school. Delusions of persecution and reference became widespread, and interfered with her functioning (as in phenotypic set 61_18). Her delusions were continuous (as in phenotypic set 65_64) but not bizarre or fragmented. She heard multiple voices talking in a chorus to her daily, telling her that people wanted to hurt her. Her hallucinations and

continued

delusions were accompanied by disorganized speech (as in phenotypic set 30_17) and prominent negative symptoms (flat affect, avolition, alogia). She was unable to work or marry and required supervision all her life, deteriorating severely over time (i.e., severe process as in phenotypic set 15_13). On mental status examination, she had childishly rude behavior, flat affect, tangential speech, poverty of abstract thinking, and poor insight and judgment about her illness and behavior.

Ms. C's clinical profile of mixed positive and negative symptoms with severe deterioration was specifically associated with the SNP set 75_67, which carried a 71% risk of schizophrenia. This SNP set is a marker of a functional complex of many genes that act in concert to regulate neurotrophic and neuroimmune functions in response to diverse environmental challenges. The gene for neurotrophin receptor-3 (NTRK3) regulates the production of neurotrophin, which promotes the growth and survival of neurons, protecting them against apoptosis in response to oxidative stress or glutaminergic excitotoxicity. NETO2 modulates the plasticity of glutamate neurotransmission at kainate receptors. GP2 (shared with 81_13) transports antigens across cell membranes and modulates adaptive immune responses. SNTG1 (shared with 81_13), STYK1, and VANGL1 play diverse roles in neuronal proliferation, differentiation, and survival in response to injury.

unrelated to the disease. The only way to make sense of these associations is to cross-match genomics with high-resolution phenomics (29). One can think of it as a "lock and key" combination (or, more precisely, many such combinations), where both pieces of information are needed to be able to interpret the results with confidence. Note that our approach complements meta-analysis (39) and/or pathway analyses (40), focusing the search on the combined genotypic-phenotypic architecture.

Despite the described constraints, we successfully identified more than 81% of the genotypic-phenotypic relationships previously found in the MGS data set in two independent samples. These samples were the only ones where both genotypic and detailed phenotypic features were available and provided by the researchers. Remarkably, the identification was performed with half of the SNPs used in the MGS study, because of the different platforms and our conservative preference to avoid external imputations. The success of our replication efforts strongly supports the validity and power resulting from combining genomic and phenomic information in association studies.

Overall, we believe our approach is a pioneering effort to specify complex but manageable patterns of gene-gene interaction underlying the polygenic risk of schizophrenia. In addition, our results hold promise for the emergence of a new era in clinical psychiatry in which person-centered treatment of complex disorders can be guided by reliable assessments of well-validated clinical syndromes and their specific causes.

AUTHOR AND ARTICLE INFORMATION

From the Department of Psychiatry and the Department of Genetics, Washington University School of Medicine, St. Louis; the Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain; St. Louis VA Medical Center, St. Louis; Roskamp Laboratory of Brain Development, Modulation, and Repair, Department of Psychiatry and Behavioral Neurosciences, University of South Florida, Tampa; Washington DC VA Medical Center, Washington, D.C.; and the Department of Psychiatry, University of Southern California, Los Angeles.

Address correspondence to Dr. Zwir (zwiri@psychiatry.wustl.edu).

Supported in part by the Spanish Ministry of Science and Technology (projects TIN2009-13950 and TIN2012-38805) and by the R.L. Kirschstein National Research Award to Dr. Zwir; by NIH for the Molecular Genetics of Schizophrenia Consortium, including NIH grants R01MH060879, R01MH067257, R01MH059588, R01MH059571, R01MH059565, R01MH059587, R01MH060870, R01MH059566, R01MH059586, R01MH061675, R01MH081800, U01MH046276, U01MH046289, U01MH046318, U01MH079469, U01MH079470, 5K08MH077220, 5R01MH052618-05, 5R01MH058693-06, and 3R01MH085548-05S1; and by the Genetic Association Information Network.

The authors thank NIMH for its foresight in initiating its Molecular Genetics Initiative in 1989 to support the collection of the data. They also thank the study participants, the investigators, and the research staff for data collection and genotyping at the study sites of the Molecular Genetics of Schizophrenia Consortium. In particular they thank Art Schaffer and Caroline Drain for 25 years of devoted service as research coordinators at Washington University; C. Zorumski for comments on an earlier draft; H. Huang and C. Gu for helpful input and discussions; and members of the Molecular Genetics of Schizophrenia Consortium for their constructive comments on earlier drafts of the manuscript. The data analyzed here were from dbGaP databases (see reference 41), to which Dr. Cloninger was given access under dbGaP study 2358, "Genetic and Non-Genetic Factors in the Genesis of Schizophrenia."

The authors report no financial relationships with commercial interests.

The Molecular Genetics of Schizophrenia Consortium includes P.V. Gejman, A.R. Sanders, J. Duan (North Shore University Health System and University of Chicago), C.R. Cloninger, D.M. Svrakic (Washington University, St. Louis), N.G. Buccola (Louisiana State University Health Sciences Center, New Orleans), D.F. Levinson, J. Shi (Stanford University, Stanford, Calif.; Dr. Shi is now at the National Cancer Institute), B.J. Mowry (Queensland Centre for Mental Health Research, Brisbane, and Queensland Brain Institute, University of Queensland, Brisbane), R. Freedman, A. Olincy (University of Colorado Denver), F. Amin (Atlanta Veterans Affairs Medical Center and Emory University, Atlanta), D.W. Black (University of Iowa Carver College of Medicine, Iowa City), J.M. Silverman (Mount Sinai School of Medicine, New York), and W.F. Byerley (University of California, San Francisco).

Received April 2, 2014; revision received June 2, and June 25, 2014; accepted June 30, 2014.

REFERENCES

1. de Vries PJ: Genetics and neuropsychiatric disorders: genome-wide, yet narrow. Nat Med 2009; 15:850–851
2. Wray NR, Visscher PM: Narrowing the boundaries of the genetic architecture of schizophrenia. Schizophr Bull 2010; 36:14–23

3. Cicchetti D, Rogosch FA: Equifinality and multifinality in developmental psychopathology. Dev Psychopathol 1996; 8I:597–600
4. Hill WG, Zhang XS: On the pleiotropic structure of the genotype-phenotype map and the evolvability of complex organisms. Genetics 2012; 190:1131–1137
5. Svrakic DM, Zorumski CF, Svrakic NM, Zwir I, Cloninger CR: Risk architecture of schizophrenia: the role of epigenetics. Curr Opin Psychiatry 2013; 26:188–195
6. Gottesman II, Shields J, Meehl PE: Schizophrenia: The Epigenetic Puzzle. New York, Cambridge University Press, 1982
7. Consortium Cross-Disorder Group of the Psychiatric Genomics Consortium: Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 2012; 45: 984–994
8. Lee SH, DeCandia TR, Ripke S, Yang J, Sullivan PF, Goddard ME, Keller MC, Visscher PM, Wray NR; Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ); International Schizophrenia Consortium (ISC); Molecular Genetics of Schizophrenia Collaboration (MGS): Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat Genet 2012; 44:247–250
9. 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
10. Yngvadottir B, Macarthur DG, Jin H, Tyler-Smith C: The promise and reality of personal genomics. Genome Biol 2009; 10:237
11. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH: Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet 2010; 11: 446–450
12. Williams SM, Haines JL: Correcting away the hidden heritability. Ann Hum Genet 2011; 75:348–350
13. Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kähler AK, Akterin S, Bergen SE, Collins AL, Crowley JJ, Fromer M, Kim Y, Lee SH, Magnusson PK, Sanchez N, Stahl EA, Williams S, Wray NR, Xia K, Bettella F, Borglum AD, Bulik-Sullivan BK, Cormican P, Craddock N, de Leeuw C, Durmishi N, Gill M, Golimbet V, Hamshere ML, Holmans P, Hougaard DM, Kendler KS, Lin K, Morris DW, Mors O, Mortensen PB, Neale BM, O'Neill FA, Owen MJ, Milovancevic MP, Posthuma D, Powell J, Richards AL, Riley BP, Ruderfer D, Rujescu D, Sigurdsson E, Silagadze T, Smit AB, Stefansson H, Steinberg S, Suvisaari J, Tosato S, Verhage M, Walters JT, Levinson DF, Gejman PV, Kendler KS, Laurent C, Mowry BJ, O'Donovan MC, Owen MJ, Pulver AE, Riley BP, Schwab SG, Wildenauer DB, Dudbridge F, Holmans P, Shi J, Albus M, Alexander M, Campion D, Cohen D, Dikeos D, Duan J, Eichhammer P, Godard S, Hansen M, Lerer FB, Liang KY, Maier W, Mallet J, Nertney DA, Nestadt G, Norton N, O'Neill FA, Papadimitriou GN, Ribble R, Sanders AR, Silverman JM, Walsh D, Williams NM, Wormley B, Arranz MJ, Bakker S, Bender S, Bramon E, Collier D, Crespo-Facorro B, Hall J, Iyegbe C, Jablensky A, Kahn RS, Kalaydjieva L, Lawrie S, Lewis CM, Lin K, Linszen DH, Mata I, McIntosh A, Murray RM, Ophoff RA, Powell J, Rujescu D, Van Os J, Walshe M, Weisbrod M, Wiersma D, Donnelly P, Barroso I, Blackwell JM, Bramon E, Brown MA, Casas JP, Corvin AP, Deloukas P, Duncanson A, Jankowski J, Markus HS, Mathew CG, Palmer CN, Plomin R, Rautanen A, Sawcer SJ, Trembath RC, Viswanathan AC, Wood NW, Spencer CC, Band G, Bellenguez C, Freeman C, Hellenthal G, Giannoulatou E, Pirinen M, Pearson RD, Strange A, Su Z, Vukcevic D, Donnelly P, Langford C, Hunt SE, Edkins S, Gwilliam R, Blackburn H, Bumpstead SJ, Dronov S, Gillman M, Gray E, Hammond N, Jayakumar A, McCann OT, Liddle J, Potter SC, Ravindrarajah R, Ricketts M, Tashakkori-Ghanbaria A, Waller MJ, Weston P, Widaa S, Whittaker P, Barroso I, Deloukas P, Mathew CG, Blackwell JM, Brown MA, Corvin AP, McCarthy MI, Spencer CC, Bramon E, Corvin AP, O'Donovan MC, Stefansson K, Scolnick E, Purcell S, McCarroll SA, Sklar P, Hultman CM, Sullivan PF; Multicenter Genetic Studies of Schizophrenia Consortium; ; Psychosis Endophenotypes International Consortium; Wellcome Trust Case Control Consortium 2: Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet 2013; 45:1150–1159
14. 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
15. Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, Lin X: Powerful SNP-set analysis for case-control genome-wide association studies. Am J Hum Genet 2010; 86:929–942
16. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X: Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 2011; 89:82–93
17. Arnedo J, del Val C, de Erausquin GA, Romero-Zaliz R, Svrakic D, Cloninger CR, Zwir I: PGMRA: a web server for (phenotype × genotype) many-to-many relation analysis in GWAS. Nucleic Acids Res 2013; 41:W142–W149
18. International Schizophrenia Consortium: Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 2008; 455:237–241
19. Consortium TSPG; Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium: Genome-wide association study identifies five new schizophrenia loci. Nat Genet 2011; 43:969–976
20. Stroup TS, McEvoy JP, Swartz MS, Byerly MJ, Glick ID, Canive JM, McGee MF, Simpson GM, Stevens MC, Lieberman JA: The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol development. Schizophr Bull 2003; 29:15–31
21. Sklar P, Pato MT, Kirby A, Petryshen TL, Medeiros H, Carvalho C, Macedo A, Dourado A, Coelho I, Valente J, Soares MJ, Ferreira CP, Lei M, Verner A, Hudson TJ, Morley CP, Kennedy JL, Azevedo MH, Lander E, Daly MJ, Pato CN: Genome-wide scan in Portuguese Island families identifies 5q31–5q35 as a susceptibility locus for schizophrenia and psychosis. Mol Psychiatry 2004; 9:213–218
22. Lieberman JA: Comparative effectiveness of antipsychotic drugs: a commentary on: Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1) and Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE). Arch Gen Psychiatry 2006; 63:1069–1072
23. Sullivan PF, Lin D, Tzeng JY, van den Oord E, Perkins D, Stroup TS, Wagner M, Lee S, Wright FA, Zou F, Liu W, Downing AM, Lieberman J, Close SL: Genomewide association for schizophrenia in the CATIE study: results of stage 1. Mol Psychiatry 2008; 13: 570–584
24. Zwir I, Shin D, Kato A, Nishino K, Latifi T, Solomon F, Hare JM, Huang H, Groisman EA: Dissecting the PhoP regulatory network of Escherichia coli and Salmonella enterica. Proc Natl Acad Sci USA 2005; 102:2862–2867
25. Zwir I, Huang H, Groisman EA: Analysis of differentially-regulated genes within a regulatory network by GPS genome navigation. Bioinformatics 2005; 21:4073–4083
26. Romero-Zaliz R, Del Val C, Cobb JP, Zwir I: Onto-CC: a web server for identifying Gene Ontology conceptual clusters. Nucleic Acids Res 2008; 36:W352–W357
27. Harari O, Park SY, Huang H, Groisman EA, Zwir I: Defining the plasticity of transcription factor binding sites by deconstructing DNA consensus sequences: the PhoP-binding sites among gamma/enterobacteria. PLOS Comput Biol 2010; 6:e1000862
28. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: Systematic determination of genetic network architecture. Nat Genet 1999; 22:281–285

29. Houle D, Govindaraju DR, Omholt S: Phenomics: the next challenge. Nat Rev Genet 2010; 11:855–866
30. Faraone SV, Blehar M, Pepple J, Moldin SO, Norton J, Nurnberger JI, Malaspina D, Kaufmann CA, Reich T, Cloninger CR, DePaulo JR, Berg K, Gershon ES, Kirch DG, Tsuang MT: Diagnostic accuracy and confusability analyses: an application to the Diagnostic Interview for Genetic Studies. Psychol Med 1996; 26:401–410
31. Holliday EG, McLean DE, Nyholt DR, Mowry BJ: Susceptibility locus on chromosome 1q23-25 for a schizophrenia subtype resembling deficit schizophrenia identified by latent class analysis. Arch Gen Psychiatry 2009; 66:1058–1067
32. Kendler KS, Karkowski-Shuman L, O'Neill FA, Straub RE, MacLean CJ, Walsh D: Resemblance of psychotic symptoms and syndromes in affected sibling pairs from the Irish Study of High-Density Schizophrenia Families: evidence for possible etiologic heterogeneity. Am J Psychiatry 1997; 154:191–198
33. Manolio TA: Genomewide association studies and assessment of the risk of disease. N Engl J Med 2010; 363:166–176
34. 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
35. Yang J, Lee SH, Goddard ME, Visscher PM: GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011; 88:76–82
36. Blanchard JJ, Cohen AS: The structure of negative symptoms within schizophrenia: implications for assessment. Schizophr Bull 2006; 32:238–245
37. Hallmayer JF, Kalaydjieva L, Badcock J, Dragovic M, Howell S, Michie PT, Rock D, Vile D, Williams R, Corder EH, Hollingsworth K, Jablensky A: Genetic evidence for a distinct subtype of schizophrenia characterized by pervasive cognitive deficit. Am J Hum Genet 2005; 77:468–476
38. Schachtner R, Lutter D, Knollmüller P, Tomé AM, Theis FJ, Schmitz G, Stetter M, Vilda PG, Lang EW: Knowledge-based gene expression classification via matrix factorization. Bioinformatics 2008; 24:1688–1697
39. Thompson JR, Attia J, Minelli C: The meta-analysis of genome-wide association studies. Brief Bioinform 2011; 12:259–269
40. Zhang K, Chang S, Cui S, Guo L, Zhang L, Wang J: ICSNPathway: identify candidate causal SNPs and pathways from genome-wide association study by one analytical framework. Nucleic Acids Res 2011; 39:W437–W443
41. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M, Kholodov M, Zbicz K, Beck J, Kimelman M, Shevelev S, Preuss D, Yaschenko E, Graeff A, Ostell J, Sherry ST: The NCBI dbGaP database of genotypes and phenotypes. Nat Genet 2007; 39:1181–1186