

Schizophrenia or Schizophrenias? The Challenge of Genetic Parsing of a Complex Disorder

Assen Jablensky, M.D., D.M.Sc.

Schizophrenia is a broad clinical entity defined by arrays of subjective symptoms, behavioral signs, and variable patterns of course. The nonuniformity of its clinical presentation prompted Bleuler, who coined the term schizophrenia, to state as early as 1911 that “it is not a disease in the strict sense, but appears to be a group of diseases ... therefore we should speak of schizophrenias in the plural” (1). During subsequent decades of research into its etiology and neurobiology, researchers have explored numerous biological indicators tentatively associated with the disorder, including neurocognitive dysfunction, brain dysmorphology, and neurochemical abnormalities. Yet none of these variables has been definitively proven to possess the sensitivity and specificity expected of a diagnostic test or biomarker. In the recent past, genetic linkage and association studies have targeted multiple candidate loci and genes, but failed to demonstrate that any specific gene variant or a combination of genes is either necessary or sufficient to cause schizophrenia. A likely conceptual impediment is that we still do not know whether schizophrenia is a single disease process with pleiotropic manifestations at the level of cerebral organization and symptoms, or a collection of etiologically divergent, only marginally overlapping, disorders (2).

While the genetic heterogeneity of this complex disorder is widely acknowledged, it is rarely (if at all) translated into a viable research strategy at the level of the phenotype. The symptoms of schizophrenia span a wide range of psychopathology and display an extraordinary amount of interindividual variability and temporal inconstancy. Diagnosis is based primarily on the interpretation of subjective experiences as reported by the patient, and while current diagnostic criteria ensure a degree of reliability, the boundaries of the phenotype are fuzzy. Schizophrenia geneticists are facing a particularly difficult task, seeking to discover specific variants and genes contributing to an overinclusive diagnostic category for which no specific biological substrate has yet been identified—most likely because of extensive heterogeneity and an admixture of different underlying disease subtypes. As a consequence, the phenomenological similarity of patients, selected for genetic and other biological research by the current diagnostic criteria, is modest at best, and may be disconcertingly low at worst.

Current genome-wide association (GWA) studies typically involve very large, nonrandom samples of cases and controls, and they tend to bypass or mitigate the phenotype problem by the “brute force” of the great numbers, which perform some

sort of “regression to the mean.” A recent example is provided by the remarkable success of the recently published Psychiatric Genomics Consortium GWA study meta-analysis (3), comprising 36,989 schizophrenia cases and 113,075 controls from more than 80 research institutions and groups. This study identified 128 independent associations from 108 genomic loci meeting stringent criteria for genome-wide significance, of which 83 had not been reported previously. Enriched associations were found for loci in physical proximity to genes expressed in the brain, including *DRD2*, genes involved in glutamatergic neurotransmission (*GRM3*, *GRIN2A*, *SRR*, *GRIAI*), and genes encoding calcium channel subunits. A highly significant association was found for a locus close to the major histocompatibility complex region on chromosome 6, involved in acquired immune response. Importantly, several of the findings tentatively converge on molecular pathways that may play a role in the pathogenesis of the disorder. Yet, although a powerful discovery tool, schizophrenia GWA studies have certain inherent limitations (4, 5): they only identify genetic associations of weak effect size; while the number of detected significant associations is positively correlated with the ever-increasing sample sizes, it may eventually reach a point of diminishing returns (6); the polygenic risk profile scores can predict case-control status but not individual disease risk; and the stringency of the significance threshold criterion (by convention, $p < 1 \times 10^{-8}$) removes from follow-up many potentially real signals that are just below the bar. Altogether, the reported variants account for only 25%–28% of the estimated heritability of schizophrenia (0.67–0.81), leaving much of the “missing heritability” unexplained.

Against this background, the report by Arnedo et al. in this issue of the *Journal* (7) proposes a radically different, purely data-driven approach to the exploration of the genotype-phenotype relationships in schizophrenia and to the resolution of the “missing heritability” problem. The authors reanalyzed the Molecular Genetics of Schizophrenia (MGS) GWA study (8) (4,196 cases and 3,827 controls), which employed a structured diagnostic interview (the Diagnostic Interview for

The study finds tentative support for the proposition that schizophrenia is not a nosological monolith but a collection of partially overlapping clinical syndromes.

Genetic Studies) to characterize the symptom profiles. Using a generalized factorization method to scan for “naturally” occurring clusters of intercorrelated single-nucleotide polymorphisms (SNPs) in the genomic data (9), they identified 723 such sets. Independently of the SNP clustering, they applied the same algorithm to perform a similar, unsupervised clustering of subjects and their symptom profiles (comprising 93 clinical features) into 342 phenotype sets, which were then linked to the SNP sets, thus forming multiple interactive genotypic networks. The networks were interconnected by polymorphisms indexing genes previously identified by GWA studies as being associated with schizophrenia, as well as genes reported as abnormally expressed in the brains of affected individuals. The risk of schizophrenia was examined for each SNP/phenotype network. While many of the networks had SNPs or subjects in common, 42 SNP sets with greater than 70% schizophrenia risk represented “disjoint” subnetworks, sharing neither SNPs nor subjects, as would be expected if schizophrenia is a heterogeneous group of disorders. One of the sets had a risk of 100%, indicating that all members were schizophrenia cases. Different SNP sets were associated with particular symptom patterns, which aggregated into eight tentative clinical syndromes, differing from one another by severity and the ratio of positive to negative symptoms. The authors proposed that the heritability of schizophrenia is not “missing” but is in fact distributed over a large number of genotypic-phenotypic subsets. In a subsequent stage of analysis, the authors replicated their computational approach on two independent case-control samples with symptom profile phenotypes (the Clinical Antipsychotic Trials of Intervention Effectiveness study [CATIE; 10] and the Portuguese Island Study [PIS; 11]) to test the robustness of their findings. Using jackknifing and leave-one-set-out procedures, the replication confirmed 17 SNP networks, each corresponding to a discrete SNP cluster and to a characteristic clinical syndrome, thus supporting the hypothesis that schizophrenia is a composite collection of partially overlapping but distinct disorders underpinned by separate genotypic networks.

The *Journal's* publication of the Arnedo et al. article online on Sept. 15, 2014, evoked a detailed critical commentary on the web site of the Schizophrenia Research Forum (12) by a group of leading investigators of the Psychiatric Genomic Consortium, who urged caution in the interpretation of the results of the study. The critique raised questions concerning particular aspects of methodology: the lack of correction for population structure and stratification (both MGS and CATIE included mixed samples of participants with European and African ancestry); a possible confounding of the interacting SNP sets by both ancestry admixture and linkage disequilibrium (LD) (many of the SNP sets map to very large LD blocks); the way 2,891 SNPs from the MGS study were selected for analysis (the SNPs were subjected to a permutation test, but since they had been selected on the basis of their p values of case-control association in the MGS, the permutation test did not yield a valid null distribution); and lack of clarity on how the replication of the CATIE and PIS samples was conducted (did

the degree of replication deviate from what could be expected by chance?).

In their response on the Schizophrenia Research Forum web site (12), the authors point out that the “person-centered” analyses employed in the study test directly the association between genotypic and phenotypic variability within individuals, thus obviating the need for group-wise correction for ethnicity. Ethnic stratification had been considered as a covariate but was found to have little overall impact on the results, as shown by the fact that many of the SNP sets contained comparable numbers of subjects of European and African ancestry. As regards the possibility of artifactual clustering of SNPs due to LD, set membership was determined by the covariation of polymorphisms within particular subgroups of subjects regardless of whether or not these polymorphisms were in LD in the total population (actually, the majority of SNPs in high-risk sets map to genomic regions that are far apart, or on different chromosomes, and therefore unlikely to be in LD). Finally, the permutation test was used to assess the approximate probability of the association between SNP sets and symptom sets, rather than to establish a null distribution, and did not include controls. In conclusion, the authors emphasize the model-free, data-driven nature of their analyses and the absence of any a priori assumptions or expectations about their outcomes. They regard GWA studies and their own approach as “complementary perspectives and procedures” (12).

What, on balance, can be learned from this study and the exchange of commentaries between two groups of highly skilled researchers? The Arnedo et al. study presents a novel and challenging complementary approach to the current design and methodology of large-scale meta-analyses of GWA data on schizophrenia. This approach is not an isolated endeavor; a conceptually related strategy of “phenotype-based genetic association study” was utilized in a 2011 study by Papiol et al. (13). Importantly, the Arnedo et al. study suggests that the perennial problem of “missing” or hidden heritability of schizophrenia may eventually be resolved by partitioning the totality of associated polymorphisms or genomic markers into “natural” subsets with particular phenotypic features. The study finds tentative support for the proposition that schizophrenia is not a nosological monolith but a collection of partially overlapping clinical syndromes, each of them associated with a relatively discrete set of genetic polymorphisms. Yet there are caveats that need to be stated. The phenotypes available to the researchers comprised clinical symptoms, as assessed by the Diagnostic Interview for Genetic Studies and the Positive and Negative Syndrome Scale, with unknown margins of error arising from both the subjective reports of patients and the interpretation of those reports by examining clinicians. Notwithstanding the availability of diagnostic criteria and structured research instruments, misclassification in the fine-grain assessment of symptoms remains a factor compounding further the heterogeneity of case-control samples collected at different sites. It is difficult to discern from Table 2 in the Arnedo et al. article phenotype patterns that

correspond to “true genetic syndromes,” and the authors’ statement that the “emerging picture is suggestive of a possible pathophysiology” is an overstatement. What is lacking is objective quantitative endophenotypes, such as cognitive tests, brain electrophysiology, and neuroimaging, which can complement symptom assessments and are likely to produce phenotype characterization at a more fundamental level. Phenotype refinement through disaggregation into clinical subtypes, and extension by covariate quantitative traits or endophenotypes (14), has so far had a limited following in schizophrenia research. Subtyping strategies supported by mounting evidence that sample stratification, using quantitative traits, can reduce heterogeneity and substantially increase power. This approach has scored successes in the genetics of other complex diseases, and its application to schizophrenia genetics will bring the disorder into the mainstream of current research on other common genetic diseases. The result could be the confirmation of Bleuler’s conjecture of the existence of many schizophrenias.

AUTHOR AND ARTICLE INFORMATION

From the Centre for Clinical Research in Neuropsychiatry and the School of Psychiatry and Clinical Neurosciences, University of Western Australia, Perth.

Address correspondence to Dr. Jablensky (assen.jablensky@uwa.edu.au).

The author reports no financial relationships with commercial interests.

Accepted November 2014.

Am J Psychiatry 2015; 172:105–107; doi: 10.1176/appi.ajp.2014.14111452

REFERENCES

1. Bleuler E: Dementia Praecox or the Group of Schizophrenias (1911). Translated by Zinkin J. New York, International Universities Press, 1950)
2. Jablensky A: Subtyping schizophrenia: implications for genetic research. *Mol Psychiatry* 2006; 11:815–836
3. Schizophrenia Working Group of the Psychiatric Genomics Consortium: Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; 511:421–427
4. Visscher PM, Brown MA, McCarthy MI, et al: Five years of GWAS discovery. *Am J Hum Genet* 2012; 90:7–24
5. Ayalew M, Le-Niculescu H, Levey DF, et al: Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. *Mol Psychiatry* 2012; 17:887–905
6. Kendler KS, O’Donovan MC: A breakthrough in schizophrenia genetics. *JAMA Psychiatry* 2014; 71:1319–1320
7. Arnedo J, Svrakic DM, del Val C, et al: Uncovering the hidden risk architecture of the schizophrenias: confirmation in three independent genome-wide association studies. *Am J Psychiatry* 2015; 172: 139–153
8. Shi J, Levinson DF, Duan J, et al: Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 2009; 460:753–757
9. Arnedo J, del Val C, de Erausquin GA, et al: PGMRA: a web server for (phenotype x genotype) many-to-many relation analysis in GWAS. *Nucleic Acids Res* 2013; 41:W142–W149
10. Sullivan PF, Lin D, Tzeng J-Y, et al: Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry* 2008; 13:570–584
11. Sklar P, Pato MT, Kirby A, et al: Genome-wide scan in Portuguese Island families identifies 5q31-5q35 as a susceptibility locus for schizophrenia and psychosis. *Mol Psychiatry* 2004; 9:213–218
12. Schizophrenia Research Forum: Study claiming eight types of schizophrenia called into question (comments Oct 3 and 6, 2014; authors’ response Oct 9, 2014) (<http://www.schizophreniaforum.org/new/detail.asp?id=2097>)
13. Papiol S, Malzahn D, Kästner A, et al: Dissociation of accumulated genetic risk and disease severity in patients with schizophrenia. *Transl Psychiatr* 2011; 1:e45
14. Greenwood TA, Lazzeroni LC, Murray SS, et al: Analysis of 94 candidate genes and 12 endophenotypes for schizophrenia from the Consortium on the Genetics of Schizophrenia. *Am J Psychiatry* 2011; 168:930–946