The Genetics of Symptom-Based Phenotypes: Toward a Molecular Classification of Schizophrenia

Pamela DeRosse1–3, Todd Lencz2–4, Katherine E. Burdick2–4, Samuel G. Siris4,5, John M. Kane2–4, and Anil K. Malhotra2–4

1Center for Translational Psychiatry, The Feinstein Institute for Medical Research, Manhasset, NY; 2Department of Psychiatry Research, The Zucker Hillside Hospital, Division of the North Shore-Long Island Jewish Health System, Glen Oaks, NY; 3Division of Continuing Psychiatric Services for Schizophrenia and Related Conditions, The Zucker Hillside Hospital, Division of the North Shore-Long Island Jewish Health System, Glen Oaks, NY; 4Department of Psychiatry and Behavioral Science, Albert Einstein College of Medicine, Yeshiva University, Bronx, NY; 5Division of Medical Research, Manhasset, NY.

Objectives: Genetic linkage studies in schizophrenia (SZ) have primarily focused on the phenotype of disease susceptibility. A limited number of studies, however, have reported suggestive linkage to specific SZ symptom domains including regions on chromosomes 6, 8, and 20. We examined these chromosomal regions for association to positive, negative, and disorganized symptom clusters, using a dense set of single-nucleotide polymorphisms (SNPs). Methods: We ascertained 178 Caucasian patients with SZ for lifetime severity of clinical symptomatology using a structured diagnostic interview. The cohort was genotyped with the Affymetrix 500K microarray, from which we selected, a priori, 4833 intragenic SNPs located within chromosomal regions previously linked to specific SZ symptom clusters. Parametric tests, corrected for multiple testing, were used to compare the effects of allelic variation within these SNPs to the lifetime severity of the specific symptom domain that had been implicated by prior linkage studies. Results: We were able to extend previous reports of linkage between chromosome 6q and both positive and disorganized symptoms. Lifetime severity of positive symptoms was significantly \( P = 2.5 \times 10^{-5} \) associated with a SNP within the origin recognition complex subunit 3-like (ORC3L) gene, a gene implicated in synaptic plasticity. Level of disorganized symptoms was significantly \( P < 6.0 \times 10^{-5} \) associated 2 SNPs within the brain-specific angiogenesis inhibitor 3 (BAI3) gene, which is highly expressed in brain during development. Conclusions: These data point toward specific candidate genes located within previously implicated linkage peaks for clinical symptomatology. Identification of functional variants within these regions and a characterization of the effect of these risk genotypes on the treatment of specific clinical symptoms are needed.

Key words: schizophrenia/symptoms/quantitative traits/ ORC3L/BAI3

The clinical presentation of schizophrenia (SZ) is complex and characterized by a myriad of symptoms that may include positive, negative, and disorganized features. Because no single symptom or symptom cluster is considered pathognomonic, the current categorical/syndromal nosology may not be optimal for elucidating the molecular mechanisms underlying the illness. Consequently, it has been suggested that the current system be replaced1 or supplemented2 by a dimensional, symptom-based approach focused on individual, subsyndromal, and quantitative phenotypes.3–7 Such a nosology is attractive because it provides a mechanism for studying the heterogeneity of SZ and may enhance our ability to identify the underlying pathophysiology of the illness.8

The notion that the diagnostic entity of SZ functions only as a boundary around a collection of heterogeneous pathophysiological processes is consistent with the view that SZ is a polygenic disorder in which many genes of small effect may be contributing to its risk. Moreover, it may provide an explanation for the failures to replicate so often seen in SZ genetics. Specifically, if multiple and variable genetic loci are associated with pathophysiological and phenotypic variability in SZ, substantial genetic overlap between healthy and broadly defined patient groups will exist. Thus, any comparison of these groups would likely mask or minimize the true risk alleles associated with more homogenous phenotypes.

Empirical support for a quantitative, symptom-based approach to elucidating the molecular underpinnings of SZ can be derived from family studies that demonstrate the heritability of symptom factors including positive, negative, and affective symptoms,9 thought withdrawal, thought insertion; thought broadcasting; and delusions...
of control and disorganization in sibling pairs (including twins) concordant for SZ. To date, however, relatively limited work has been conducted to identify, at the molecular level, the genetic variants associated with specific clinical phenomena. Gene-symptom relationships have emerged primarily from follow-up studies of putative SZ risk genes, with only a handful of replicated findings. For example, significant relationships have been identified between single-nucleotide polymorphisms (SNPs) in the Disrupted in Schizophrenia 1 (DISC1) gene and the severity of delusions in patients with SZ. Based on these data, it seems likely that susceptibility genes for SZ may not only increase the risk for the illness but also influence the clinical presentation of the illness. Moreover, a number of studies have reported data indicating that specific genetic variants may act to modify the clinical presentation of illness without increasing the overall risk for the illness. These include reported relations between negative symptoms and variation in BDNF, DAT1, and hkCa3; positive symptoms and variation in DRD4, DRD2, CCK-A, and 5-HTT; and disorganization and variation in DRD2.

In contrast to candidate gene studies, genetic linkage approaches have the capacity to detect novel loci located within relatively large chromosomal regions. To date, only a few studies have utilized linkage approaches to identify quantitative trait loci (QTLs) for SZ symptoms. Brzustowicz et al assessed 28 genetic markers spanning chromosome 6 for linkage to the positive, negative, and general psychopathology subscales of the Positive and Negative Syndrome Scale (PANSS) in 10 moderate sized affected families (n = 183). Despite no significant linkage to diagnosis, significant evidence for linkage to specific PANSS subscales was observed at several loci on chromosome 6p. Kendler et al conducted a similar analysis on 4 genomic regions on chromosome 5q, 6p, 8p, and 10p that had been previously linked to the broad diagnostic category of SZ in a large sample of 265 affected families (n = 1408) from the Irish Study of High-Density Schizophrenia Families. Linkage to 8p22-8p21 was reported for a Kraepelinian constellation of symptoms, including negative thought disorder, affective deterioration, and poor outcome. Wilcox et al conducted a genome-wide linkage scan of symptoms in SZ, in a study of 51 families (n = 136) derived from the National Institute of Mental Health–funded Genetics Initiative on Schizophrenia. These analyses revealed suggestive linkage to chromosomes 6, 9, and 20 for the disorganized symptoms trait and to chromosome 12 for the negative symptoms trait. Finally, Fanous et al recently conducted a genome-wide scan in 270 Irish high-density families (n = 755), using clinically homogeneous phenotypes derived from a latent class analysis of the Operational Criteria Checklist for Psychotic Illness. This analysis implicated several regions of suggestive linkage to clinical phenotypes including a region on chromosome 20 linked to deficit syndrome, a clinical phenotype characterized by severe negative symptoms.

Unfortunately, follow-up linkage disequilibrium studies have not been reported, and no genes in these previously identified linkage regions have been identified that significantly influence the clinical presentation of SZ. Thus, the present study was designed to densely map the previously implicated linkage regions to detect the specific genes in these regions that may be acting to modify the clinical presentation of SZ.

Methods

Participant Sample

The study sample was comprised of 178 patients who met criteria for a SZ-spectrum disorder (SZ, n = 158; schizoaffective disorder, n = 13; or schizophreniform disorder, n = 7), based on the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, (SCID). All participants were recruited from the Zucker Hillside Hospital (ZHH), a division of the North Shore-Long Island Jewish Health System (NSLIJHS), in Glen Oaks, NY, and provided written informed consent to a protocol approved by the Institutional Review Board of the NSLIJHS.

Clinical Assessment

Each subject was assessed with a SCID administered by trained and reliable raters. Information obtained from the SCID was supplemented by a review of medical records and interviews with family informants when possible. The interview, clinical information from available medical records, and information from family informants were then compiled into a narrative case summary and presented at a diagnostic consensus conference. Diagnoses were then determined by a consensus among a minimum of 3 expert diagnosticians from the ZHH faculty.

Because symptom severity often varies during the course of illness, a lifetime symptom severity rating rather than a cross-sectional rating was used. Lifetime symptom ratings were derived from SCID data and included overall ratings on negative symptoms, positive symptoms, and disorganized symptoms. The overall negative symptom rating included ratings on avolition, alopecia, and affective flattening. The overall positive symptom rating included ratings on delusions (referential, paranoid, grandiose, somatic, control, thought broadcasting, bizarre, and other delusions) and hallucinations (auditory, visual, tactile, and other hallucinations). Finally, the overall disorganization symptom rating included ratings on disorganized speech and disorganized behavior. Each SCID item used in these ratings were rated on a continuous scale where 1 = absent, 2 = subthreshold, and 3 = present. The overall rating for each symptom cluster

P. DeRosse et al.
was calculated by summing the scores of all items within a cluster.

Genotyping

Genomic DNA was extracted from whole blood, and genotype data were derived using the Affymetrix 500K array (Affymetrix, Santa Clara, CA) and are described in detail by Lencz et al.31 We restricted our analyses to SNPs located within the chromosomal regions previously linked to quantitative symptom traits in SZ. Moreover, allelic variation at each SNP within a region was only tested for association to symptom severity in the specific domain to which it had been previously linked.

Selection of SNPs within linkage regions was based on the statistical results reported for each microsatellite marker that had been assessed in the individual linkage studies. We selected SNPs within regions encompassed by neighboring microsatellite markers that were nominally associated with a specific symptom domain (P < .0513; P < .0133; logarithm of the odds (LOD) > 225,27,28). Each nominally associated marker was entered into the University of California at Santa Cruz genome browser (UCSC) database to identify the corresponding location of that marker on the physical chromosome map (Build 35, May 2004). The physical positions of the most telomeric and centromeric boundaries were then used to limit the region of the chromosome we investigated.

Based on the report of Brzustowicz et al.,25 SNPs on chromosome 6 encompassed by markers D6S477–D6S1270 were assessed for association to positive symptoms. Based on the report of Kendler et al.,26 SNPs on chromosome 20 encompassed by markers D20S115–D20S481 were assessed for association to disorganized symptoms. Finally, SNPs on chromosome 6p encompassed by markers D6S426–D6S1270 were assessed for association to disorganized symptoms. Based on the report of Fanous et al.,28 SNPs on chromosome 20 encompassed by markers D20S115–D20S481 were assessed for association to negative symptoms based on the findings reported by Fanous et al.28

For each linkage region identified by the aforementioned studies, we included an additional 100 kb on each side of the marker boundaries to ensure optimal coverage of the linkage region. Further, to reduce the number of statistical tests being carried out, only intragenic SNPs were selected for analysis. Of note, although Wilcox et al27 identified single markers on chromosome 9, 12, and 20, with LOD scores greater than 2, these regions were not examined in the present study due to the lack of a second significant boundary to delineate a chromosomal region.

Based on the aforementioned methodology of SNP selection and the availability of SNPs on the Affymetrix array, we assessed a total of 5747 SNPs in the region on chromosome 6p linked to positive symptoms and 1399 SNPs in the region on chromosome 6p linked to negative symptoms by Brzustowicz et al.25 Further, the region on chromosome 8p derived from the Kendler et al26 study yielded a total of 3979 SNPs for analysis. The Wilcox et al27 linkage region on chromosome 6p yielded 5574 SNPs for analysis. Finally, 2211 SNPs on chromosome 20pq were assessed to test the Fanous et al28 linkage region. Overall, a total of 12 438 SNPs were available on the Affymetrix array for the present analyses of which 4833 were intragenic and tested for association to symptom severity. The number of intragenic SNPs, as well as information about the chromosomal regions, source citations (including associated symptom cluster), and number of SNPs assessed in each region, are shown in table 1.

Statistical Analyses

As detailed in table 1, a total of 5 hypotheses were tested in the present study. Specifically, we tested for association between positive symptoms and allelic variation in SNPs from a region of chromosome 6q, disorganized...
Table 2. SNPs Within Previously Identified Linkage Regions That Were Significantly Associated With Symptom Severity

<table>
<thead>
<tr>
<th>Gene</th>
<th>Location</th>
<th>Position</th>
<th>dbSNP</th>
<th>Alleles</th>
<th>Symptom</th>
<th>Associated Allele</th>
<th>t</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORC3L</td>
<td>6q15</td>
<td>88403068</td>
<td>rs16879646</td>
<td>G/A</td>
<td>Positive</td>
<td>A</td>
<td>4.28</td>
<td>2.50 × 10⁻⁵</td>
</tr>
<tr>
<td>BAI3</td>
<td>6q12</td>
<td>69757118</td>
<td>rs1415031</td>
<td>G/C</td>
<td>Disorganized</td>
<td>C</td>
<td>4.08</td>
<td>5.69 × 10⁻⁵</td>
</tr>
<tr>
<td>BAI3</td>
<td>6q12</td>
<td>69768622</td>
<td>rs9446083</td>
<td>G/A</td>
<td>Disorganized</td>
<td>A</td>
<td>4.08</td>
<td>5.90 × 10⁻⁵</td>
</tr>
</tbody>
</table>

**Note:** The first 5 columns indicate the gene symbol, its chromosomal location, physical position, the SNP within the gene that was significantly associated with a symptom cluster, and the major and minor alleles at that SNP (minor allele is in bold). Column 6 indicates the symptom cluster to which the SNP was associated, and column 7 indicates the allele associated with increased severity on the symptom cluster. Statistical comparisons are presented in columns 8 and 9. SNP, single-nucleotide polymorphism; BAI3, brain-specific angiogenesis inhibitor 3.

Results

We tested regions on chromosome 6, 8p, and 20pq for association with specific SZ symptoms domains and identified 3 intragenic SNPs that met Bonferroni-corrected criteria for association to symptom severity. Specifically, we identified a single SNP associated with lifetime severity of positive symptoms in our cohort that was located within the region of 6q that had previously been linked to positive symptoms. Twenty-nine additional SNPs in BAI3 that were available for analysis were tested but were not associated with the severity of positive symptoms.

Further, at 6q12, the minor alleles at 2 SNPs, rs1415031 and rs9446083, were significantly associated with lifetime severity of disorganized symptoms (rs1415031: \( \bar{x}_G = 4.520 \pm 0.12; t = 4.08; P = 5.69 \times 10^{-5} \); and rs9446083: \( \bar{x}_A = 5.22 \pm 0.1780 \) vs \( \bar{x}_G = 4.41 \pm 0.13; t = 4.08; P = 5.90 \times 10^{-5} \)). A comparison of disorganized symptom severity by genotype is presented in the bottom panel of figure 1. Because of the high linkage disequilibrium (LD) between these 2 SNPs (\( D' = 0.83 \)) and to avoid redundancy, only the data for rs1415031 are shown. Both of these SNPs are located in the intronic region 3' to exon 9 in the gene encoding brain-specific angiogenesis inhibitor 3 (BAI3). Eighty-nine additional SNPs in BAI3 that were available for analysis were tested for association to disorganized symptoms. Of these, 18 (20%), spanning almost the entire coding region of the gene, were nominally associated with disorganized symptoms (\( P < 0.05 \)). These SNPs, along with their physical position and associated \( P \) values are shown in table 3.

Discussion

We report an association between a SNP on chromosome 6q and positive symptoms and 2 SNPs on chromosome 6q and disorganized symptoms in SZ. On chromosome 6q, rs16879646 was significantly associated with positive symptoms in SZ. These results are consistent with the findings of Brzustowicz et al. who reported significant linkage of this region to negative symptom severity in SZ. Moreover, on chromosome 6q, rs1415031 and rs9446083 were significantly associated with the severity of disorganization in SZ. These results are consistent with the findings of Wilcox et al. who reported suggestive linkage of this region to disorganization in SZ.

The SNP identified in the region of chromosome 6q that had been linked to positive symptoms by Brzustowicz
et al.\textsuperscript{25} is located within the \textit{ORC3L} gene. \textit{ORC3L} is a component of the 6-subunit protein complex known as the ORC. \textit{ORC3L} is a 77-kb gene comprised of 20 exons and is located at 6q15. Although the most well-documented aspect of ORC function is related to the formation of prereplication complexes upon exit from mitosis, recent data implicate ORC in sister chromatin cohesion, cytokinesis, and neural dendritic branching.\textsuperscript{32}

Although little is known about the function of human \textit{ORC3L}, the drosophila homolog, \textit{Latheo}, codes for a protein that is localized to the presynaptic terminals of differentiated motor neurons and appears to modulate mechanisms necessary for behavioral plasticity.\textsuperscript{33} Mutations in \textit{Latheo} have been shown to disrupt associative learning through its effect on Ca\textsuperscript{2+} and activity-dependent synaptic plasticity.\textsuperscript{34} Alterations in synaptic plasticity have long been suggested to play a role in both the development of schizophrenic symptoms, particularly positive symptoms, and the deterioration process in \textit{SZ}.\textsuperscript{35} Thus, these data suggest that variation in \textit{ORC3L} may influence the lifetime severity of positive symptoms by mediating Ca\textsuperscript{2+} and activity-dependent synaptic plasticity.

The SNPs identified in the region of chromosome 6q that had been linked to disorganization by Wilcox et al.\textsuperscript{27} are located within the gene encoding \textit{BAI3}. \textit{BAI3} is a p53-target gene encoding a 7-span transmembrane protein and is thought to be a member of the secretin, g-protein–coupled receptor family.\textsuperscript{36} \textit{BAI3} is located at chromosome 6q12 and is a 750-kb gene comprised of 31 exons. Very little is known about the function of human \textit{BAI3}, and to date, a relatively limited amount of data regarding the function of \textit{BAI3} homologs have been reported in other species. Recent work in mice, however, indicates that the protein coded for by mouse \textit{BAI3} is expressed almost exclusively in the brain with the highest levels of expression of \textit{BAI3} during the early neonatal period. Approximately 10 days after birth, \textit{BAI3} decreases steadily until adulthood and relatively little expression of \textit{BAI3} is seen in the adult brain.\textsuperscript{37} Thus, it seems likely that \textit{BAI3} plays a role in the neurodevelopmental process and may influence the clinical symptomatology of \textit{SZ} through its involvement in

<table>
<thead>
<tr>
<th>dbSNP</th>
<th>Position</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs544398</td>
<td>69632524</td>
<td>0.04</td>
</tr>
<tr>
<td>rs1880177</td>
<td>69646677</td>
<td>0.03</td>
</tr>
<tr>
<td>rs7760666</td>
<td>69678581</td>
<td>0.01</td>
</tr>
<tr>
<td>rs2184723</td>
<td>69703807</td>
<td>0.02</td>
</tr>
<tr>
<td>rs1932615</td>
<td>69713729</td>
<td>0.04</td>
</tr>
<tr>
<td>rs3823064</td>
<td>69738338</td>
<td>0.02</td>
</tr>
<tr>
<td>rs3798979</td>
<td>69738381</td>
<td>0.02</td>
</tr>
<tr>
<td>rs1415031*</td>
<td>69757118</td>
<td>5.70 × 10\textsuperscript{-5}</td>
</tr>
<tr>
<td>rs3798995</td>
<td>69757481</td>
<td>0.02</td>
</tr>
<tr>
<td>rs9446083*</td>
<td>69768622</td>
<td>5.90 × 10\textsuperscript{-5}</td>
</tr>
<tr>
<td>rs9456474</td>
<td>69777487</td>
<td>0.01</td>
</tr>
<tr>
<td>rs9446085</td>
<td>69783083</td>
<td>0.01</td>
</tr>
<tr>
<td>rs1889878</td>
<td>69793014</td>
<td>0.01</td>
</tr>
<tr>
<td>rs7743332</td>
<td>69803749</td>
<td>0.01</td>
</tr>
<tr>
<td>rs688606</td>
<td>69840895</td>
<td>0.04</td>
</tr>
<tr>
<td>rs7759645</td>
<td>69874050</td>
<td>0.03</td>
</tr>
<tr>
<td>rs1147510</td>
<td>69912418</td>
<td>0.01</td>
</tr>
<tr>
<td>rs7739401</td>
<td>69915651</td>
<td>0.02</td>
</tr>
<tr>
<td>rs779467</td>
<td>69988140</td>
<td>0.01</td>
</tr>
<tr>
<td>rs779462</td>
<td>70001377</td>
<td>0.02</td>
</tr>
</tbody>
</table>

\textit{Note: SNP, single-nucleotide polymorphism; BAI3, brain-specific angiogenesis inhibitor 3.}

\textbf{Fig. 1.} Comparison of Lifetime Severity of Psychotic Symptoms in Patients With Schizophrenia by Genotype. The upper panel represents the data for the single-nucleotide polymorphism in \textit{ORC3L} that is significantly associated with lifetime severity of positive symptoms. The lower panel represents the data for the rs1415031 in brain-specific angiogenesis inhibitor 3 (\textit{BAI3}) that was significantly associated with lifetime severity of disorganized symptoms. Error bars indicate SDs.
neurodevelopment. This analysis is consistent with the prevailing explanatory model of SZ, which posits that the disease is the behavioral outcome of an aberration in neurodevelopmental processes that begins long before the onset of clinical symptoms.38

Several limitations of the current study should be noted. First, the method of genotyping employed in the present study limited the markers within ORC3L and BAII that could be assessed, and thus, it is possible that the SNPs we have identified are not directly associated with symptom severity. Rather, the SNPs we have identified may be in high LD with SNPs within the coding regions of these genes that are directly associated with symptom severity. Further, as with any genetic association study, the present findings may represent false positives. Although we believe that the use of a strict Bonferroni correction has substantially limited the potential for type I errors, additional independent replications will be required to confirm the present findings.

The present data support previous reports of linkage between 6q and both positive and disorganized symptoms and suggest that there are specific candidate genes located within these previously implicated linkage peaks that are associated with clinical symptomatology. These data add to the accumulating evidence demonstrating the potential utility of a symptom-based approach to the genetics in SZ. Given the clinical heterogeneity and molecular complexity of the illness, however, it is unlikely that any single genetic paradigm will account for a large degree of the variance to be explained. With the advent of large-scale genotyping platforms, hypothesis-free genome-wide searches using traditional association methods as well as more contemporary methods such as Whole Genome Homozygosity Association and copy number variation should permit discovery of additional symptom QTLs, although larger samples may be required to overcome the necessary corrections for multiple testing.

Funding

Donald and Barbara Zucker Foundation; internal funding from the NSLIJHS; Stanley Foundation (to A.K.M.); National Alliance for Research on Schizophrenia and Depression (to A.K.M.), National Institutes of Health (MH065580 to T.L., MH074543 to J.M.K., MH001760 to A.K.M.; National Institutes of Health National Alliance for Research on Schizophrenia and Depression (to A.K.M.), National Institutes of Health (MH065580 to T.L., MH074543 to J.M.K., MH001760 to A.K.M.); Stanley Foundation (to A.K.M.); Donald and Barbara Zucker Foundation; internal funding from the NSLIJHS; Stanley Foundation (to A.K.M.), National Institutes of Health National Alliance for Research on Schizophrenia and Depression (to A.K.M.), National Institutes of Health (MH065580 to T.L., MH074543 to J.M.K., MH001760 to A.K.M.; Stanley Foundation (to A.K.M.); Donald and Barbara Zucker Foundation; internal funding from the NSLIJHS; Stanley Foundation (to A.K.M.); National Institutes of Health National Alliance for Research on Schizophrenia and Depression (to A.K.M.); Donald and Barbara Zucker Foundation; internal funding from the NSLIJHS; Stanley Foundation (to A.K.M.)

Acknowledgments


References


40. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*. 2008;320:539–543.