Background: Twin and adoption studies have consistently implicated genes in the etiology of schizophrenia. Molecular genetic studies have found some consistent support for linkage to many regions of the genome. Despite these encouraging results, none of these findings have achieved genome-wide levels of statistical significance, and none have been consistently replicated.

Methods: This report is a follow-up of a genome scan that analyzed linkage to the diagnosis of schizophrenia in a series of sibling pairs in the National Institute of Mental Health Genetics Initiative for Schizophrenia data. In this report, we use the same sample to assess linkage to three quantitative traits developed from the Scale for the Assessment of Negative Symptoms and the Scale for the Assessment of Positive Symptoms: positive, negative, and disorganized symptoms.

Results: We show suggestive linkage to chromosomes 6, 9, and 20 for the disorganized trait and to chromosome 12 for the negative trait. We also show weak association with PAH (phenylalanine hydroxylase) on that chromosome.

Conclusions: The findings on chromosome 6 replicate some prior findings, the other loci are novel. A larger sample would provide more power to detect both linkage and association for this complex disorder.

Key Words: Schizophrenia, linkage, quantitative traits, Scale for the Assessment of Negative Symptoms, Scale for the Assessment of Positive Symptoms, genetics

Introduction

Twin and adoption studies have consistently implicated genes in the etiology of schizophrenia (Faraone et al 1996). Molecular genetic studies have found some consistent support for linkage to ten regions of the genome (1q21–q22, 5q11.2–q13.3, 6q13–q26, 6p23, 8p21, 11q14–q21, 13q32, 15q14, 18p, 22q11–q13). Despite these encouraging results, none of these findings have achieved genome-wide levels of statistical significance, and none have been consistently replicated. These ambiguities, which are seen for most psychiatric disorders, have motivated the search for novel approaches to molecular genetic studies of these conditions (Ott 1990; Risch 1990; Faraone et al 1999; Tsuang 2000).

Leal (2001) and others in the workshop on phenotypes and genetic analysis of complex traits raise the questions, “Which phenotypes should be chosen to maximize the heritability of a trait? What is the validity of current diagnostic schemes? Is it beneficial to use quantitative measures for psychiatric traits? How can current measurements used to phenotype individuals be used to assign a quantitative score?”

There is considerable clinical variability in the clinical presentation of schizophrenia. Kendler et al (2000) used three nested qualitative traits: narrow, intermediate, and broad affectation. Using these phenotypic classifications, a signal on chromosome 8p was replicated, but findings on 5q, 6p, and 10p failed to be replicated. Pulver et al (2000) used two stratified phenotypes, schizophrenia spectrum personality disorders and psychotic affective disorders, in genome scans. Their strongest finding was also on 8p21.

Brzustowicz et al (1997) used the total scores on the Positive symptom, Negative symptom, and General Psychopathology scales of the PANSS (Positive and Negative Syndrome Scale). They conducted both qualitative and quantitative trait analyses for a set of 28 markers on chromosome 6p. Brzustowicz et al found no evidence for linkage when qualitative traits were used; however, they found some evidence for linkage for positive symptoms, and directional findings ($p < .10$) for general psychopathology in a 22cM region spanning D6S1280 to D6S1270.

This report is a follow-up of a genome scan that analyzed linkage to the diagnosis of schizophrenia in a series of sibling pairs. In this report, we use the same sample to assess linkage to three quantitative traits: positive, negative, and disorganized symptoms. We hypothesized that these quantitative traits would provide more
precise definitions of the phenotype and thereby, greater power to detect linkage.

Methods and Materials

The National Institute of Mental Health (NIMH) funded the Genetics Initiative on Schizophrenia to create a national resource of standardized clinical data and DNA samples available to the scientific community. To produce these resources, families were ascertained by cooperative agreements between NIMH and investigators at Washington University, Harvard University, and Columbia University. Blood samples were shipped to the NIMH Cell Repository at the Coriell Institute for Medical Research in Camden, New Jersey, which created lymphoblast cell lines.

The institutional review boards of all participating institutions approved the research protocol.

Subjects

Probands affected with either DSM-III-R schizophrenia or schizoaffective disorder, depressed type were identified by systematic screening of patients in psychiatric hospitals and clinics. If a case had at least one living first-degree relative with either disorder, the family was retained for further examination. Among probands, any manic syndrome had to be brief, occurring for less than 30% of the total duration of psychosis. A diagnosis of schizophrenia or schizoaffective disorder was excluded if there was any evidence that drug abuse explained the persistent psychotic features.

Families were excluded if both parents were schizophrenic. Pedigree extension was sequential. We included all available first-degree relatives of the probands and then all first-degree relatives of any relatives who had schizophrenia, schizoaffective disorder, depressed, delusional disorder, manic subtype of schizoaffective disorder, brief reactive psychosis, schizopreniform disorder, psychosis not otherwise specified, and schizotypal personality disorder. Cloninger et al (1998) described the diagnostic distribution.

Table 1 gives the number of families, number of family members, demographics, and the distribution of sibship size for the families used in the analysis. This sample is the subset of the original NIMH sample for whom complete, or nearly complete, data from the Scales for the Assessment of Negative and Positive Symptoms (SANS, SAPS) were available.

Diagnostic Assessment

The structured interview was the Diagnostic Interview for Genetic Studies (DIGS) (Nurnberger et al 1994). Test–retest reliabilities of DIGS-based diagnoses were shown to be excellent within sites and across sites (Nurnberger et al 1994). We supplemented structured interview data by medical records and a semi-structured itemized assessment of psychopathology in family members, called the Family Instrument for Genetic Studies (FIGS). Best estimate diagnoses were made by two experienced psychiatrists or psychologists based on all available information. If there was any disagreement between the two independent clinicians, a third diagnostician was asked to evaluate the diagnosis.

We assessed quantitative traits using the SANS and SAPS (Andreasen and Olsen 1982). Five constructs are measured by the 25 items in the SANS: Affective Flattening, Alogia, Avolition Apathy, Anhedonia and Asociality, and Inattention. The 34 items in the SAPS assess four constructs: Hallucinations, Delusions, Bizarre Behavior, and Positive Formal Thought Disorder. Fifty of the 59 items assess facets of the constructs, and nine are “global” items intended to provide a general rating of functioning on each of the underlying dimensions.

Genotyping

Millennium Pharmaceuticals (Cambridge, MA) conducted the genome scan using 459 markers spaced at an average of 10 cM. The markers were selected from the CHLC-6 set, and supplemental markers were added from the Genethon map (Evry Cedex, France). The markers were di-, tri-, and tetra-nucleotide repeats that can be reliably scored using automated methods.

The polymerase chain reactions (PCR) were set up with 5.0 μL genomic DNA (4 ng/μL), 5.05 μL primer cocktail, and 4.95 μL Taq cocktail. The PCR cycling was 95°C for 30 sec, 35 cycles, 95°C for 30 sec, 72°C for 30 sec (25°C for 30 sec, 72°C for 60 sec) for 30 cycles, 72°C for 10 min. The gels were run on Applied Biosystems (ABI; Foster City, CA) 377 DNA sequencers using ABI Prism 377 data collection software. The data were analyzed with the ABI Prism GeneScan 2.0.2 with Genotyper 1.1.1 (Cloninger et al 1998).

In our prior work, we restricted the analyses by ethnicity due to observed allelic heterogeneity. In the present report, we used the combined sample but offset the European American sample by 0.0001 cM. This was done to gain the statistical power of the larger sample size and accommodate the allelic heterogeneity. There were 76–78 informative sibling-pairs for whom both genotypic and phenotypic information were available.

Statistical Methods

QUANTITATIVE TRAITS. First, missing data were imputed for the SANS and SAPS for some cases. One approach to missing data is to use list-wise deletion of observations with any missing items. In this case, more than half of the sample (57%) would have been omitted from the analyses. It is probable that
these data are not missing at random. This method was deemed unacceptable for the present analyses. An individual was omitted from the analyses only if information for either the SANS, SAPS, or both was completely missing. Otherwise, based on the null hypothesis that the original model is correct, missing data were replaced with the individual-specific mean on the items constituting the original factor.

Principal components analysis with an oblique rotation was used to determine the number of factors to retain for the principal factor analyses. The “global” ratings were included in these analyses. Based on the eigenvalues of the factors, the scree plot of the eigenvalues, the variance explained by the solutions, and the interpretability of each of the constructs, a 12-factor solution was chosen. The covariance structure for the 12-factor model was used to estimate the three-factor, second-order model. The second-order model was derived using an oblique rotation. The exploratory analyses were done using SAS software version 6.12 (SAS Institute 1989). PRELIS version 2.3 was used to compute the polychoric correlation matrix and LISREL version 8.3 (Joreskog 1993) was used for these analyses.

**LINKAGE.** Estimates of the marker allele frequencies were computed for each ethnic group separately. Analyses were performed separately for each ethnic group to investigate potential disease locus heterogeneity between the two groups. The ethnicities were analyzed together to increase power and to look for common genetic effects. The combined analysis maintained separate allele frequencies for each ethnic group by offsetting European American markers .0001 cM on the genetic map from the original position. We used the model-free linkage analysis program, SIBPAL2, from the S.A.G.E., beta version 7.0 (available at http://darwin.cwru.edu/pub/sage.html). In SIBPAL2, the quantitative traits are modeled using single multiple regression in full siblings as a function of identical by descent (IBD) sharing based on multipoint IBD information (Elston et al 2000).

LOD scores were computed from p values using the following transformation:

$$LOD = \left( \frac{\text{probit}(1 - p\text{-value})}{\sqrt{2*\log_{10}}} \right)^2$$

**Results**

The results of the factor analyses (cf. Methods) are described in Table 2, which shows the SANS and SAPS items we used to define the three quantitative traits. The “Positive” factor is composed of delusions and hallucinations: “Non-Auditory Hallucinations, Other Delusions, Delusions of Control,” and “Auditory Hallucinations.” The “Negative” factor is composed of “Flat Affect, Alogia, Avolition/Apathy,” and “Anhedonia/Asocial.” The “Disorganized” construct contains “Thought Disorder, Disorganized Speech and Affect, Bizarre Behavior,” and “Inattention.”

The linkage results are presented in Figure 1, which shows the log-odds for linkage (LOD) scores for each chromosome. Results for the Positive Symptom Score are represented with a plus sign (“+”). The Negative Symptom Score is represented with a minus sign (“-”). The Disorganized Score is represented with an asterisk (“*”).

Lander and Krugylak (1995) proposed that, for allele-sharing methods among siblings and half-siblings, suggestive linkage occurs at an LOD score of 2.2 and significant linkage occurs at an LOD score of 3.6. Using these criteria, none of our results meet the criteria for significant linkage; however, there are regions on chromosomes 6p, 6q, 9p, 12q, and 20q that meet the criteria for suggestive linkage. Table 3 summarizes the results for all markers with LOD scores greater than 2.0.

There are two apparent peaks for the disorganized phenotype on chromosome 6 at marker D6S426 (LOD = 3.34) and midway between markers D6S1053 and D6S445 (76 cM, LOD = 3.44). D6S426 spans the centromere and is located at approximately 50–60 cM. D6S1053 is located at approximately 70–80 cM and is also known as GATA64D02. CA and GT repeats also characterize marker D6S455 (80–90 cM).

A maximum LOD score of 2.97 is observed at marker D12S1300, at approximately 104 cM on chromosome 12. The LOD score for the gene encoding phenylalanine hydroxylase (PAH) at 109.5 cM is 2.12.

Association analyses were conducted for both loci on this chromosome, because PAH is a gene with known biological function and a plausible role in the etiology of schizophrenia. The multi-allelic Family-Based Association Tests (FBAT) test for quantitative traits was used (Rabinowitz and Laird 2000). Under an additive model, the results were nonsignificant at both loci and roughly equivalent [D12S1300: $\chi^2(4) = 8.40, p = .078$; PAH: $\chi^2(6) = 11.50, p = .074$]. Under a dominant model, there is significant association with the PAH gene and no association with the anonymous marker [D12S1300: $\chi^2(4) = 6.36, p = .174$; PAH: $\chi^2(6) = 14.31, p = .026$].

Table 2. Factor Structure Used to Create Three Quantitative Traits

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
<th>Disorganized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Auditory Hallucinations</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>Other Delusions</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>Delusions of Control</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>Auditory Hallucinations</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>Flat Affect</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>Alogia</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Avolition/Apathy</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>Anhedonia/Asocial</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Thought Disorder</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>Disorganized Speech &amp; Affect</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>Bizarre Behavior</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Inattention</td>
<td>40</td>
<td></td>
</tr>
</tbody>
</table>

Table entries are factor loadings. Conceptually, factor loadings can be thought of as the correlation between the original variable and the new factor. All entries have been multiplied by 100.
The maximum LOD score for disorganized symptoms on chromosome 20 is 3.04, 3 cM distal to D20S478 (also known as GATA42A03). D20S481 (also known as GATA47F05) is 5 cM distal to the maximum LOD score.

**Discussion**

Our genome scan of 78 sibling-pairs from the schizophrenia consortium of the NIMH Genetics Initiative found
suggestive evidence for linkage to five regions on chromosomes 6p, 6q, 9p, 12q, and 20q using positive, negative, and disorganized symptom quantitative traits. Our findings on chromosome 6 replicate some of the other findings in regions on 6p and 6q (Brzustowicz et al 1997). We show both suggestive linkage (Lander and Kruglyak
1995) and association to the PAH gene on chromosome 12. This is consistent with existing hypotheses about the role of this amino acid in the etiology of schizophrenia; however, our maximum LOD score was 5 cM proximal to the gene. There are no other reports of linkage to the regions we report on chromosomes 9p and 20q.

These findings are markedly different from the earlier reports from this sample. The present results are reported for the combined European American and African American data. The original reports from these data were for each of the sub-samples separately, due to observed allelic heterogeneity. Affection was defined by a DSM-III-R diagnosis of schizophrenia or schizoaffective disorder, depressed. In the prior report, suggestive linkage was reported for two markers on 10p for the European American sub-sample. Other regions with p values less than .01 were 2q12, 10q11.2, and 10q21 (Faraone et al. 1998, 1999). In the African American sample, four regions showed nominal linkage, with p values less than .05, 6q16–6q24, 8pter–8q12, 9q32–9q34, and 15p13 (Kaufmann et al. 1998). These findings were markedly different from those observed in the European American pedigrees.

Using the combined sample, Freedman (2001b) showed significant disequilibrium at the alpha7-nicotinic acetylcholine receptor subunit gene (CNRNA7) on 15q13–14, a finding we did not replicate.

In the present report, the only concordance with prior findings is on chromosome 6. We found suggestive linkage to the disorganized trait at 6q15, adjacent to the region showing some evidence of linkage in the African American sample. The differences may be due to a number of factors: the combined sample was used, providing greater statistical power; the dependent variables were quantitative rather than qualitative traits; and the current analyses employed a sibling-based allele-sharing method in SIBPAL, rather than GENEHUNTER.

Our study adds to a growing literature studying quantitative measures of psychopathology in schizophrenia pedigrees. Kendler et al. (2000) used three qualitative traits. Our findings, using a different phenotype and different analytic strategy, replicate the negative findings of Kendler et al. on 6p for the regions typed in common.

Pulver et al. (2000) also used two stratified phenotypes, schizophrenia spectrum personality disorders and psychotic affective disorders. Their strongest finding was on 8p21. We did not find any signal on chromosome 8.

Bruzustowicz et al. (1997) used the total scores on the Positive symptom, Negative symptom, and General Psychopathology scales of the PANSS in a quantitative trait analysis. Our data include 17 of the markers in the screening set they used. Bruzustowicz et al report significant linkage results for Positive symptoms, and directional findings for General Psychopathology as quantitative trait in a 22 cM region spanning D6S1280 to D6S1270 (also known as GATA29C09). This region generally overlaps our second peak. Our maximum LOD score is 5 cM distal to D6S1053 (LOD = 3.44) for Disorganized symptoms.

The peak of the Bruzustowicz et al. finding is at GATA29C09, 4cM proximal to D6S1053. The findings are in the same region, but the phenotypes are somewhat different. The PANSS does not measure Disorganized symptoms as such. Disorganized symptoms include Thought Disorder, Disorganized Speech and Affect, Bizarre Behavior, and Inattention. These constructs are likely part of the Positive scale in the PANSS. Norman et al. (1996) reported high correlations between the positive and negative scales of the SANS/SAPS and the PANSS. So, although there are no findings in our report for our definition of Positive symptoms, it is likely that the definition of Positive symptoms in the PANSS is broader and includes the constructs we call Disorganized symptoms. It is noteworthy that we also replicated the negative findings near 6pter. Bruzustowicz included 5 markers implicated in earlier studies in the region spanning 9–34 cM. They also found no linkage to either qualitative or quantitative traits in this region near 6pter.

Our analysis of chromosome 6 showed results over a broad region extending from D6S426 to D6S1270 (about 34 cM, 6p21–6q15). This region spans the centromere. Two peaks were observed, one at 6p21 and the other between 6q12.1 and 6q14.2. Dinucleotide repeats, CA and GT in 6q, characterize both peaks.

Wang et al. (1995) showed linkage at 6p22 in families of Irish descent. The follow-up study by Straub et al. (1995) employed a larger, expanded sample and resulted in weaker evidence for linkage. In the same year, Antonarakis et al. (1995) and Schwab et al. (1995) reported linkage at 6q22–24. Maziaide et al. (1997) failed to replicate this finding in a Canadian sample. Levinson et al. (2000), in a multicenter linkage study, examined candidate regions on 5q, 6q, 10p, and 13q. The strongest evidence for linkage

<table>
<thead>
<tr>
<th>Marker</th>
<th>Location</th>
<th>Factor</th>
<th>LOD score</th>
</tr>
</thead>
<tbody>
<tr>
<td>D6S426</td>
<td>6p21</td>
<td>Disorganized</td>
<td>3.34</td>
</tr>
<tr>
<td>D6S1053</td>
<td>6q11.2–6p12.1</td>
<td>Disorganized</td>
<td>3.00</td>
</tr>
<tr>
<td>+ 5 cM</td>
<td>—</td>
<td>Disorganized</td>
<td>3.44</td>
</tr>
<tr>
<td>D6S445</td>
<td>6q14.2</td>
<td>Disorganized</td>
<td>3.05</td>
</tr>
<tr>
<td>D6S1270</td>
<td>6q15</td>
<td>Disorganized</td>
<td>3.03</td>
</tr>
<tr>
<td>D9S288</td>
<td>9pter</td>
<td>Disorganized</td>
<td>3.14</td>
</tr>
<tr>
<td>D12S1300</td>
<td>12q22</td>
<td>Negative</td>
<td>2.97</td>
</tr>
<tr>
<td>PAH-1-H</td>
<td>12q23.2</td>
<td>Negative</td>
<td>2.12</td>
</tr>
<tr>
<td>D20S478</td>
<td>20q11.2</td>
<td>Disorganized</td>
<td>2.79</td>
</tr>
<tr>
<td>+3 cM</td>
<td>—</td>
<td>Disorganized</td>
<td>3.04</td>
</tr>
</tbody>
</table>

LOD, log-odds for linkage.
was observed on 6q. This replicated the earlier findings of Levinson et al (1998) and Martinez et al (1999). Cao et al (1997) first reported linkage at D6S474 and later at D6S424 with a larger sample. Marker D6S424 is located at 6q16.6 and is characterized by both CA and GT repeats. Martinez replicated the finding at this marker. Marker D6S474 is approximately 13 cM distal to the second peak (D6S445) observed in the present work. D6S474 is approximately 27 cM distal to these findings.

Kaufmann et al (1998) reported linkage to the DSM-III-R qualitative phenotype in these data in a broad region extending from D6S445 to D6S310. The highest LOD score in the present analyses is at D6S445 for the Disorganized phenotype.

Boin et al (2001) showed an association between a polymorphism in tumor necrosis factor A (TNFA) and schizophrenia in a case-control study. TNFA is located between 34.6 and 44.2 Mb. Our finding at marker D6S426 is located 8–20 Mb distal to TNF.

Schwab et al (2000) described a region extending from D6S271 to D6S1613. Our finding on 6q overlaps this region. Our results are proximal to most of those reported on 6q and span the centromere and include some prior findings on 6p.

Although we report five positive findings, some of which are consistent with prior research, we failed to replicate prior findings on chromosomes 1, 5, 8, 10p (in this sample), 11, 13, 15 (also in this sample), 18, or 22.

We acknowledge for each trait, our sample size was smaller than would be optimal; however, our use of quantitative traits allowed the inclusion of all informative siblings in each analysis. Our quantitative traits are second-order factor scores. It is possible that the simpler second-order structure obfuscates meaningful phenotypic differences found in traits derived from the first-order analyses. It does, however, retain the advantage of requiring fewer statistical tests (and hence greater control of Type I error) than the 12-factor first-order model.

Although the SANS and SAPS are generally accepted clinician ratings for symptoms associated with the negative and positive symptoms of schizophrenia, they provide only one source of diagnostic data. A multi-assessment system including longitudinal clinical assessments, family informants, and medical records would likely provide a better measure of the phenotype.

It is likely that the genes conferring susceptibility to schizophrenia interact with one another, either in an epistatic or polygenic manner. In the same sample, Freedman (2001a) used a two trait-locus linkage analysis and showed that individuals inherit liability for the disease through at least two different loci. The power to detect any one of these genes of small effect was low in this sample. A larger sample would have given us more power for a genome scan of this complex disorder. Thus, our work cannot exclude any regions of the genome as potentially harboring susceptibility genes for schizophrenia.

Although our finding of both suggestive linkage and association with the gene encoding phenylalanine are intriguing, our LOD score and p value are not high enough for us to assert with confidence that these findings are true positives.

Despite these limitations, our findings are worth further scrutiny in larger samples using quantitative traits of schizophrenic psychopathology.

Preparation of this article was supported in part by the National Institute of Mental Health (NIMH) Grant R01 MH59624 to Dr. Ming T. Tsuang. The original data collection and genotyping was supported by NIMH grants 1 R01 MH41874–01, 5 U01 MH46318, and 1 R37 MH43518 to Dr. Ming T. Tsuang, U01 MH46289 to Dr. Charles Kaufmann, U01 MH46276 to Dr. C. Robert Cloninger, and R01 MH44292 to Drs. Jurg Ott and Charles Kaufmann, and contracts with Millennium Pharmaceuticals (which are sponsored by Wyeth-Ayerst Pharmaceuticals). Some of the results in this article were obtained by using the program package S.A.G.E., which is supported by a U.S. Public Health Service Resource Grant (1 P41 RR03655) from the Division of Research Resources.

The National Institute of Mental Health Genetics Initiative for schizophrenia is a multi-site study performed by three independent research teams in collaboration with staff from the NIMH. The NIMH Collaborators include David Shore, M.D., Debra Wynne, M.S.W., Steven O. Moldin, Ph.D., Darrell G. Kirch, M.D. (1989–1994), Kate A. Berg, Ph.D. (1990–1994), Nancy E. Maestri, Ph.D. (1992–1994); the NIMH Senior Scientific Consultant is Darrel A. Regier, M.D., M.P.H. The Principal Investigators and Co-Investigators from the three sites are: Harvard University, Boston, Massachusetts (U01 MH46318), Ming T. Tsuang, M.D., Ph.D., D.Sc., Stephen Faraone, Ph.D., and John Pepple, Ph.D.; Washington University, St. Louis, Missouri (U01 MH46276), C. Robert Cloninger, M.D., Theodore Reich, M.D., and Dragan Svrakic, M.D.; Columbia University, New York, New York (U01 MH46289), Charles Kaufmann, M.D., Dolores Malaspina, M.D., and Jill Harkavy Friedman, Ph.D. Blood samples are sent to the NIMH Cell Repository at Coriell Institute for Medical Research. Clinical data are stored in the NIMH Data Management Center at SRA Technologies, Inc (Atlanta, GA).

References