Genomewide Linkage Analyses of Bipolar Disorder: A New Sample of 250 Pedigrees from the National Institute of Mental Health Genetics Initiative


We conducted genomewide linkage analyses on 1,152 individuals from 250 families segregating for bipolar disorder and related affective illnesses. These pedigrees were ascertained at 10 sites in the United States, through a proband with bipolar I affective disorder and a sibling with bipolar I or schizoaffective disorder, bipolar type. Uniform methods of ascertainment and assessment were used at all sites. A 9-cM screen was performed by use of 391 markers, with an average heterozygosity of 0.76. Multipoint, nonparametric linkage analyses were conducted in affected relative pairs. Additionally, simulation analyses were performed to determine genomewide significance levels for this study. Three hierarchical models of affection were analyzed. Significant evidence for linkage (genomewide $P < .05$) was found on chromosome 17q, with a peak maximum LOD score of 3.63, at the marker D17S928, and on chromosome 6q, with a peak maximum LOD score of 3.61, near the marker D6S1021. These loci met both standard and simulation-based criteria for genomewide significance. Suggestive evidence of linkage was observed in three other regions (genomewide $P < .10$), on chromosomes 2p, 3q, and 8q. This study, which is based on the largest linkage sample for bipolar disorder analyzed to date, indicates that several genes contribute to bipolar disorder.

Introduction

Bipolar disorder is a serious mood disorder, in which individuals suffer from episodes of extreme depression and mania. It affects ~1% of the population and is about equally frequent in both sexes. The cost to both society and the affected individual is high: nearly one-half of individuals with bipolar disorder attempt suicide (Jamison 1999). Worldwide, it currently accounts for 14 million years of healthy life lost owing to mortality and disability, nearly as much as schizophrenia (World Health Organization, World Health Report 2002).

Family, twin, and adoption studies have provided evidence for a strong genetic component to bipolar disorder (Taylor et al. 2002). Family studies have consistently demonstrated that the risk for bipolar disorder is elevated among first-degree relatives of bipolar probands (Tsuang and Faraone 1990). Twin studies demonstrate higher concordances for the disorder among MZ twins, as compared with DZ twins, with an estimated heritability >80% (summarized by Craddock and Jones [1999]). Finally, the limited number of adoption studies that have examined bipolar disorder have provided further support for a genetic component to the disorder, as indicated by increased risk of affective disorder in the biological parents, but not the adoptive parents, of adoptees who became affected with bipolar disorder (Mendlewicz and Rainer 1977; Wender et al. 1986). Together, these studies converge to suggest that genetic influences play a large role in the etiology of bipolar disorder.

The substantial evidence for a genetic component to bipolar disorder has launched a considerable effort to identify the gene(s) involved in the disorder. Early studies adopted a single-gene approach, which did not prove...
successful, despite initial enthusiasm (Egeland et al. 1987; Kelsoe et al. 1989). In the majority of families, bipolar disorder is now thought to be influenced by multiple genes, as well as environmental influences. Gene-gene and gene-environment interactions likely further complicate attempts to understand the etiology of this complex disorder. Locus heterogeneity is also thought to be present, in which several distinct genes contribute to the disorder, perhaps even within the same family. Despite these obstacles, several promising loci have been identified by linkage studies. These include regions on chromosomes 4p, 10p, 12q, 13q, 16p, 18p, 18q, 21q, 22q, and Xq (for a more extensive review, see Nurnberger and Foroud [2000]). Some of these chromosomal regions meet strict criteria for significant evidence of linkage (Lander and Kruglyak 1995). However, narrowing the linkage regions and identifying candidate genes have proven difficult, and no genes have yet been conclusively demonstrated to affect risk for bipolar disorder.

The complex etiology of bipolar disorder, and the difficulties inherent in detecting influential genes, led to the creation of the National Institute of Mental Health (NIMH) Genetics Initiative for Bipolar Disorder. The goal of this multisite collaboration was to ascertain and evaluate a large sample of families informative for genetic linkage analyses of bipolar disorder. As part of a four-site collaboration, two independent samples were ascertained and analyzed (NIMH Genetics Initiative Bipolar Group 1997; McInnis et al., in press). The first sample consisted of 540 subjects from 97 families, and the second sample consisted of 353 individuals from 56 families. Genomewide scans in these families have been reported elsewhere (Detera-Wadleigh et al. 1997; Edenberg et al. 1997; Rice et al. 1997; Stine et al. 1997; Dick et al. 2002). In the next stage of the study, the consortium was expanded from 4 to 10 sites. Here, we report analyses from the first sample collected by the 10-site collaboration, consisting of 1,152 individuals from 250 families. These families are completely independent of the previously collected NIMH families with bipolar disorder, and no part of this data set has been previously published. We find two regions with LOD scores exceeding simulated genomewide significance levels ($P < .03$) and several others with LOD scores strongly suggestive of linkage ($P < .10$).

**Materials and Methods**

**Family Ascertainment and Assessment**

Multiplex families with bipolar disorder were ascertained at 10 sites: Indiana University (with satellite sites at University of Louisville and at Wayne State University in Detroit), Johns Hopkins University, the NIMH Intramural Research Program, Rush-Presbyterian Medical Center in Chicago, University of California at Irvine, University of California at San Diego, University of Chicago, University of Iowa, University of Pennsylvania, and Washington University in St. Louis. Specific, uniform ascertainment rules were used at all sites. Families were identified by screening admissions at local treatment facilities or by advertisement through advocacy groups, Web sites, and professional organizations. Participants were asked to give informed consent for interview, a blood specimen for DNA and cell lines, and permission to contact relatives; the study was approved at institutional review boards at each participating site. Families were ascertained when a proband with bipolar I (BPI) affective disorder had a sibling affected with BPI or schizoaffective disorder, bipolar type (SABP). Probands who were the offspring of a bilineal mating (defined as “both parents affected with BPI or SABP”) were not included in genetic analyses. Additional first-degree relatives were ascertained, and, in some cases, families were extended to include affected relatives (with BPI, SABP, BPII [Bipolar II Affective Disorder with Recurrent Major Depression, consisting of at least two episodes meeting criteria, with an intervening period of at least 2 mo of euthymia or a period of hypomania or mania], or UPR [Recurrent Major Depression, also known as “Unipolar Depression, Recurrent type”]). To maintain consistency with the previous NIMH Genetics Initiative bipolar samples, BPI and SABP were diagnosed by use of DSM-IIIIR criteria (1994), and BPII and UPR were diagnosed by use of research diagnostic criteria (Feighner et al. 1972). The vast majority of the sample (93%) self-reported their ethnicity as Caucasian; 3.5% reported their race as African American, and the remaining 3.5% reported other ethnicities.

Subjects were assessed by use of the Diagnostic Instrument for Genetic Studies (Nurnberger et al. 1994). This is a polydiagnostic instrument, developed for the assessment of mood disorders and related conditions. It exhibits excellent test-retest reliability for the disorders of primary interest. Interviewers generally had clinical experience and underwent an on-site training program prior to their involvement in the study (Nurnberger et al. 1994). The Family Instrument for Genetic Studies was also administered. This instrument was used to obtain information on symptoms among relatives. Medical records were obtained from 63.5% of the individuals diagnosed with major affective disorders (BPI, SABP, BPII, and/or UPR). Final diagnoses were made by two independent clinicians, by incorporation of all available information. In the event of a disagreement regarding diagnosis between the two clinicians, a third clinician reviewed the case as a tie breaker. More extensive details on the ascertainment and assessment of families are available in a previous report (NIMH Genetics Initiative Bipolar Group 1997).
Three hierarchical definitions of affected status were employed in this analysis. These definitions are supported by family-study data and were selected to maximize the likelihood that genes contributing to heritable disorders could be identified. BPII and major depression are reported more frequently in relatives of probands with BPI than in relatives of control subjects in multiple family studies that employ direct examination of relatives (see review in Nurnberger and Berrettini [1998]). Schizophrenia is generally found to be genetically related to both schizophrenia and bipolar disorder in family studies, but the bipolar subtype is more closely related to mood disorders (see review in Gershon et al. [1988]). All these disorders appear to be heritable, on the basis of data from twin studies (McCabe 1975; Bertelsen 1979; Kendler et al. 1992, 1993), but variation in severity may be related to the presence or absence of specific genetic factors. Thus, it is generally felt that a complete genetic analysis of family data requires the use of multiple models of affected status. The use of diagnostic hierarchies is not unique to bipolar disorder; they have been used in studies of schizophrenia (Kendler et al. 1996), alcoholism (Williams et al. 1999; Foroud et al. 2000), and many other disorders (Narod et al. 1995; Pankratz et al. 2002).

Model 1 included individuals with BPI or SABP. Model 2 included individuals affected under model 1, plus individuals with the less severe BPII. Finally, model 3, the broadest model of affection, included individuals diagnosed under models 1 and 2, plus individuals with UPR. There were 588 genotyped individuals affected under model 1, 639 genotyped individuals affected under model 2, and 737 genotyped individuals affected under model 3. Table 1 summarizes the number of affected relative pairs of each type for each model of affection. Here, we report results from genomewide linkage analyses of affected relative pairs, using the three models of affection.

Genotyping

The genome scan was performed at the Center for Inherited Disease Research by use of automated fluorescent microsatellite analysis. PCR products were sized on an ABI 3700 Sequencer. The marker set used was a modification of the Cooperative Human Linkage Center version 9 marker set (391 markers, average spacing 9 cM, average heterozygosity 0.76). The error rate, based on 17,707 paired genotypes, was 0.05%. The overall missing data rate for the 471,032 total genotypes was 3.75% per genotype. All genotyping was performed blind to critical status.

The marker genotype data were used to verify the reported family relationships among the subjects by use of the computer program PREST (McPeek and Sun 2000). Pedigrees were checked for non-Mendelian inheritance by use of the programs CRIMAP (Green 1990) and USRM13 (Boehnke 1991). An average of 18 genotypes per autosomal chromosome was removed because of Mendelian inconsistencies. Marker order and map positions were obtained from the Marshfield electronic database.

Statistical Analyses

We employed nonparametric, multipoint methods of linkage analysis, utilizing affected relative pairs. The program Merlin (Abecasis et al. 2002) was used to analyze the extent of allele sharing among all affected relative pairs. All affected individuals in a family were analyzed simultaneously for allele sharing by descent. Marker allele frequencies were estimated by use of the default in Merlin, among all individuals. Merlin produces modified Kong-and-Cox LOD scores and associated single-point P values (Kong and Cox 1997). Merlin was also used to conduct simulation analyses to estimate genomewide significance levels for the study. The method employed performs gene-dropping simulations that replace input data with simulated chromosomes conditional on the family structure, marker spacing, allele frequencies, and missing data patterns of the study’s original input files. One thousand simulations were performed on the data to determine the genomewide significance levels for the study.

Results

The LOD score graphs from the affected-relative-pair analyses, for all chromosomes and all three models of affection, are shown in figure 1. On the basis of simulation studies with our pedigree structures, a LOD score of 2.69 corresponds to a genomewide significance level of .05 in this sample. This threshold is lower than the 3.6 proposed by Lander and Kruglyak for interpreting a LOD score as significant (Lander and Kruglyak 1995). However, their threshold was obtained by simulation of a genome scan that used 100 sibling pairs, with genotype data on an infinitely dense genetic map. Our simulations are specific to the pedigrees and marker map used in our sample. The large sample size and more realistic genetic information under Each Model of Affection

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Model 1 (N = 588)</th>
<th>Model 2 (N = 639)</th>
<th>Model 3 (N = 737)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siblings</td>
<td>317</td>
<td>358</td>
<td>446</td>
</tr>
<tr>
<td>Half siblings</td>
<td>5</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Grandparent-grandchild</td>
<td>8</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>Avuncular</td>
<td>54</td>
<td>72</td>
<td>99</td>
</tr>
<tr>
<td>Cousins</td>
<td>12</td>
<td>21</td>
<td>34</td>
</tr>
</tbody>
</table>

Table 1: Number of Affected Relative Pairs with Genotype and Phenotype Information under Each Model of Affection
Figure 1  LOD score graphs from affected-relative-pair analyses across all chromosomes and diagnostic models (see key in figure). Tick labels on graphs indicate the placement of markers. Horizontal dashed line indicates the genomewide significance level, determined by simulation in this sample.

As measured against the threshold obtained in our study’s simulations, two chromosomal regions met genomewide significance for affected-relative-pair analyses. Chromosome 17 yielded a maximum LOD score of 3.63, for affection model 2, at 126 cM, at the marker D17S928. This was the most distal marker on the map of chromosome 17q. To further evaluate evidence of linkage to this distal region of the chromosome, we also conducted single-point linkage analyses. Although the
LOD scores were expectedly reduced across the chromosome, with single-point mapping, evidence of linkage was still observed at the marker D17S928, with a maximal LOD score of 2.6, under model 2. Chromosome 6 yielded a maximum LOD score of 3.61, at 114 cM, near the marker D6S1021, under model 3. These regions also exceeded the threshold for significance recommended in Lander and Kruglyak’s (1995) guidelines for interpreting linkage results for complex disorders.

Three additional chromosomal regions yielded LOD scores that did not meet genomewide significance but were suggestive of linkage both by our simulations ($P < .10$) and Lander and Kruglyak’s (1995) guidelines. Chromosome 2p yielded a maximum LOD score of 2.62, at 94 cM, near the marker D2S1394, under model 3. Chromosome 3q yielded a maximum LOD score of 2.54, under model 1, at 151 cM, near the marker D3S1764. Finally, chromosome 8q yielded a maximum LOD score of 2.46, under model 1, at 151 cM, near the marker D8S256.
Discussion

This study reports results from a new sample of 250 families with bipolar disorder who have not been previously analyzed. We identified several regions of interest, some of which overlap with and provide further support for previous linkage findings. Our strongest findings were on chromosome 17q, with a LOD score of 3.63, at the marker D17S928, and on chromosome 6q, with a maximum LOD score of 3.61, near the marker D6S1021. The chromosome 17 finding is just distal to an area of possible linkage reported in the previous collection of 153 pedigrees from the NIMH Bipolar Genetics Initiative. That study reported its maximal result on chromosome 17 at the marker D17S1531 (NPL = 2.07; LOD = 1.31) (McInnis et al., in press), using the intermediate disease definition, model 2. Chromosome 17 was also implicated as one of the top regions of interest in a recent genome scan of the Wellcome Trust UK-Irish pedigrees with bipolar disorder (Bennett et al. 2002). The same marker that resulted in the maximum LOD score in our sample yielded a multipoint LOD score of 1.38, with 59.5% allele sharing. Possible candidate genes in this region include phosphodiesterase 6G and pyrroline-5-carboxylate reductase, both known to be expressed in CNS tissue and thyroid hormone–binding protein p55.

Chromosome 6 also yielded evidence of linkage in the previous sample of NIMH pedigrees with bipolar disorder. The results from these two studies are ~25 cM apart. Although the exact location of the peak is not identical, simulation studies have demonstrated that non-parametric analyses of complex disease yield substantial variation in the location estimate of susceptibility loci (Roberts et al. 1999). The maximum NPL score in the original NIMH study of 153 families with bipolar disorder was 2.65, at D6S311, under model 3, with a corresponding LOD score of 1.53 (McInnis et al., in press). As in the current study, all three affection status models demonstrated some evidence of linkage to chromosome 6q. Chromosome 6q also showed nominal significance in the Wellcome Trust UK-Irish pedigrees with bipolar disorder, with their peak LOD score obtained with a marker (D6S434) located 3 cM from our maximum LOD score (Bennett et al. 2002).

Chromosome 8q, which yielded suggestive evidence of linkage in our sample, with a maximum LOD score of 2.46, under model 1, was also a region with $P < .05$ observed under multiple models in the recent meta-analysis of bipolar genome scans (Segurado et al. 2003 [in this issue]). Additionally, the regions on chromosomes 8q and 2p that had suggestive evidence of linkage in this study were among the top regions of linkage in a recent study of 65 pedigrees with bipolar disorder (McInnis et al. 2003). Chromosome 8q yielded a LOD score of 2.1, at D8S258, the same marker yielding a maximum LOD score in the sample reported here. Additionally, chromosome 2 yielded a LOD score of 1.5 at 2p12, the same region implicated in this study.

We pursued a strategy to maximize our ability to detect linkage. This included the use of a large sample of well-characterized families and analysis of a series of hierarchical disease models. We have not explicitly adjusted our threshold of significance to take into consideration the multiple models that were evaluated. However, since the models were nested, it was not surprising that coincident evidence of linkage was typically found for all chromosomal regions nominated in this study. Chromosomal regions in which the broadest model (model 3) gave a higher LOD score than models 1 and 2 (e.g., chromosome 2) may be interpreted as providing greater evidence of a locus contributing to susceptibility to both unipolar and bipolar disorders. In contrast, chromosomal regions in which the broadest model gave a lower LOD score than models 1 and 2 (e.g., chromosome 17) support a locus with more limited effect, contributing primarily to susceptibility for the more severe disease phenotype (i.e., bipolar rather than unipolar disorder).

Several regions that were identified in the previous NIMH study with linkage to bipolar disorder did not emerge with significant linkage in this study. Chromosomes 16p and 20p had the strongest evidence of linkage in the first NIMH study (Dick et al. 2002; McInnis et al., in press); however, there was no evidence of linkage to these regions when the sample reported here was used. Additionally, it is striking that regions that have been implicated by several other research groups and are currently thought to be promising, such as 13q, 22q (Badner and Gershon 2002), and chromosome 18 (Berrettini et al. 1994), did not emerge with significant findings in our study. We are currently pursuing additional analyses using family-specific LOD scores to further examine these and other chromosomal regions. As reported by Suarez et al. (1994), replication of a specific linkage result is considerably more difficult in a second sample, and varying frequencies of genes predisposing or protecting against disease can lead to apparent inconsistency across samples. Therefore, lack of replication of a linkage finding does not necessarily lead to the conclusion that the initial linkage finding was spurious. Rather, ascertainment of an independent sample will often lead to differing population substructures. Thus, although obtaining large sample sizes is of certain importance for enhancing our ability to detect genes involved in complex disorders, doing so may not resolve all of the complications presented by these disorders.

In conclusion, this report presents the results of the most recent sample of pedigrees with bipolar disorder, collected at 10 sites around the United States, as part...
of a collaboration, formed by the NIMH, that has been ongoing for 15 years. It is the largest sample of pedigrees with bipolar disorder collected to date, which allows for the potential identification of genes of small effect that were not previously detectable in smaller samples. We find two chromosomal regions that meet stringent criteria for genomewide significance ($P < .05$) on chromosomes 17q and 6q and three regions with suggestive evidence of linkage ($P < .10$) on chromosomes 2p, 3q, and 8q. This study is the first genome scan that we are aware of that has yielded multiple chromosomal regions that meet genomewide significance criteria for linkage to bipolar disorder. These findings support the widely held belief that bipolar disorder is influenced by multiple genes, and they suggest that, with a large enough sample, it is possible to detect multiple genetic regions of influence. Converging evidence of linkage to these regions from other samples leads us to be hopeful that we are a step closer to identifying susceptibility genes for bipolar disorder.

Acknowledgments

We gratefully acknowledge support from the NIMH (R01s MH59545 [to J.I.N.], MH60068 [to W.B.], MH59567 [to J.R.K.], R0159553 [to W.H.B.]). Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to Johns Hopkins University, contract N01-HG-65403. Preparation of this manuscript was also supported by AA13358 (to D.D.) and AA00285 (to T.F.). We also acknowledge John Hennion, M.S., Jennifer Khalid, R.N., and other interviewers and best-estimate diagnosticians, for subject ascertainment and assessment, as well as Howard Edenberg, Ph.D., for helpful suggestions on a previous draft of this manuscript. We are grateful to the patients and family members who generously participated in this research and to the treatment facilities and other organizations that collaborated with us in identifying families.

Electronic-Database Information

URLs for data presented herein are as follows:

Center for Inherited Disease Research, http://www.cidr.jhmi.edu
(for detailed information on laboratory methods and markers)
Marshfield Center for Medical Genetics, http://research.marshfieldclinic.org/genetics/

References

Gershon ES, DeLisi LE, Hamovit J, Nurnberger JJ Jr, Maxwell ME, Schreiber J, Dauphinais D, Dingman CW, Guroff JJ


Erratum

In the July 2003 issue of the Journal, the article entitled, “Genomewide Linkage Analyses of Bipolar Disorder: A New Sample of 250 Pedigrees from the National Institute of Mental Health Genetics Initiative,” by Dick et al. (73: 107–114), includes the following error in linkage calculations: 12 pairs of MZ twins were inadvertently analyzed as siblings in the data set. Analysis of the corrected data set reduces the evidence of linkage on 6q and 17q from the “genomewide significant” to the “genomewide suggestive” level (maximum LOD 2.2 and 2.4, respectively). However, we note that a combined analysis of all 399 pedigrees with bipolar disorder from the National Institute of Mental Health Genetics Initiative (A. L. Hinrichs, S. Bertelsen, and T. Reich, unpublished data) continues to show genomewide significance for 6q (maximum LOD of 3.8 at 113 cM). A revised table and figure are available at the following Web site: http://ipr.iupui.edu/data/. The authors apologize for this error.