Apparent Replication of Suggestive Linkage on Chromosome 16 in the NIMH Genetics Initiative Bipolar Pedigrees


1Indiana University School of Medicine, Indianapolis, Indiana
2Johns Hopkins University, Baltimore, Maryland
3University of Chicago, Chicago, Illinois
4Washington University, St. Louis, Missouri

Analyses of a replication sample of families collected as part of the National Institute of Mental Health (NIMH) Genetics Initiative for bipolar disorder provide further evidence for linkage to a region of chromosome 16. Families who had a bipolar I (BPI) proband and at least one BPI or schizoaffective, bipolar type (SABP) first-degree relative were ascertained for the purpose of identifying genes involved in bipolar affective disorder. A series of hierarchical models of affected status was used in linkage analyses. Initial genetic analyses of chromosomes 3, 5, 15, 16, 17, and 22, completed at Indiana University in 540 subjects from 97 families, suggested evidence of linkage to chromosomes 5, 16, and 22 [Edenberg et al., 1997: Am J Med Genet 74:238–246]. Genotyping was subsequently performed on these chromosomes in a replication sample of 353 individuals from 56 families. Nonparametric linkage analyses were performed using both affected relative and sibling pair methods. Analyses in the new sample on chromosome 16, using the broadest model of affected status, corroborate previously reported suggestive linkage to the marker D16S2619. Combining the initial and replication samples further increased the evidence of linkage to this region, with a peak lod score of 2.8.

KEY WORDS: bipolar disorder; genome scan; nonparametric linkage analysis; relative pair analysis; sib pair analysis

INTRODUCTION

Approximately 1% of the population has bipolar affective disorder, which causes severe mood disturbance; affected individuals typically experience both depressive and manic states. Family, twin, and adoption studies provide evidence for a significant genetic component in bipolar disorder [Craddock and Jones, 1999]; however, the underlying genetic architecture remains unknown. It is possible that both epistasis and genetic heterogeneity contribute to disease susceptibility, complicating efforts to identify and replicate linkage to any particular chromosomal region. Despite this, several promising areas of linkage have been identified; these include regions on chromosomes 4p, 10p, 12q, 13q, 16p, 18p, 18q, 21q, 22q, and Xq [for a complete review, see Nurnberger and Foroud, 2000].

As part of the National Institute of Mental Health (NIMH) Genetics Initiative, a four-site collaboration was initiated to collect the largest sample of uniformly ascertained and assessed families informative for genetic analyses of bipolar disorder [NIMH Genetics Initiative Bipolar Group, 1997]. A genome screen was conducted on an initial sample of 540 individuals from...
97 families. Three hierarchical definitions of affected status were employed: model 1 included individuals with bipolar I (BPI) or schizoaffective disorder, bipolar type (SABP); model 2 included individuals affected under model 1, plus individuals with the less severe bipolar II (BPII) with recurrent depression; and model 3, the broadest model of affected status, included individuals diagnosed under models 1 and 2 plus individuals with unipolar recurrent depression (UPR). Using these nested models of disease definition, the strongest linkage findings observed in the initial sample were on chromosomes 1, 6, 7, 10, 16, and 22 [NIMH Genetics Initiative Bipolar Group, 1997]. Here we report the results of additional analyses performed on chromosomes 3, 5, 15, 16, 17, and 22 in a replication sample.

MATERIALS AND METHODS

Family Ascertainment and Assessment

The NIMH Genetics Initiative for bipolar disorder consisted of four sites at the time these data were collected: Indiana University, John Hopkins University, the NIMH Intramural Research Program, and Washington University at St. Louis. Multiplex bipolar families were ascertained at each of the four sites. The replication sample employed identical ascertainment and assessment strategies as the initial sample. Families were identified by screening admissions at local treatment facilities. They were admitted into the study if they met the following criteria: 1) the proband identified at the treatment facility must have BPI and 2) a second affected first-degree family member with either BPI or SABP must be available. Probands who were the offspring of a bilineal mating (defined as both parents affected with BPI or SABP) were not enrolled into the study.

Subjects were assessed using the Diagnostic Instrument for Genetic Studies (DIGS) [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 disorders of primary interest. Interviewers generally had some clinical training, and underwent an on-site training program before their involvement in the study; details regarding training were described previously [Nurnberger et al., 1994]. The Family Instrument for Genetic Studies (FIGS) was also administered. This instrument was used to obtain information on symptoms among relatives. Medical records were requested for individuals who reported some psychiatric treatment. Final diagnoses were made by two independent clinicians, incorporating 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 the preliminary report of the initial sample [NIMH Genetics Initiative Bipolar Group, 1997].

To select the families and samples to be included in the replication sample, all available families not previously genotyped were reviewed. Families were selected for genotyping if they included at least one sibling pair considered affected under any model of diagnosis. Other affected family members were included whenever they were available. When possible, intervening biological relatives of the affected individuals were also genotyped to increase the power to reconstruct identity-by-descent allele sharing among affected individuals. Parents of affected siblings pairs were genotyped whenever they were available to improve the ability to estimate identity-by-descent marker allele sharing between siblings; when parents were unavailable, unaffected siblings were genotyped. The replication sample included 353 individuals from 56 families. It consisted of 143 individuals with a model 1 diagnosis, comprising 98 relative pairs and 59 sibling pairs (29 sibling pairs with both parents genotyped). An additional 41 individuals had a BPII diagnosis, yielding 191 relative pairs and 119 sibling pairs (58 with both parents genotyped) affected under model 2. Finally, 44 individuals had a UPR diagnosis, forming 302 affected model 3 relative pairs and 182 affected model 3 sibling pairs (81 with both parents genotyped).

As with the initial sample, the genotyping and analysis of chromosomes was divided among the four collaborating sites; genotyping for chromosomes 3, 5, 15, 16, 17, and 22 was performed at Indiana University. The methods used for genotyping and Mendelian error detection were as described previously [Edenberg et al., 1997]. The marker allele frequencies were estimated from the sample using the program USERM13 [Boehnke, 1991]; the intermarker distances were estimated using CRIMAP [Green, 1990]. A total of 84 markers were genotyped in the replication sample on these six chromosomes. The average distance between markers was 11.3 cM. The markers averaged 8.3 alleles, for an average heterozygosity of 72.5%.

The evidence of linkage to chromosomes 5 and 16 proved to be the most promising in both the initial and replication samples. To further evaluate these putative susceptibility loci, and maximize the power of the available data, additional analyses were performed analyzing the initial and replication samples jointly. The combined sample included nearly 900 genotyped individuals from 153 families. For chromosomes 5 and 16 only, the marker allele frequencies and intermarker distances were re-estimated using the data from the combined sample, using the methods described above.

Statistical Methods

We employed nonparametric, multipoint methods of linkage analysis, utilizing both affected relative pairs and affected sibling pairs. Our primary method of analysis was affected relative pairs. The program Genehunter [Kruglyak et al., 1996] was used to analyze the extent of allele sharing among all affected relative pairs. Rather than utilizing pairwise comparisons of affected individuals in a family, the more powerful “all” function in Genehunter was used to examine all affected individuals in a family simultaneously for allele sharing by descent. Because Genehunter has been shown to be overly conservative [Kong and Cox, 1997], lod scores
were computed using the modified program proposed by Kong and Cox [1997], Genehunter Plus. To further corroborate potential regions of linkage, the program ASPEX [Hinds and Risch, 1999] was used to conduct analyses limited to affected sibling pairs. The linkage analyses were performed using only those affected siblings with both parents genotyped, (sib_ibd), and subsequently the analyses were repeated with all affected siblings included regardless of parental genotyping (sib_phase). Limiting the analyses to only those affected sibling pairs with genotyped parents (sib_ibd) allows for unambiguous estimation of identity-by-descent (IBD). While this type of analysis results in greater accuracy in the estimate of marker allele sharing among affected siblings, this occurs at the expense of a reduction in the sample size. To complement these analyses of the most informative subset of the genetic data, additional linkage analyses were performed utilizing information from all available sibling pairs regardless of the availability of parental genotyping. While the sample size is maximized in these analyses, estimates of marker allele sharing are often based on identity-by-state (IBS) rather than IBD, and are therefore less powerful.

A consideration in the analyses was the weighting of sibships consisting of more than two affected siblings. The analysis may be performed forming all possible pairs of affected siblings \( n(n-1)/2 \), where \( n \) = number of affected siblings in a nuclear family); however, the linkage test statistic may be biased by families with large affected sibships. Conversely, using only independent sibling pairs (\( n-1 \)) may be an overly conservative technique. To ensure the consistency of all linkage results, affected sibling pair linkage analyses were performed using all possible sibling pairs, as well as only the independent sibling pairs. Therefore, in total, four different sibling pair methods of analysis were conducted. For clarity of presentation, we only include results from analyses of all possible pairs here; across all chromosomes, results from analyses using only independent pairs were consistent with results from all possible pairs, unless otherwise indicated.

RESULTS

Table I shows chromosomal regions yielding multipoint lod scores > 1.5 in the replication sample using the affected relative pair method of analysis. For these regions, the lod scores obtained using affected sibling pairs are also shown. On chromosome 5, analyses using affected relative pairs yielded a maximum lod score of 1.7 for model 3 near the marker D5S1469. Sibling pair analyses further supported evidence for linkage in this region, with a maximum lod score of 2.8 for model 3 at marker D5S207, the marker adjacent to D5S1469. Model 2 also yielded evidence of linkage to this region across several methods of analysis; a maximum lod score of 1.9 was obtained at marker D5S207 when affected sibling pairs with genotyped parents were analyzed. In the initial sample of NIMH bipolar pedigrees, there was evidence of increased allele sharing across all models of affected status at marker D5S1456 [Edenberg et al., 1997]; this region is approximately 20–40 cM distal to the linked region in the replication sample. Therefore, we combined the initial and replication samples for joint analysis to further explore whether these two reports of linkage represent the same underlying susceptibility locus. Joint analyses suggested that this is not the case. The evidence for linkage to chromosome 5 decreased in the combined sample, with lod scores < 1.0 across the entire chromosome using affected relative pairs. Similarly, the evidence for linkage to marker D5S207 decreased in the combined sample using the affected sibling pair methods, with a maximum lod score of 2.0 obtained using the more conservative sample of sibling pairs with genotyped parents. However, the evidence for increased allele sharing among affected siblings at marker D5S1456 was somewhat enhanced in the combined sample as compared to the initial sample. For model 3, a maximum lod score of 1.6 was obtained using the sib_ibd method near marker D5S1456, with 59% allele sharing; this is increased from the 55% allele sharing observed at this marker in the initial sample [Edenberg et al., 1997].

On chromosome 16, a maximum lod score of 2.1 was obtained in the replication sample at marker D16S2619 for model 3 using affected relative pairs (Fig. 1). Lod scores of 2.3 were also obtained using the sib_ibd method at markers D16S2619 and D16S749. Because there was also significant evidence of allele sharing at marker D16S2619 in the initial sample, we combined the initial and replication samples for joint analysis. Using affected relative pairs, a region of nearly 25 cM had four markers (D16S748, S16S2619, D16S764, and D16S749) that all yielded lod scores above 2.0 for model 3 (Fig. 2). The highest lod scores occurred at markers D16S748 (lod = 2.6) and D16S749 (lod = 2.8). All methods of sibling pair analysis also yielded evidence of increased allele sharing in this region, although, as in the replication sample, the exact position

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Model</th>
<th>Lod score</th>
<th>Location</th>
<th>Sib IBD</th>
<th>Location</th>
<th>Sib Phase</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>2</td>
<td>1.64</td>
<td>157cM</td>
<td>1.89</td>
<td>64.8</td>
<td>150cM</td>
<td>1.23</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>1.69</td>
<td>157cM</td>
<td>2.83</td>
<td>65.8</td>
<td>150cM</td>
<td>1.03</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>2.13</td>
<td>36cM</td>
<td>2.28</td>
<td>62.4</td>
<td>36cM</td>
<td>1.69</td>
</tr>
</tbody>
</table>

All possible pairs.
and magnitude of the maximum lod score varied between methods. The highest lod score obtained in sibling analyses (2.1) was at marker D16S748, using the sib phase method for model 3.

Chromosomes 3, 15, 17, and 22 did not yield any regions with lod scores ≥1.5 in the replication sample using the affected relative pair method. On chromosome 17, a lod score of 2.2 was obtained at marker D17S1294 for model 3 using the sib ibd method. However, sibling pair analyses of only independent pairs did not yield evidence of linkage to this area.

**DISCUSSION**

Analyses from the replication sample and combined NIMH data sets provide further evidence for a susceptibility locus for affective disorders on 16p. In the initial sample, there was significantly increased allele sharing at marker D16S2619 and the flanking markers for model 2 (BPI, SABP, or BPII) [Edenberg et al., 1997]. Significantly increased allele sharing was also found at the nearby marker D16S749 for model 3 (model 2 plus UPR) [Edenberg et al., 1997]. Affected relative
pair methods in the initial sample yielded a lod score of 1.7 at this marker for model 3 [Foroud et al., 2000]. Despite the modest size of the replication sample, it provides further evidence of linkage to these markers using both affected relative and sibling pair methods. Maximizing the available data, the combined sample increased the evidence of linkage with the maximum lod score approaching 3.0; four markers in this region exhibited lod scores greater than 2.0. Although none of the samples individually reached the level of significance suggested by Lander and Kruglyak [1995] for a genome scan, we are encouraged by the consistency of evidence for linkage to this region across the samples.

The strongest evidence of linkage in this region in the initial sample was for model 2, whereas the replication and combined samples provide the strongest evidence of linkage for model 3, the broadest definition of affection. This more recent evidence suggests that this region may contain a gene conferring susceptibility to affective disorders in general, rather than to the more specific, severe diagnosis of bipolar disorder.

Chromosome 16 was first implicated as playing a potential role in bipolar disorder by Ewald et al. [1995]. Using data from two large Danish families, a lod score of 2.5 was obtained under a recessive model of inheritance at marker D16S510 [Ewald et al., 1995]. Subsequently, parametric analyses on a Costa Rican family segregating for bipolar disorder yielded evidence of linkage at D16S521, approximately 10 cm distal to the Ewald finding [McInnes et al., 1996]. The region of linkage reported here is approximately 15–30 cm proximal to the Ewald linkage, according to the Marshfield Genetic Map; however, it is quite possible that these linkages all represent the same underlying genetic factor. 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].

More recently, a lod score > 3.0 has been reported on chromosome 16 in 40 Finnish families with bipolar disorder using a dominant model of inheritance [Ekholm et al., 2001]. The marker yielding linkage, D16S769, is approximately 10 cm more centromeric than the peak lod score obtained in the NIMH pedigrees according to the Marshfield Genetic Maps. No markers were genotyped in the NIMH pedigrees in the exact location of linkage with the Marshfield Genetic Map places these markers in the same region where we found increased allele sharing in the original report [Edenberg et al., 1997]. Although this difference is within the resolution of these mapping techniques, in this case, combining the data sets for analysis suggested that these two reports of linkage do not represent the same underlying susceptibility locus. Sibling analyses in the combined sample increased the evidence for linkage to the region reported in the initial sample, but decreased the evidence for linkage to the region implicated in the replication sample. Despite this, the region near marker D5S207, with linkage in the replication sample, remained the strongest region of linkage in the combined sample. A recent report using Monte Carlo Markov Chain methods to analyze a complex, multigenotypical Costa Rican pedigree segregating for severe bipolar disorder also found evidence of significantly increased allele sharing at five consecutive markers on 5q [Garner et al., 2001]. This linkage had not been detected in prior parametric analyses of the Costa Rican family [McInnes et al., 1996]. The Marshfield Genetic Map places these markers in the same region where we find increased allele sharing in the NIMH replication sample. This region has also been linked to bipolar disorder in both parametric and nonparametric analyses of five related pedigrees from the Sanguenay Lac-St. Jean region of Quebec [Shink et al., 1998]. Finally, a previous report also presented evidence for linkage to D5S62, near the GABRA1 gene [Coon et al., 1993]. We specifically tested a microsatellite marker at GABRA1 and did not find evidence for linkage (lod < 1.0), although there was increased allele sharing.
at nearby markers more centromeric than GABRA1. Thus, this region on chromosome 5 remains of interest, although the evidence for linkage is ambiguous. It is possible that several genes of very small effect are contained in this region.

It is interesting that, in the initial sample, the highest lod score was obtained on chromosome 22, near marker D22S533 (lod = −2.5 for model 3). There was no evidence of linkage to chromosome 22 in the replication sample using any method of analysis. Additionally, a region of chromosome 3 that showed excess allele sharing in the initial sample, at markers D3S2405 and D3S3038, did not show evidence of linkage in the replication sample. There are several reasons why reports of linkage may vary between samples; these include genetic heterogeneity, incomplete penetrance, and the presence of phenocopies. Additionally, as reported by Suarez et al. [1994], replication of a specific linkage result is considerably more difficult in a second sample, requiring a substantially larger sample size than was used in the original sample. Therefore, it is possible that inconsistency across samples is the result of the varying frequency of genes predisposing or protecting against disease.

In conclusion, we have corroborating evidence of linkage to chromosome 16 in a second sample of families collected as part of the NIMH Genetics Initiative on bipolar disorder. The initial and replication samples both demonstrate increased allele sharing at marker D16S2619 and flanking markers. Combining these samples further increased the evidence of linkage in this region, with a maximum lod score of 2.8 for the broadest model of affected status. Family ascertainment is ongoing for a new sample of bipolar families, collected as part of an eight-site collaboration across the United States. Genotyping is currently underway for 1272 individuals from 271 families, and a replication sample of the same size is scheduled for genotyping the following year. These datasets should be informative in further investigating, and localizing, the potential susceptibility gene on 16p.

REFERENCES


