

Published in final edited form as:

Mol Psychiatry. 2012 July ; 17(8): 818–826. doi:10.1038/mp.2011.89.

Genome Wide Linkage Analysis of 972 Bipolar Pedigrees Using Single Nucleotide Polymorphisms

Judith A Badner¹, Daniel Koller², Tatiana Foroud², Howard Edenberg^{2,3}, John I Nurnberger Jr⁴, Peter P Zandi^{5,6}, Virginia L. Willour⁵, Francis J McMahon⁷, James B Potash⁵, Marian Hamshere⁸, Detelina Grozeva⁸, Elaine Green⁸, George Kirov⁸, Ian Jones⁸, Lisa Jones⁹, Nicholas Craddock⁸, Derek Morris¹⁰, Ricardo Segurado¹⁰, Mike Gill¹⁰, Dessa Sadovnick¹¹, Ronald Remick¹², Paul Keck¹³, John Kelsoe¹⁴, Muhammad Ayub¹⁵, Alan MacLean¹⁶, Douglas Blackwood¹⁶, Chun-Yu Liu¹, Elliot S Gershon¹, William McMahon¹⁷, Gholson Lyon¹⁷, Reid Robinson¹⁷, Jessica Ross¹⁸, and William Byerley¹⁸

¹Department of Psychiatry, University of Chicago, Chicago, IL USA ²Department of Medical and Molecular Genetics, Indiana University School of Medicine ³Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN USA ⁴Institute of Psychiatric Research, Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN USA ⁵Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine and Bloomberg School of Public Health, Baltimore, MD USA ⁶Department of Mental Health, Johns Hopkins University School of Medicine and Bloomberg School of Public Health, Baltimore, MD USA ⁷Unit on the Genetic Basis of Mood and Anxiety Disorders, Mood and Anxiety Disorders Program, National Institute of Mental Health, National Institutes of Health, US Dept of Health and Human Services, Bethesda, MD, USA ⁸MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK ⁹Neuropharmacology & Neurobiology, University of Birmingham, Birmingham, UK ¹⁰Trinity College, Dublin, Ireland ¹¹Department of Medical Genetics, University of British Columbia, Vancouver, CA ¹²Department of Psychiatry, St. Paul's Hospital, Vancouver, CA ¹³Department of Psychiatry, University of Cincinnati, Ohio, USA ¹⁴Department of Psychiatry, University of California San Diego, La Jolla, CA ¹⁵School of Medicine and Health, University of Durham, Durham, UK ¹⁶Division of Psychiatry, University of Edinburgh, Edinburgh, UK ¹⁷Department of Psychiatry, University of Utah Medical Center, Salt Lake City, UT USA ¹⁸Department of Psychiatry, University of California San Francisco, San Francisco CA USA

Abstract

Because of the high costs associated with ascertainment of families most linkage studies of Bipolar I disorder (BPI) have used relatively small samples. Moreover, the genetic information content reported in most studies has been less than 0.6. While microsatellite markers spaced every 10 centimorgans typically extract most of the genetic information content for larger multiplex families, they can be less informative for smaller pedigrees especially for affected sib pair kindreds. For these reasons we collaborated to pool family resources and carry out higher density genotyping. Approximately 1100 pedigrees of European ancestry were initially selected for study and were genotyped by the Center for Inherited Disease Research using the Illumina Linkage Panel 12 set of 6090 SNPs. Of the ~1100 families, 972 were informative for further analyses and mean information content was 0.86 after pruning for LD. The 972 kindreds include 2284 cases of

Correspondence: Dr. William Byerley, Box 0984 - MGN, 401 Parnassus Ave, LangPorter, University of California, San Francisco, San Francisco, CA. 94143-0984, william.byerley@ucsf.edu.

Supplementary Information accompanies the paper on the Molecular Psychiatry website

BPI disorder, 498 individuals with Bipolar II disorder (BPII) and 702 subjects with Recurrent Major Depression. Three affection status models were considered: ASM1 (BPI and schizoaffective disorder, BP cases (SABP) only), ASM2 (ASM1 cases plus BPII) and ASM3 (ASM2 cases plus Recurrent Major Depression). Both parametric and non-parametric linkage methods were carried out. The strongest findings occurred at 6q21 (Nonparametric Pairs Lod 3.4 for rs1046943 at 119 cM) and 9q21 (Nonparametric Pairs Lod 3.4 for rs722642 at 78 cM) using only BPI and SA, BP cases. Both results met genome-wide significant criteria, although neither was significant after correction for multiple analyses. We also inspected parametric scores for the larger multiplex families to identify possible rare susceptibility loci. In this analysis we observed 59 parametric lods of 2 or greater, many of which are likely to be close to maximum possible scores. While some linkage findings may be false positives the results could help prioritize the search for rare variants using whole exome or genome sequencing.

Keywords

Bipolar disorder; Whole genome linkage; Single Nucleotide Polymorphisms; 6q21; 9q21; 2q12

Introduction

Bipolar I disorder (BPI), also termed Manic-Depressive Illness (MDI), is a complex neuropsychiatric disorder afflicting approximately 1% of the worldwide population (APA 2000). Onset of illness typically occurs in late adolescence to early adulthood when individuals present with acute mania, severe depression or mixed mood states. For most persons the disorder is lifelong with recurring episodes of mania and/or depression. BPI causes substantial morbidity with high costs to the individuals as well as society, and an estimated 10% of those afflicted die by suicide. As mania and depression are seemingly antipodal neurobiological states BPI disorder is one of the most enigmatic of medical illnesses.

Although fundamental neurobiological alterations underlying illness are not yet known, family, twin and adoption studies indicate that BPI is genetically-based with heritability estimates of 80%.¹ These studies also show that BPI displays variable expressivity, reduced penetrance and unknown modes of inheritances. Results of some segregation analyses implicate major loci in a subset of pedigrees.^{2,3} As the disorder is relatively common BPI is likely to have substantial genetic heterogeneity, both allelic and non-allelic (locus). Similar to other complex genetic disorders such as diabetes, autoimmune disorders, cardiovascular disease and Alzheimer Disease, the genetic architecture is likely to be comprised of both common and rare variants.

For most of the past few decades linkage analysis has been the only genome-wide method available for mapping genetic illnesses. It is robust to allelic heterogeneity (although not to locus heterogeneity) and is optimal for mapping rare variants with relatively large effect sizes. Over 25 genome-wide linkage studies of BPI have been completed using mostly small/nuclear pedigrees derived from outbred populations and results are compatible with the presence of large effect size rare alleles with substantial locus heterogeneity.⁴ Meta-analyses have produced some significant findings although these results were based on molecular methods with lower information content than methods currently available.⁵ Genome-wide linkage analyses of uniquely large pedigrees - mostly from isolated populations - have also implicated a number of chromosomal regions,^{6,7,8,9,10,11} many generating lods of 3 or more in individual pedigrees. Firm replication has not been achieved for any of these implicated regions but this is not surprising since exceedingly large samples are required for replication.¹²

Association analysis is more robust than linkage given a wider range of locus heterogeneity (though not allelic heterogeneity) and is a powerful method for mapping common alleles with small effect sizes. By the mid 2000s the completion of the second-generation haplotype map of the human genome (HapMap2) coupled with the advent of low-cost genotyping platforms for interrogation of several hundred thousand single nucleotide polymorphisms (SNPs), made genome-wide association studies GWAS feasible. GWAS has been tremendously successful for mapping common variants for over 40 complex disorders.^{13, 14, 15} While GWAS findings represent a significant biomedical advance, the implicated loci explain no more than 5 to 10% of the genetic liability for any disorder. Most associated SNPs give no insight into disease mechanisms. Initial GWAS for BPI using modest sample sizes (~a few thousand cases and controls) produced some suggestive results. Subsequent meta-analyses (comprised of several thousand cases and controls) yielded significant evidence for association to a few loci including *CACNA1C* and *ANKK3*.^{16, 17, 18, 19, 20, 21, 22} These results represent a significant advance in unraveling the genetics of bipolar disorder, however, the identified loci to date explain only 1 to 2% of disease liability.

In contrast to GWAS, sample sizes for linkage studies have been comparatively small, due in part to the high costs and effort necessary for the ascertainment of multiplex pedigrees (versus individual cases and controls). As a result, almost all linkage studies have been significantly underpowered. In addition, the genetic information content for the vast majority of linkage studies has been under 60%. For these reasons we established a collaboration to pool family resources to carry out higher density genotyping. We report genome-wide linkage analyses for 972 pedigrees genotyped with the Illumina 6K SNP array. The families reported here include 2284 individuals with BPI disorder (140 are schizoaffective, bipolar type -(SABP)), 498 with BPII disorder and 702 with Recurrent Major Depression.

Materials and Methods

Subjects

The sample is described in Supplementary Table 1. For this study, only European-American samples were genotyped. Three different affection status models (ASMs) are used: ASM1: Bipolar I and SABP; ASM2: ASM1 and bipolar II; and ASM3: ASM2 and recurrent major depression. For each model, only individuals without significant mood, anxiety, or psychotic disorders were coded as unaffected. The remaining individuals were coded as unknown. Supplementary Table 2 lists the number of relative pairs that are concordantly affected or unaffected, or discordant for each of the 3 ASMs. Supplementary Table 3 shows the distribution of the number of affected individuals per pedigree.

All subjects underwent semi-structured interviews and hospital records were obtained where available. All information for each subject was evaluated by two experienced clinicians whom made consensus diagnoses. When the two clinicians disagreed, a third clinician derived the final diagnoses. Individuals from the NIMH Genetics Initiative were interviewed using the Diagnostic Interview for Genetic Studies (DIGS).²³ DSM-III-R criteria were used for BPI diagnoses and Research Diagnostic Criteria (RDC) were used for BPII and Recurrent Major Depression Diagnoses.²⁴ Subjects ascertained at Cardiff University and Trinity College Dublin were interviewed using the Schedule for Clinical Assessment in Neuropsychiatry²⁵ and DSM-IV criteria were applied for all diagnoses. Individuals identified by Johns Hopkins University (JHU) were assessed either using the Lifetime Version of the Schedule for Affective Disorders and Schizophrenia (SADS-L) or the DIGS.²⁶ Research Diagnostic Criteria were used for all JHU diagnoses. Individuals evaluated by the University of California San Diego (UCSD) were interviewed using the

Structured Clinical Interview for DSM-III-R (SCID) except those derived from the first family ascertained for which the SADS-L was used.²⁷ The UCSD Diagnoses were made using DSM-III-R criteria, modified to require a 2-day minimum duration for hypomania. Individuals at Edinburgh were assessed using the SADS-L and DSM-III-R criteria were used.²⁸ NIMH-Intramural/University of Chicago families were assessed using SADS-L.²⁹ Families from the UCSF/Utah site were interviewed using the SADS-L and RDC were used for all diagnoses.³⁰

Genotyping

Genotyping was performed by the Center for Inherited Disease Research (CIDR) using the Illumina Linkage Panel 12 set of 6090 SNPs. Five thousand six hundred seventy (5670) autosomal and X-linked SNPs passed CIDR's quality control (QC). The investigators assessed the SNPs for missingness, allele frequency and Hardy-Weinberg equilibrium using PLINK (<http://pngu.mgh.harvard.edu/purcell/plink/>) analysis.³¹ Mendelian errors were assessed by both PLINK and Pedcheck (<http://watson.hgen.pitt.edu/register/docs/pedcheck.html>) analyses,³² and if there were errors in a pedigree, the SNP was set to missing for all members in the pedigree. Unlikely recombinants were assessed by Merlin (<http://www.sph.umich.edu/csg/abecasis/merlin/index.html>) analysis³³ and problematic genotypes were set to missing. Five thousand six hundred forty two (5642) autosomal and X-linked SNPs were included in the linkage analysis. Missingness was less than 3% for SNPs and less than 2% for genotyped individuals, frequency was at least 1%, and Hardy-Weinberg Equilibrium p -value > 0.001 for these SNPs.

Pedigree relationship checks

PLINK analysis was used to identify unexpected relatedness between pedigrees, using Identity By Descent (IBD) scores across all the SNPs and gender misspecification by looking at homozygosity of X-linked SNPs. PREST (<http://fisher.utstat.toronto.edu/sun/Software/Prest/>) analysis³⁴ was used to test whether the observed IBD scores were consistent with reported relationships within pedigrees. For the PREST analysis, a subset of 4969 autosomal SNPs with low linkage disequilibrium (LD) (pairwise $r^2 < 0.5$) was used. Eight duplicate pairs were found and any duplicate that was inconsistent with the rest of the pedigree was excluded. In two instances, partial pedigree overlaps were identified between 2 sites (UCSD and JHU, NIMH and UCSF/Utah). These pedigrees were joined, using the observed IBD scores to identify the relationships. If phenotype information for the overlapping samples did not agree between sites, then the individual's phenotype was set to unknown (1/5 overlapping samples). In another pair of pedigrees from 2 sites, a parent in an NIMH pedigree was consistent with being a child in an NIMH-Intramural/Chicago pedigree. These 2 pedigrees were merged. Two UCSF/Utah pedigrees had founders that were siblings and these pedigrees were merged. For 22 samples, genetic gender was not consistent with reported gender. If the IBD scores were otherwise consistent with the reported relationship, the gender was changed; otherwise, the individual was discarded. Pedigree structures were corrected when relationships determined by IBD were unambiguous; otherwise, samples were discarded. PLINK IBD and PREST analyses were performed after modifications to assess the accuracy of the pedigree relationships.

Ancestry Analysis

Principal Components Analysis, as implemented in EIGENSOFT,³⁵ was used to identify subjects of non-European-American ancestry. The genotype data for the founders of each of the pedigrees was combined with genotypes from the HapMap sample (founders only) for the CEPH, Yoruban, Japanese, and Chinese sample. HapMap genotypes that did not overlap with the linkage sample were discarded. The first 2 components of the PCA analysis were

plotted (Supplementary Figure 1). None of the samples clustered with the Yoruban or Asian samples and almost all of the samples clustered with the CEPH sample. Individuals from 10 families (4 NIMH, 2 JHU, 3 Cardiff, 1 UCSD) were identified as outliers and these 10 families were discarded from the analysis.

Linkage Analysis

Parametric and non-parametric analyses were performed for 3 affection status models (ASMs) using two different methods of analysis. All SNPs were analyzed with Merlin version 1.1.2³³ using the correction for LD by identifying clusters of SNPs with pairwise LD of at least 0.1. Large pedigrees were trimmed using PedShrink software (http://mayoresearch.mayo.edu/schaid_lab/software.cfm), which trims pedigrees to a specified bit size and prioritizes trimming uninformative individuals. All informative individuals were analyzed with MORGAN version 3.0³⁶ (<http://www.stat.washington.edu/thompson/Genepi/pangaea.shtml>) which uses Markov chain Monte Carlo methods for linkage mapping in large pedigrees. A subset of 4679 SNPs with pairwise $r^2 < 0.1$ was used in the MORGAN linkage analysis.

Parametric heterogeneity (HLOD) scores under dominant model (risk allele frequency 0.01, penetrances 0.01, 0.70, 0.70) and recessive model (risk allele frequency 0.1, penetrances 0.01, 0.01, 0.70) were calculated in Merlin and MORGAN/lm_markers. These models are powerful for rare, relatively highly penetrant susceptibility alleles. Nonparametric linkage analyses were performed using 1) the S_{All} statistic under the exponential model³⁷ implemented in Merlin which can be powerful in datasets that show very strong linkage signals or which include larger pedigrees; 2) the S_{Pairs} statistic under the linear model³⁷ implemented in Merlin, which may be more powerful for relatively common alleles; 3) the Spairs statistic implemented in MORGAN/lm_ibdtests; and 4) the Slambda statistic implemented in MORGAN/lm_ibdtests which incorporates affected and unaffected individuals. For the MORGAN nonparametric tests, the normality-based tests, rather than assessing significance through permutation of affection status, were used as many families did not have unaffected relatives.

Sex-specific genetic maps were obtained from Rutgers Map Interpolator software (<http://compgen.rutgers.edu/old/map-interpolator>) that estimates genetic distances from the physical locations of the SNPs, using the Rutgers combined linkage physical map of the human genome.³⁸

Thresholds for suggestive and significant linkage for each analysis were estimated from the data using autoregressive models to estimate the correlation between standard normal statistics at adjacent map points and then using this correlation to estimate study-specific critical values (http://www.wpic.pitt.edu/wpiccompgen/robust_estimation_genome_scan.htm).³⁹

Covariate Analyses

Information on age of onset, presence of psychosis, presence of mood-incongruent psychosis, and suicide attempts was present for most of the families studied. The effects of these covariates on linkage were analyzed using Ordered Subsets Analysis⁴⁰ as implemented in FLOSS (Flexible Ordered Subset Analysis).⁴¹ Covariate designation was pedigree-specific. For age of onset, pedigrees were classified based on the youngest age of onset within a pedigree. For the other covariates, pedigrees were classified by the presence of at least one individual who was positive for the covariate. Also, the number of affected individuals within a pedigree was analyzed as a covariate as pedigrees with many affected individuals may be segregating for different loci (possibly with rarer, more highly penetrant

alleles) than pedigrees with few affected individuals. Nonparametric linkage (NPL) scores for each pedigree were calculated by Merlin using the S_{pairs} statistic. Families were ordered by covariate scores. Families with the same covariate score had the same rank. Thus, for binary covariates, there were only 2 subsets of families. Multipoint linkage was performed on the subset of families with the k-smallest or k-largest scores. The permutation test compared the maximum ordered subset linkage score for the covariate score by ordering of the families with the maximum ordered subset linkage scores obtained for random orderings of the families. Permutation controls for multiple testing of all regions of a single chromosome but not for multiple chromosomes or multiple covariates.

Age of onset, presence of psychosis, presence of mood-incongruent psychosis, and suicide attempts were also analyzed as phenotypes. For binary covariates, individuals were considered affected if they were positive for the covariate and unknown otherwise. These covariates were analyzed with Merlin using the S_{All} statistic under the exponential model and the S_{pairs} statistic under the linear model. Age of onset was analyzed with variance components analysis as implemented in Merlin. This analysis partitions the variance of the trait into the following components: major locus, a polygenic component and non-genetic variation.

Imprinting Analyses

Imprinting was analyzed using GeneFinder⁴² (<http://people.virginia.edu/~wc9c/genefinder/index.html>) with linkage extensions. This analysis uses generalized estimating equations to test whether sharing of maternal alleles among affected sib pairs is significantly different from sharing of paternal alleles. For this analysis, extended IBD scores were calculated using Merlin.

Results

Nonparametric and parametric linkage analyses

The critical lod scores for suggestive (occurring once by chance in a single genome scan), significant (probability of 0.05 of occurring in a single genome scan by chance), and significant criteria with Bonferroni correction for 24 different analyses, are presented in Supplementary Table 4. Bonferroni correction is likely to be overly conservative as the analyses are not independent of each other. Suggestive criteria scores range from 1.9–2.3, significant criteria scores range from 3.2–3.6, and criteria scores after correcting for multiple testing are 4.5–5.0. Information content was estimated using Merlin's entropy measure, where 1 is completely informative. Mean information content was 0.87 (s.d. 0.026) for all the SNPs and 0.86 (s.d. 0.028) for SNPs after pruning for LD.

Supplementary Figures 2, 3, and 4 show the Merlin results for ASM1, ASM2, and ASM3 respectively. Table 1 presents all results that meet suggestive criteria for significance. Results that met significant criteria are found at 6q21 for ASM1 (Pairs Lod 3.4 for rs1046943 at 119 cM) and 9q21 for ASM1 (Pairs Lod 3.4 for rs722642 at 78 cM). Suggestive results were found in 10 additional regions. No results were significant after correcting for multiple analyses.

The MORGAN linkage results are presented in Supplementary Figures 5, 6, and 7 for ASM1, ASM2, and ASM3 respectively. Table 2 presents all the results that met the suggestive criteria for significance. No results met the significant criteria. With the exception of 15q22 for ASM2 (Slambda lod 3.0 for rs782944 at 60 cM), all of the regions presented in Table 6 also met a minimum suggestive criteria for significance in the Merlin analyses.

Table 3 presents the results from the MORGAN parametric analyses for which the lod score for a single pedigree exceeds 2. While most of these results are likely to be false positives, inspection of results for larger families may suggest regions in which a rare, highly penetrant locus exists, particularly when 3 or more families show elevated lod scores in the same region. Such regions include chromosome 4 (152–214 cM) and chromosome 19 (23–58 cM).

The exploratory covariate analyses performed using FLOSS identified genome-wide significant regions of enhanced linkage for families with more affected individuals, younger ages of onset, and lack of psychosis. However, no results were significant when corrected for the number of phenotypes tested. All genome-wide significant results are presented in Table 4. Families with younger ages of onset demonstrated enhanced evidence of linkage for chromosome 2 at 118 cM (ASM2 – Base NPL 3.24, Subset NPL 5.18, $p=0.0007$, 429/785 families). Linkage to chromosome 8 at 82 cM was enhanced in families without psychosis (ASM3 – Base NPL 2.02, Subset NPL 3.76, $p=0.0015$, 173/943 families). Families with a larger number of affected individuals showed increased evidence for linkage on chromosome 12 at 97 cM (ASM2 – Base NPL 2.40, Subset NPL 4.36, $p=0.0015$, 27/872 families).

The results for the exploratory analyses using the covariates as phenotypes include: 1) psychosis showed a Exp lod score of 2.23 on chromosome 12 at 122 cM, an Exp lod score of 2.27 on chromosome 15 at 23 cM, a Pairs lod of 2.27 on chromosome 18 at 64 cM and a Pairs lod of 2.1 on chromosome 18 at 83 cM and 2) mood-incongruent psychosis showed a Pairs lod of 2.03 on chromosome 18 at 33 cM. Neither analysis met genome-wide significance. Finally, there was no evidence of imprinting from the GENEFINDER analyses (results not shown).

Discussion

The strongest linkage findings occurred at 6q21 (Pairs Lod 3.4 for rs1046943 at 119 cM) and 9q21 (Pairs Lod 3.4 for rs722642 at 78 cM) using only BPI and SABP cases, diagnostic classifications that are presumably the least heterogeneous of phenotypes. Both results met genome-wide significant criteria but were not significant when corrected for multiple analyses. These regions did not show enhanced evidence of linkage for any of the subgroups defined by age, psychosis, suicide, or number of affected individuals within a pedigree. The parametric lod scores in these regions are substantially lower than the non-parametric lod scores, suggesting low penetrance as data from unaffected individuals appear to be detracting from the linkage results. The 6q21 linkage region was also identified in a meta-analysis of 11 linkage studies performed by McQueen et al.⁵ Because there is a large overlap between samples included in this study and the meta-analysis, these findings are not independent. Given the overlap between the two studies, it is noteworthy that this study found no evidence of linkage to 8q, a region identified in the meta-analysis. However, the meta-analysis did find suggestive evidence of linkage to 9q at 46–48 cM, which is 30 cM proximal to our finding. Under ASM1 the next two highest results occurred at 2q (Pairs Lod 3.1 at 118) and 17q (Pairs lods 2.7 at 109 cMs). Under ASM2 or ASM3 suggestive results were found in 10 additional regions (Table 1). No results were significant after correcting for multiple analyses.

Association analysis has been performed for the 6q21 region and bipolar disorder by Fan et al.⁴³ 3,047 SNPs were analyzed in a case-control sample of 1064 individuals and a family-based sample of 256 nuclear families. A replication study of 151 SNPs in 622 cases and 1181 controls, involving SLC22A16, DDO, PREP, NT5DC1, GPR6, and the region around rs794854, was also performed. Evidence of association with SLC22A16 in all 3 samples,

although not consistently with the same SNPs, was found. SLC22A16 encodes an organic cation/carnitine transporter. Organic ion transporters transport various medically and physiologically important compounds, including pharmaceuticals, toxins, hormones, neurotransmitters, and cellular metabolites. For 9q21, evidence of association to NTRK2, a tyrosine kinase receptor gene which binds with brain derived neurotrophic factor, was found in a whole genome association study of bipolar disorder African-Americans.²¹ However, there was no association found in the European-American sample in this region.

Copy Number Variants (CNVs) have been analyzed by Zhang et al.⁴⁴, in a sample of individuals drawn from the NIMH families in this study. No evidence of association with a specific CNV was found. However, singleton deletions that were present in cases but not controls were found in 6q21 (1 case), 6q26 (2 cases, no overlap), and 9q21 (1 case). Grozeva et al.⁴⁵ analyzed CNVs in the The Wellcome Trust Case Control Consortium sample, which includes about 50 cases that are in our linkage sample²⁵. Significant evidence of association between bipolar disorder and CNVs was not found in this study. CNVs present in cases but not controls were found on 6q21 (2 cases), 6q22 (2 cases, no overlap), and 9q21 (2 cases, no overlap).

Exploratory covariate analyses found enhanced evidence of linkage in the subset of families with younger age of onset at 2q12, lack of psychosis at 8q13, and larger numbers of affected individuals at 12q21. Mathieu et al.⁴⁶ found evidence of linkage to 2q14, 10–30 cM away from our finding, in the portion of their sample with age of onset less than 22 years but not in the subsample with older age of onset. The finding on 8q13 is approximate 70 cM proximal to the 8q linkage region reported in McQueen et al.⁵ and thus probably is not the same region. Analyses using covariates as phenotypes found no genome-wide significant findings.

We also inspected scores for the larger multiplex families and observed 59 parametric lod scores of 2 or greater, many of which are close to maximum possible scores (Table 3). While some of these lod scores may represent false positive linkage findings, there were multiple regions in which at least 1 pedigree had a lod score of at least two. In particular chromosome 4q (152–214 cM) and chromosome 19 (23–58 cM) are regions in which 3 or more pedigrees showed elevated lod scores and may represent regions in which a rare, relatively penetrant locus may be segregating in a small proportion of families. The 12q21 area may also represent such a region as there was increased evidence for linkage in families with a larger number of affected individuals. These data could be an invaluable asset for mapping rare variants via genome-wide sequencing. Large families, producing relatively high lod scores, are more likely to contain rare alleles with relatively large effect sizes. Focusing on regions showing elevated lod scores will help reduce the number of rare variants to be examined. In addition, co-segregation of the rare variant may be tested within large families, which will further filter the number of variants for follow-up.

Despite a large sample of 972 pedigrees of European ancestry and almost optimal extraction of genetic information (0.87) there were no linkage findings that met genome-wide significance after correction of multiple testing. However, the genetic models are interdependent and statistical correction may be overly conservative under these circumstances. Thus, some of the linkage findings may represent true positives. A number of factors could explain the inability to obtain robust and replicable findings in this linkage study as well as in the field. First, the assumption that rare large-effect size alleles underlie some of the genetic variation of bipolar disorder may be incorrect. In contrast to linkage GWAS has found evidence for a few common loci that appear to be quite significant. However, these loci explain only a small percent of the genetic variation. Given that both rare and common loci have been found for most complex genetic diseases it is likely that

both rare and common variation predispose to BPI. Second, assuming rare variants exist, it appears that there is a large range of locus heterogeneity and that even large samples may not have sufficient power. Linkage results reported over the last two decades are compatible with this. Under a model of extensive locus heterogeneity uniquely large pedigrees will be invaluable for mapping loci as the ability to ascertain thousands of smaller nuclear/sib pair families are cost prohibitive. For this study, seven separate sites contributed families and it is possible – if not probable – that varying fractions of families linked to particular loci are found in each of the samples. There were also different ascertainment and diagnostic criteria among some of the sites and this may have increased heterogeneity. Another factor that hinders the ability to obtain replicable linkage results is that very large sample sizes are required for replication.¹² Several thousands of pedigrees may be required and it is unlikely that this will ever be possible. While the era of large-scale linkage studies is probably over, linkage analysis of uniquely large pedigrees may still be worthwhile.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institutes of Health (NIH/NIMH) research grant R01-MH077314 (JAB, WB). Wellcome Trust 045267/Z/WRE/MB/JAT (MG, RS)

References

1. Merikangas KR, Low NCP. The epidemiology of mood disorders. *Curr Psychiatry Rep.* 2004 Dec; 6(6):411–421. [PubMed: 15538988]
2. Pauls DL, Bailey JN, Carter AS, Allen CR, Egeland JA. Complex segregation analyses of old order Amish families ascertained through bipolar I individuals. *Am J Med Genet.* 1995 Aug; 60(4):290–297. [PubMed: 7485263]
3. Spence MA, Flodman PL, Sadovnick AD, Bailey-Wilson JE, Ameli H, Remick RA. Bipolar disorder: evidence for a major locus. *Am J Med Genet.* 1995 Oct; 60(5):370–376. [PubMed: 8546148]
4. Serretti A, Mandelli L. The genetics of bipolar disorder: genome 'hot regions,' genes, new potential candidates and future directions. *Mol Psychiatry.* 2008 Aug; 13(8):742–771. [PubMed: 18332878]
5. McQueen MB, Devlin B, Faraone SV, Nimgaonkar VL, Sklar P, Smoller JW, et al. Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. *Am J Hum Genet.* 2005 Oct; 77(4):582–595. [PubMed: 16175504]
6. Blackwood DH, He L, Morris SW, McLean A, Whitton C, Thomson M, et al. A locus for bipolar affective disorder on chromosome 4p. *Nat Genet.* 1996 Apr; 12(4):427–430. [PubMed: 8630499]
7. Adams LJ, Mitchell PB, Fielder SL, Rosso A, Donald JA, Schofield PR. A susceptibility locus for bipolar affective disorder on chromosome 4q35. *Am J Hum Genet.* 1998 May; 62(5):1084–1091. [PubMed: 9545396]
8. Jasinska AJ, Service S, Jawaheer D, DeYoung J, Levinson M, Zhang Z, et al. A narrow and highly significant linkage signal for severe bipolar disorder in the chromosome 5q33 region in Latin American pedigrees. *Am J Med Genet B Neuropsychiatr Genet.* 2009 Oct; 150B(7):998–1006. [PubMed: 19319892]
9. Ekholm JM, Pekkarinen P, Pajukanta P, Kieseppä T, Partonen T, Paunio T, et al. Bipolar disorder susceptibility region on Xq24-q27.1 in Finnish families. *Mol Psychiatry.* 2002; 7(5):453–459. [PubMed: 12082562]
10. Green E, Elvidge G, Jacobsen N, Glaser B, Jones I, O'Donovan MC, et al. Localization of bipolar susceptibility locus by molecular genetic analysis of the chromosome 12q23-q24 region in two pedigrees with bipolar disorder and Darier's disease. *Am J Psychiatry.* 2005 Jan; 162(1):35–42. [PubMed: 15625199]

11. Venken T, Claes S, Sluijs S, Paterson AD, van Duijn C, Adolfsson R, et al. Genomewide scan for affective disorder susceptibility Loci in families of a northern Swedish isolated population. *Am J Hum Genet.* 2005 Feb; 76(2):237–248. [PubMed: 15614721]
12. Suarez, BK.; VEP Hampe, CL. Problems of replicating linkage claims in psychiatry. In: Gershon, ESCC., editor. *Genetic approaches to mental disorders.* American Psychiatric Press, Inc; 1994. p. 23-46.
13. Ioannidis JPA, Thomas G, Daly MJ. Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet.* 2009 May; 10(5):318–329. [PubMed: 19373277]
14. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nat Rev Genet.* 2009 Apr; 10(4):241–251. [PubMed: 19293820]
15. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet.* 2010 Jun; 11(6):415–425. [PubMed: 20479773]
16. Consortium WTCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007 Jun; 447(7145):661–678. [PubMed: 17554300]
17. Ferreira MAR, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet.* 2008 Sep; 40(9):1056–1058. [PubMed: 18711365]
18. Sklar P, Smoller JW, Fan J, Ferreira MAR, Perlis RH, Chambert K, et al. Whole-genome association study of bipolar disorder. *Mol Psychiatry.* 2008 Jun; 13(6):558–569. [PubMed: 18317468]
19. Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, et al. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry.* 2008 Feb; 13(2):197–207. [PubMed: 17486107]
20. Scott LJ, Muglia P, Kong XQ, Guan W, Flickinger M, Upmanyu R, et al. Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proc Natl Acad Sci U S A.* 2009 May; 106(18):7501–7506. [PubMed: 19416921]
21. Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, Berrettini W, et al. Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry.* 2009 Aug; 14(8):755–763. [PubMed: 19488044]
22. Ollila HM, Soronen P, Silander K, Palo OM, Kiesepää T, Kaunisto MA, et al. Findings from bipolar disorder genome-wide association studies replicate in a Finnish bipolar family-cohort. *Mol Psychiatry.* 2009 Apr; 14(4):351–353. [PubMed: 19308021]
23. Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, et al. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch Gen Psychiatry.* 1994 Nov; 51(11):849–859. discussion 863-4. [PubMed: 7944874]
24. Edenberg HJ, Foroud T, Conneally PM, Sorbel JJ, Carr K, Crose C, et al. Initial genomic scan of the NIMH genetics initiative bipolar pedigrees: chromosomes 3, 5, 15, 16, 17, and 22. *Am J Med Genet.* 1997 May; 74(3):238–246. [PubMed: 9184305]
25. Lambert D, Middle F, Hamshere ML, Segurado R, Raybould R, Corvin A, et al. Stage 2 of the Wellcome Trust UK-Irish bipolar affective disorder sibling-pair genome screen: evidence for linkage on chromosomes 6q16-q21, 4q12-q21, 9p21, 10p14-p12 and 18q22. *Mol Psychiatry.* 2005 Sep; 10(9):831–841. [PubMed: 15940300]
26. Zandi PP, Badner JA, Steele J, Willour VL, Miao K, MacKinnon DF, et al. Genome-wide linkage scan of 98 bipolar pedigrees and analysis of clinical covariates. *Mol Psychiatry.* 2007 Jul; 12(7): 630–639. [PubMed: 17505464]
27. Kelsoe JR, Spence MA, Loetscher E, Foguet M, Sadovnick AD, Remick RA, et al. A genome survey indicates a possible susceptibility locus for bipolar disorder on chromosome 22. *Proc Natl Acad Sci U S A.* 2001 Jan; 98(2):585–590. [PubMed: 11149935]
28. Macgregor S, Visscher PM, Knott SA, Thomson P, Porteous DJ, Millar JK, et al. A genome scan and follow-up study identify a bipolar disorder susceptibility locus on chromosome 1q42. *Mol Psychiatry.* 2004 Dec; 9(12):1083–1090. [PubMed: 15249933]

29. Berrettini WH, Goldin LR, Martinez MM, Maxwell ME, Smith AL, Guroff JL, et al. A bipolar pedigree series for genomic mapping of disease genes: diagnostic and analytic considerations. *Psychiatr Genet*. 1991; 2:125–160.
30. Coon H, Jensen S, Hoff M, Holik J, Plaetke R, Reimherr F, et al. A genome-wide search for genes predisposing to manic-depression, assuming autosomal dominant inheritance. *Am J Hum Genet*. 1993 Jun; 52(6):1234–1249. [PubMed: 8503452]
31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007 Sep; 81(3):559–575. [PubMed: 17701901]
32. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet*. 1998 Jul; 63(1):259–266. [PubMed: 9634505]
33. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*. 2002 Jan; 30(1):97–101. [PubMed: 11731797]
34. McPeck MS, Sun L. Statistical tests for detection of misspecified relationships by use of genome-screen data. *Am J Hum Genet*. 2000 Mar; 66(3):1076–1094. [PubMed: 10712219]
35. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006 Aug; 38(8):904–909. [PubMed: 16862161]
36. Wijmsman EM, Rothstein JH, Thompson EA. Multipoint linkage analysis with many multiallelic or dense diallelic markers: Markov chain-Monte Carlo provides practical approaches for genome scans on general pedigrees. *Am J Hum Genet*. 2006 Nov; 79(5):846–858. [PubMed: 17033961]
37. Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet*. 1997 Nov; 61(5):1179–1188. [PubMed: 9345087]
38. Matisse TC, Chen F, Chen W, Vega FMDL, Hansen M, He C, et al. A second-generation combined linkage physical map of the human genome. *Genome Res*. 2007 Dec; 17(12):1783–1786. [PubMed: 17989245]
39. Bacanu SA. Robust estimation of critical values for genome scans to detect linkage. *Genet Epidemiol*. 2005 Jan; 28(1):24–32. [PubMed: 15372617]
40. Hauser ER, Watanabe RM, Duren WL, Bass MP, Langefeld CD, Boehnke M. Ordered subset analysis in genetic linkage mapping of complex traits. *Genet Epidemiol*. 2004 Jul; 27(1):53–63. [PubMed: 15185403]
41. Browning BL. FLOSS: flexible ordered subset analysis for linkage mapping of complex traits. *Bioinformatics*. 2006 Feb; 22(4):512–513. [PubMed: 16368771]
42. Liang KY, Chiu YF, Beaty TH. A robust identity-by-descent procedure using affected sib pairs: multipoint mapping for complex diseases. *Hum Hered*. 2001; 51(1–2):64–78. [PubMed: 11096273]
43. Fan J, Ionita-Laza I, McQueen MB, Devlin B, Purcell S, Faraone SV, et al. Linkage disequilibrium mapping of the chromosome 6q21-22.31 bipolar I disorder susceptibility locus. *Am J Med Genet B Neuropsychiatr Genet*. 2010 Jan; 153B(1):29–37. [PubMed: 19308960]
44. Zhang D, Cheng L, Qian Y, Alliey-Rodriguez N, Kelsoe JR, Greenwood T, et al. Singleton deletions throughout the genome increase risk of bipolar disorder. *Mol Psychiatry*. 2009 Apr; 14(4):376–380. [PubMed: 19114987]
45. Grozeva D, Kirov G, Ivanov D, Jones IR, Jones L, Green EK, et al. Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. *Arch Gen Psychiatry*. 2010 Apr; 67(4):318–327. [PubMed: 20368508]
46. Mathieu F, Dizier MH, Etain B, Jamain S, Rietschel M, Maier W, et al. European collaborative study of early-onset bipolar disorder: Evidence for genetic heterogeneity on 2q14 according to age at onset. *Am J Med Genet B Neuropsychiatr Genet*. 2010 Dec; 153B(8):1425–1433. [PubMed: 20886542]

Table 1

Suggestive results for each Merlin analysis. Most significant analysis for each chromosomal region is bolded.

ASM	SNP	Position	Chr	Pos	Dom- HLOD	Dom- Alpha	Rec- HLOD	Rec- Alpha	Exp- LOD	Pairs- LOD
1	Rs1448110	106770806	2	118	0.1	0.03	1.2	0.07	2.2	3.1
2	Rs1448110	106770806	2	118	0.0	0.01	1.4	0.06	2.2	3.0
3	Rs1878449	155107054	4	157	0.8	0.04	2.4	0.05	0.8	0.8
1	Rs1046943	109890634	6	119	0.9	0.08	0.2	0.03	3.2	3.4*
3	Rs839556	142913571	6	150	0.0	0.00	3.0	0.07	0.5	0.9
1	Rs722642	80485739	9	78	0.6	0.06	1.0	0.06	3.3	3.4*
2	Rs722642	80485739	9	78	0.2	0.03	0.4	0.03	2.5	2.4
2	Rs1408817	114488135	10	130	1.0	0.04	2.7	0.07	0.4	0.5
2	Rs2029463	1618481	11	4	2.2	0.07	0.2	0.02	0.4	0.1
2	Rs8045185	1441376	16	5	0.0	0.00	2.4	0.06	-0.1	0.0
1	Rs17781597	68734433	17	109	1.3	0.08	0.5	0.04	2.7	2.7
2	Rs1872087	69048320	17	111	1.0	0.06	0.6	0.04	2.3	2.2
3	rs884205	58205837	18	91	0.8	0.04	2.2	0.06	2.0	2.3
2	rs565973	61233578	18	96	1.1	0.06	2.1	0.07	2.0	1.9
2	rs1028555	5285541	20	16	2.7	0.09	1.0	0.05	1.9	1.9
2	rs1381100	58273271	20	101	0.5	0.05	1.0	0.05	2.4	2.7

* Genome-wide significant

Table 2

Suggestive results for each MORGAN analysis. Most significant analysis for each chromosomal region is bolded.

ASM	SNP	Base pairs	Chr	Kosambi cM	Dom- HLOD	Dom- Alpha	Rec- HLOD	Rec- Alpha	Rec- Alpha	Spairs- Lod	Slambda- Lod
2	rs955799	105705117	2	118	0.1	0.02	1.7	0.07	0.07	2.6	0.7
1	rs1046943	109890634	6	119	1.2	0.09	0.3	0.03	0.03	2.9	0.0
3	rs839556	142913571	6	150	0.0	0	2.7	0.07	0.07	0.5	0.2
1	rs713478	80501940	9	78	1.4	0.08	1.1	0.06	0.06	2.8	2.0
2	rs1408817	114488135	10	130	0.9	0.04	2.8	0.07	0.07	0.5	1.3
2	rs782944	59125769	15	60	0.0	0	0.0	0	0	0.2	3.0
1	rs17781597	68734433	17	109	1.2	0.08	0.6	0.05	0.05	2.4	0.6
2	rs1028555	5285541	20	16	2.3	0.08	1.1	0.05	0.05	0.9	0.0

Table 3

Individual families with parametric lod scores of at least 2.0 in the MORGAN parametric analyses

ASM	Genetic_Model	Site	Ped	Chr	Loc	Lod
1	Recessive	NIMH	10102	2	105	2.15
1	Recessive	NIMH	21007	4	183	2.16
1	Dominant	UCSF	1561	4	197	2.01
1	Dominant	UCSD	1607	5	18	2.46
1	Dominant	UCSF	1561	10	43	2.08
1	Dominant	NIMH	11156	15	48	2.01
2	Recessive	JHU	31015	2	100	2.09
2	Recessive	NIMH	10102	2	131	2.4
2	Dominant	UChicago	48	4	56	3.19
2	Recessive	JHU	31015	4	141	2.09
2	Dominant	UCSD	1607	5	18	2.38
2	Dominant	Cardiff	5501	10	128	2.25
2	Recessive	JHU	31015	11	124	2.09
2	Recessive	NIMH	10113	12	18	2.1
2	Dominant	Cardiff	5501	12	127	2.25
2	Dominant	UChicago	1482	12	145	2.7
2	Dominant	NIMH	11156	15	52	2.01
2	Recessive	JHU	31302	16	1	2.33
2	Dominant	UChicago	68	19	0	2.14
2	Recessive	JHU	31064	19	23	2.09
2	Dominant	NIMH	13101	19	58	2.05
2	Dominant	NIMH	11107	22	74	2.42
3	Dominant	NIMH	11150	2	1	2.48
3	Recessive	NIMH	11101	2	12	2.01
3	Recessive	UChicago	9000	2	68	2.14
3	Recessive	JHU	31015	2	101	2.41
3	Recessive	NIMH	10102	2	131	2.35
3	Dominant	UCSF	1615	2	229	2.07
3	Recessive	NIMH	11158	2	252	2.06

ASM	Genetic_Model	Site	Ped	Chr	Loc	Lod
3	Recessive	NIMH	11150	3	0	2.12
3	Recessive	NIMH	11158	3	168	2.06
3	Dominant	NIMH	21007	3	213	2.1
3	Recessive	UCSF	1526	4	43	2.04
3	Recessive	JHU	31015	4	152	2.42
3	Recessive	NIMH	11158	4	160	2.06
3	Dominant	NIMH	11150	4	214	2.98
3	Dominant	UCSD	1607	5	14	2.95
3	Recessive	UCSF	1508	5	85	2.23
3	Dominant	NIMH	10158	5	130	2
3	Recessive	NIMH	11158	6	5	2.06
3	Recessive	UChicago	1442	6	71	2.33
3	Recessive	NIMH	11150	9	69	2.14
3	Recessive	UCSD	5807	10	1	2.33
3	Dominant	Cardiff	5501	10	129	2.25
3	Recessive	UCSF	1561	10	141	2.83
3	Dominant	UChicago	643	11	37	2.12
3	Recessive	JHU	31015	11	126	2.41
3	Dominant	NIMH	11107	12	103	2.69
3	Dominant	Cardiff	5501	12	126	2.25
3	Dominant	UCSF	1528	13	59	2.19
3	Recessive	UChicago	48	15	8	2.14
3	Recessive	UChicago	9000	18	76	2.05
3	Dominant	UCSF	1509	18	102	2.91
3	Recessive	JHU	31064	19	23	2.08
3	Recessive	NIMH	12173	19	35	2.17
3	Recessive	NIMH	11101	20	67	2.04
3	Recessive	NIMH	11158	20	67	2.06
3	Recessive	UChicago	1516	22	35	2.31
3	Dominant	NIMH	11107	22	67	2.57

Table 4

Genome-wide significant FLOSS results

ASM	Chr	Covariate	Loc	New Max ¹	Old Max ²	Delta Max ³	p-value ⁴	No. families in subgroup/No. families in analysis	Subgroup
2	2	Age	118	5.18	3.24	1.94	0.0007	429/785	Younger
3	8	Psychosis	82	3.76	2.02	1.74	0.0015	173/943	Not Psychotic
2	12	Num of ASM2	97	4.36	2.4	1.96	0.0015	27/872	More ASM2 diagnoses

¹New Max – Maximum NPL score in subgroup of families analyzed.

²Old Max – Maximum NPL score in all families with covariate information at Loc.

³Difference between New Max and Old Max.

⁴p-value – Significance of increase in NPL score for subset obtained by permutation, controlling for all possible subsets and size of chromosome.