Bipolar Disorders

WILEY

BIPOLAR DISORDERS
AN INTERNATIONAL JOURNAL OF PSYCHIATRY AND NEUROSCIENCES

ORIGINAL ARTICLE OPEN ACCESS

Genetic Linkage for Bipolar Disorder to the Distal Region of Chromosome 19q: A Large Family Whole Genome Sequencing Study

Albert H. C. Wong[1,2,3,4] | Andrew D. Paterson[2,5,6] | Nicole M. Roslin[7] | Elizabeth Tetzlaff[8] | Andrew B. Sedrak[1,3] | Young Zhou[1,3] | Gaurav Sharma[2] | Paul Grof[1,2] | Anne Duffy[9,10]

[1]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada | [2]Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada | [3]Department of Pharmacology and Toxicology, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada | [4]Institute of Medical Science, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada | [5]Program in Genetics & Genomic Biology, the Hospital for Sick Children Research Institute, Toronto, Ontario, Canada | [6]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada | [7]The Centre for Applied Genomics, the Hospital for Sick Children, Toronto, Ontario, Canada | [8]Telfer School of Management, University of Ottawa, Ottawa, Canada | [9]Department of Psychiatry, Oxford University, Oxford, UK | [10]Department of Psychiatry, Queen's University, Kingston, Ontario, Canada

Correspondence: Albert H. C. Wong (albert.wong@utoronto.ca)

Received: 26 November 2024 | **Revised:** 23 June 2025 | **Accepted:** 22 April 2026

Keywords: bipolar | depression | genetics | linkage analysis | lithium | mood disorder

ABSTRACT

Background: Bipolar disorder (BD) is among the most heritable psychiatric disorders, with a genetic architecture likely consisting of both common genetic variants with small effects and rare variants with strong effects in certain families or populations. Genome-wide association studies (GWAS) with increasingly large sample sizes have identified many susceptibility genes, but common variants in these genes do not have a clear pathophysiological pathway to BD. Genetic linkage studies have the potential to identify rare causal variants in certain families.

Aims: We sought to determine the chromosomal regions linked with BD in a specific family that has many members affected by the lithium-responsive subtype of BD.

Materials and Methods: We performed genome-wide genetic linkage analysis of a family identified through a lithium-responsive BD index patient with many relatives also affected with lithium-responsive BD-related mood disorders: three with BD I, four with BD II, and one with a major depressive episode.

Results: WGS (whole genome sequence) data were obtained for 12 members of the lithium-responsive BD pedigree including the eight affected subjects. Both parametric and nonparametric linkage analyses with the narrow BD phenotype and the broader phenotype including all eight with mood disorders provided evidence of linkage to the same region of chromosome 19. The maximum nonparametric linkage score was 3.89 for the broad phenotype, which exceeds typical thresholds for genome-wide significance.

Discussion: We identified a region of chromosome 19 that has not previously been linked to BD. Nor have significant GWAS variants been found in this region. It is possible that this family has different genetic origins for lithium-responsive BD than other patients studied previously. The family we analyzed is part of a larger cohort of BD patients and their family members, and genetic linkage analysis of additional families could be informative.

Conclusion: These results provide a starting point for investigating genes in this chromosomal region that may be involved in the pathophysiology of the lithium-responsive subtype of BD.

1 | Introduction

Bipolar disorder is a severe, heterogeneous, and highly heritable mental illness, characterized by periods of mania and depression. Bipolar disorder ranks among the leading causes of premature death and disability and has a high suicide rate [1]. Lithium is the most effective prophylactic agent for a subgroup of patients with classical manic depressive features, but only one third of bipolar patients diagnosed in accordance with current DSM criteria respond to lithium [2]. The cause of bipolar disorder is largely genetic, with heritability estimated at > 80% [3]. Three diagnoses, bipolar I, bipolar II and cyclothymia, appear to be phenotypic variants sharing the same genetic vulnerability [4], with mania being the common inherited phenotype [3]. As with most common human diseases, bipolar disorder has a heterogenous genetic origin, likely caused by a combination of common and rare genetic variants with variable effects on disease susceptibility [5]. Genome-wide association studies (GWAS) and linkage analyses of multiplex families are complementary approaches [6], allowing the identification of both rare and common genetic variants for bipolar disorder [7].

There have been many genetic linkage studies of bipolar disorder, and two meta-analyses summarize the overall findings [8, 9]. Based on comparisons between two meta-analytic methods: multiple scan probability (MSP) and genome scan meta-analysis (GSMA), and the results of GWAS, it appears that the genetic architecture of bipolar disorder is a mix of heterogeneous and polygenic loci. The earlier meta-analysis included data from well-known linkage studies with isolated populations such as the Old Order Amish, the central valley of Costa Rica, or the Saguenay River/Lac St. Jean region in Quebec, Canada. No genome-wide regions of significant linkage to bipolar disorder were identified [9]. In the more recent meta-analysis, the most significant MSP results were for the 5p14.3-q23.4 region, although nominally significant results were seen on ten different chromosomes including 5, 6, and 14. For the GSMA analysis, none survived Bonferroni multiple testing correction, but the most nominally significant results were seen on chromosomes 3 and 10 [8]. Whole-exome sequencing studies have not identified consistent associations that survive multiple testing correction [10, 11], but large GWAS [12] have identified several genetic susceptibility variants for bipolar disorder [13, 14].

Our group pioneered the use of long-term response to lithium to identify a more homogenous subtype of BD [15]. Lithium responsive bipolar disorder has distinctive clinical features that run in families [16], and may constitute a different genetic sub-type of the disorder [17]. Our approach has been adopted by other groups [18], and has been used in a GWAS on bipolar disorder to identify the novel risk gene *SESTD1* [14]. A *SESTD1* SNP on chromosome 2q31.2 (rs116323614, $p = 2.74 \times 10^{-8}$) had an odds ratio of 3.14 (95% CI 2.10–4.70) when 1639 lithium responders were compared to 889 controls. *SESTD1* encodes for a protein that regulates phospholipids, which are a potential target of lithium [19]. However, the SNP rs116323614 susceptibility allele has a frequency of ~3% in European populations, so can account for only a small proportion of bipolar disorder cases.

Here we describe the results of a whole genome sequence linkage study in a family with many members with bipolar disorder that share a lithium-responsive profile. This family is part of a large research cohort identified through bipolar patients with stable and well-characterized phenotypes based on decades of repeated clinical assessment and prospective protocol-based treatment [20], making it more likely that they carry a rare variant associated with the subtype of lithium responsive bipolar disorder.

2 | Methods

2.1 | Subject Recruitment, Clinical Assessments and Sample Collection

The pedigree in this genetic study was part of a prospective family study of bipolar parents, their adult relatives and high-risk offspring [21]. All probands and relatives signed a written informed consent form and the ethical conduct of this study was approved by the HSREB at Queen's University. The family was identified through an adult member with Bipolar I Disorder diagnosed with the SADS-L [22] interview by a research psychiatrist and confirmed on blind consensus review using all available clinical information by the research team, including at least one other psychiatrist. Consenting adult family members were similarly assessed with the SADS-L research interview and consensus diagnostic reviews of all available clinical information. Lithium response of the proband and any lithium-treated relative was confirmed using the Alda scale based on years of prospective observation [23].

The family studied was selected because it included many members with lithium-responsive bipolar disorder and relatively few other diagnoses, thorough clinical characterization of the affected members, and available DNA (Figure 1). The clinical phenotype was classified in two ways for linkage analysis, with a narrow phenotype defined as BD I or II (7 affected individuals), and a broader mood disorders phenotype that included BD I or II and major depression (8 affected individuals). All DNA samples were collected with the Oragene DNA OG-500 saliva kit and extracted with the prepIT•L2P kit (DNA Genotek, Ontario, Canada).

2.2 | Whole Genome Sequencing

Whole genome sequencing was performed at the McGill Genome Centre. Libraries were normalized, pooled, and denatured in 0.05 N NaOH and neutralized using HT1 buffer. The pool was loaded at 225 pM on an Illumina NovaSeq S4 lane using the Xp protocol as per the manufacturer's recommendations. Samples were run for 2 × 150 cycles (paired-end mode). A phiX library was used as a control and mixed with libraries at 1%. Base calling was performed with RTA v3 (Illumina, California, USA). The program bcl2fastq2 v2.20 was used to demultiplex samples and generate fastq reads [24]. Reads were aligned to the human genome build GRCh38 using bwa-mem (version 7.15), with a coverage of 30×. Mapped reads were further refined using GATK (v3.8) [25] and Picard program suites (v2.90) [26] to improve reads near indels (GATK indel realignment), improve quality scores (GATK base recalibration), and mark duplicate reads with the same paired start locations (Picard mark duplicates).

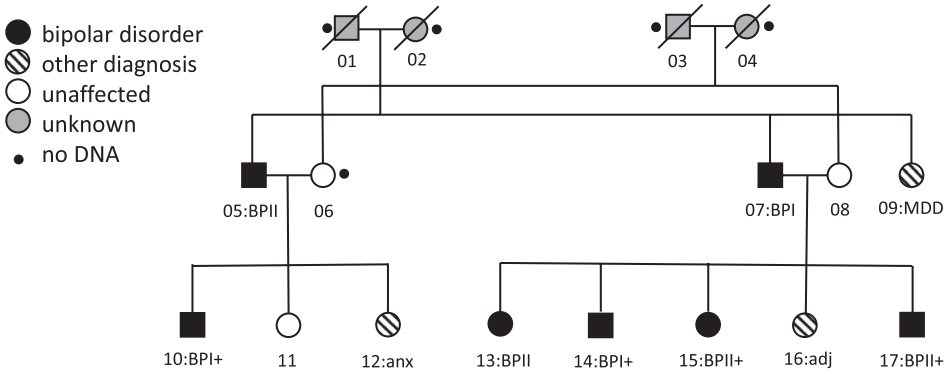


FIGURE 1 | Pedigree of the family analyzed. Family members without DNA are shown to illustrate the relationship between those who had DNA sequencing. Three family members are known to be unaffected, two of which had DNA sequencing. Abbreviations: BP1, bipolar disorder I; BPII, bipolar disorder II; MDD, major depressive disorder; anx, anxiety disorder not otherwise specified; adj, adjustment disorder. + indicates additional diagnosis of substance use disorder (see Table S1).

Calls generated using GATK haplotype caller [25] for SNVs and indels were further processed with functional annotations using snpEff [27] (4.3). To build a gemini database using the GRCh38 reference, we used vcfanno combined with vcf2db. Gemini (v0.11.1a) was used to query variants [28].

2.3 | Variant Filtering for Linkage Analysis

Starting with the multi-sample VCF file, 15% of single nucleotide variants (SNVs) were removed because they did not pass bioinformatic (VQSR) filtering. Since the goal was to obtain a set of high-quality SNPs for linkage analysis, very stringent filtering of genotypes was applied as follows. Autosomal and X chromosome variants were extracted. Genotypes were set to missing if the depth (DP) <15 or >80 on the autosomes, females on chromosome X and males in the pseudoautosomal regions (PARs), or if DP <7 or >40 for males on chromosome X in the non-PAR. Genotypes were also set to missing if the quality (GQ) <30 for autosomal SNVs, chromosome X SNVs for females, and chromosome X PARs for males; the threshold was 15 for chromosome X non-PAR SNVs for males. After setting these genotypes to missing, SNVs were removed if the call rate was $<100\%$. Filtering was performed using bcftools version 1.9.

2.4 | Sample Quality Checking

For each sample, the transition/transversion rate was calculated (Figure S1), as was the proportion of heterozygous autosomal genotypes, and examined for outliers. We also performed a principal components analysis and compared the subjects studied to samples of known ancestry (Figure S2). Genetic sex was inferred using the ratio of the median depths on chromosomes X or Y to the median on chromosome 20 [29], and compared to the sex based on pedigree information. Kinship coefficients between pairs of individuals were estimated using KING version 2.2.4 [30] and checked against the pedigree. Violation of the Mendelian rules of inheritance was checked using PedStats 0.6.10, and any SNV with ≥ 1 observed error was removed. Population inference was performed by performing a principal components analysis (PCA) on individuals 05 and 08 along with publicly available Illumina Omni2.5 genotypes from the 1000 Genomes Project (ref: https://www.biorxiv.org/content/10.1101/078600v1). PCA was performed using Eigensoft 6.0 [31, 32]. Inbreeding coefficients were estimated for all samples using FSuite [33], with 1000 Genomes project samples of Western European ancestry (GBR + CEU) used to calculate allele frequencies.

2.5 | Generation of Variants Set Linkage Analysis

A total of 2,771,963 variants remained after quality cleaning. To reduce computational complexity and avoid spurious results due to linkage disequilibrium (LD), the set of variants was pruned substantially to leave only a subset of SNVs for linkage analysis. This same set of variants were used for all linkage analyses, regardless of phenotype definition. Most subjects were closely related: there were two individuals from the pedigree who are unrelated. Therefore, external data were used to choose the set of SNVs. Genotypes generated on the Illumina Omni2.5 array for the 1000 Genomes samples were used, specifically 95 CEU and GBR samples of Western European ancestry. We started with 436,700 SNVs that were present in both the observed data and the 1000 Genomes data, matching on chromosome, base pair position (hg38) and observed alleles. SNVs were selected if they met the following criteria: minor allele frequency (MAF) >0.2 (87,635 SNVs remained), unique positions on the genetic map, provided by lllumina (81,367 SNVs remained), low pairwise LD, $r^2<0.2$ on a chromosome (26,946 SNVs remained), further pruning based on the genetic map, keeping variants if the inter-marker spacing was at least 0.2 cM (9632 SNVs remained). This set of 9632 SNVs was the final set of variants used in the linkage analyses. The MAF and LD pruning were performed using PLINK version 1.9. b3.42, while the pruning based on the genetic map was performed using FSuite version 1.0.3 [33].

2.6 | Parametric Linkage Analysis

Parametric linkage analysis was performed on the pedigree using narrow and broad phenotype classifications. The maximized maximum LOD score (MMLS) [34] procedure was used, which has been shown to be powerful in situations where the true mode of inheritance is not known [35]. Briefly, analyses were performed using dominant and recessive models with

weak parameters (disease allele frequency and penetrance), and we selected whichever model had the higher LOD score. Then the parameters were optimized to maximize the LOD score. Although a correction for multiple testing should also be performed, it is not meaningful in the current scenario, in which we were searching for regions with the highest evidence for linkage, not reporting significance of LOD scores. For the initial screen, the following parameters were used for the dominant MMLS model: disease allele frequency of 0.01 and penetrances of 0.002, 0.5, and 0.5 for carriers of 0, 1, and 2 copies of a putative disease-causing allele, respectively. The recessive MMLS model used a disease allele frequency of 0.14 and penetrances of 0.002, 0.002, and 0.5. In both cases, the parameters corresponded to a disease prevalence of approximately 1.2%.

Simulations were performed to determine the maximum possible LOD score for this family and for the two MMLS models. Pedigrees were simulated under a range of genetic models using SLINK 3.0.2 [36], preserving the pedigree structure and affection status, and analyzed under the generating model using Merlin 1.1.2. Since SLINK cannot handle the marriage loop (sibship exchange), the program MEGA2 4.7.1 was used to break the loop in the pedigree simulation, and then re-form the loop before analysis [37, 38]. The maximum LOD score obtained from the analysis of 500 simulated pedigrees was declared the maximum LOD score for any particular model. Simulations were performed under dominant and recessive models, each with penetrances of 50%, 60%, 70%, 80%, 90%, 95% and 99%. The phenocopy rate was fixed at 0.02% for all models. Three different disease allele frequencies were used: 0.001, 0.01 and 0.1 for the dominant model, and 0.05, 0.15 and 0.3 for the recessive model, corresponding to approximate disease prevalences of 0.2% and 10%, respectively, for fully penetrant models.

2.7 | Non-Parametric Linkage Analysis

The pedigree was analyzed under two different phenotype definitions using non-parametric linkage analysis. The narrow phenotype included individuals with bipolar I or II disorder. Individuals 06, 08 and 11 were considered to be unaffected and past the usual age at onset. All other individuals were assigned unknown status, resulting in 7 affected, 3 unaffected and 7 unknown subjects. The broad phenotype included all individuals in the narrow phenotype plus individual 09, who had a single major depressive episode, resulting in 8 affected, 3 unaffected, and 6 unknown subjects. This type of analysis does not assume a specific genetic model, and only includes the segregation of alleles in affected individuals in the analysis (i.e., unaffected and unknown individuals are only used to help phase the genotypes, and do not contribute to the score). The Kong and Cox extension for evaluating significance was used, under the exponential model [39]. Both NPLpairs and NPLall statistics were calculated. All analyses were performed using Merlin version 1.1.2.

2.8 | Annotation of Linkage Regions

A TCAG small variant annotation pipeline, rev27.4, was used. This pipeline employed ANNOVAR analysis to functionally annotate small variants in a Variant Call Format (VCF) which represents SNPs and indels within the linkage regions derived from parametric and non-parametric analyses. Excel v16.7 was used to compare the genotypes of affected and non-affected family members, and we focused only on variants in which all affected subjects had the same genotype not present in unaffected subjects. Regions of interest containing these variants were mapped with the NCBI Genome Viewer (GRCh38. p14), from which genes were identified and published work related to brain function, development, or psychiatric disorders was examined.

3 | Results

Quality analysis of the single nucleotide variant (SNV) data revealed that two individuals had lower quality scores and lower depth than the others (Table S1), but the genotypes that were available seemed to provide reliable information, based on kinship estimates and rates of Mendelian errors. No evidence of sample mix-up or contamination was present. The genotypes were consistent with the provided pedigree structure, and two unrelated individuals clustered with reference individuals known to have Western European ancestry. Estimated inbreeding coefficients were not significantly different from zero for all individuals.

3.1 | Parametric Linkage Analysis With the Narrow Phenotype: Bipolar Disorder I and II

Parametric linkage analysis using a narrow phenotype definition showed a maximum observed LOD score of 1.86 on chromosome 9 using the MMLS recessive model, with the simulated maximum score being 1.88. A score of 1.83 was also observed on chromosome 19 (Figure 2). For the MMLS dominant model, the simulated maximum was 1.68, which was observed on chromosome 2, and a score of 1.61 was observed in a different region of chromosome 19. The median information content was 0.99, indicating that the genotypes provided nearly complete information about the segregation of alleles in the pedigree. Since both dominant and recessive models achieved the maximum simulated scores, both models were analyzed under more strict parameters: for the recessive model, the largest simulated LOD score was 2.47, obtained when the disease allele frequency was 0.07% and 99% penetrance, corresponding to a disease prevalence of 0.5%, out of the models considered. For the dominant model, the largest simulated LOD score was 2.09, using a disease allele frequency of 0.0025 and 99% penetrance, also corresponding to a disease prevalence of 0.5%. Analysis of the observed genotypes under these maximized models resulted in larger LOD scores, but similar trends as the MMLS models: LOD scores of 2.44 and 2.41 were observed on chromosomes 9 and 19, respectively, for the recessive model, and scores of 1.68 and 1.61 were observed on chromosomes 9 and 19, respectively, for the dominant model (Table 1).

In parametric linkage analysis using a broader phenotype definition, the maximum LOD score under the MMLS model was 2.20 on chromosome 19 with the recessive model, equal to the maximum simulated score (Figure 3). A score of 2.17 was also seen on chromosome 15. The largest LOD score under the dominant

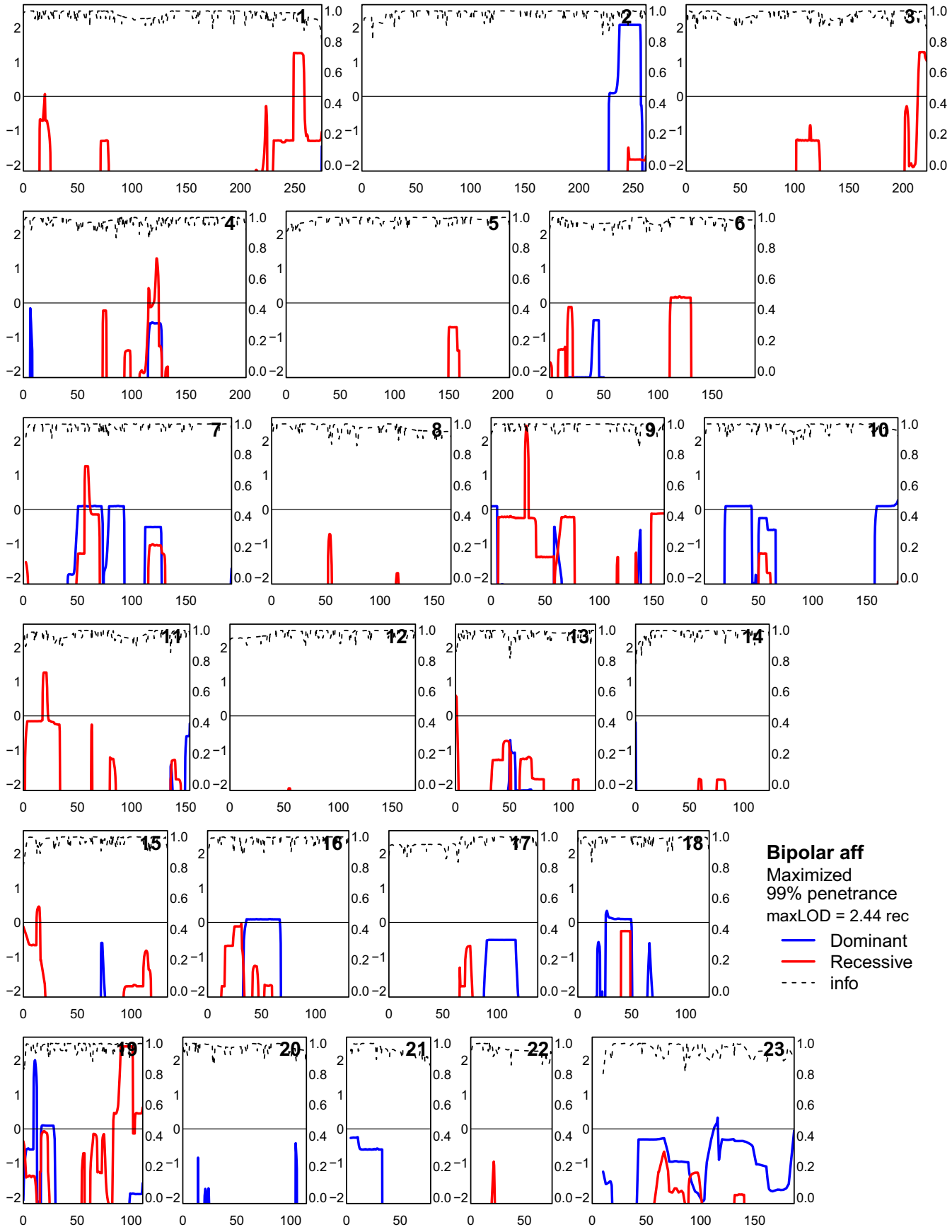


FIGURE 2 | Parametric LOD scores for the narrow bipolar disorder phenotype definition, solid lines, and information content, dashed lines. LOD scores use the axes on the left, and information scores use the axes on the right. LOD scores lower than −2 are not displayed.

TABLE 1 | Summary of genetic linkage analysis including both parametric and non-parametric results of the narrow phenotype (bipolar disorder I and II).

Non-Parametric Analysis							
	Genetic position (cM)			Physical position (bp, hg38)			
Chr	Start	End	Size	Start	End	Size	Max nplall
19	16.19	22.88	6.69	4,783,012	7,000,390	2,217,378	2.68
19	97.69	110.72	13.03	53,738,438	58,269,137	4,530,699	3.28

Parametric Analysis								
	Genetic position (cM)			Physical position (bp, hg38)				
Chr	Start	End	Size	Start	End	Size (Mb)	Model	MMS LOD	Max LOD
2	236.36	256.97	20.6	230,418,996	239,913,783	9.5	D	1.68	2.09
9	31.11	34.73	3.6	14,694,213	16,395,521	1.7	R	1.86	2.44
19	9.26	12.84	3.6	2,949,953	3,755,964	0.8	D	1.61	2.01
19	88.45	101.80	13.4	51,464,613	54,539,353	3.1	R	1.83	2.41

model was 1.81, also equal to the maximum simulated score, on chromosomes 2, 10, 18, and 19. More stringent model parameters did not change the overall trends: under the recessive model (disease allele frequency 0.07, 99% penetrance), the largest LOD scores were 2.79 on chromosome 19 and 2.76 on chromosome 15; under the dominant model (disease allele frequency 0.0025, 99% penetrance), the largest LOD score was 2.08 on chromosome 2, 10, 18, and 19.

3.2 | Non-Parametric Analysis With the Narrow Phenotype: Bipolar Disorder I and II

In the non-parametric linkage analysis of the narrow phenotype, the trends for the NPLpairs and NPLall scoring functions were similar (Figure 4), and so only NPLall will be discussed. A maximum NPL score of 3.28 was observed on chromosome 19, and a smaller peak with a maximum score of 2.68 was also present on chromosome 19 (Table 1). No other chromosome had a score > 2. The maximum score of 3.28 equalled the theoretical maximum for this family as calculated by Merlin. The boundaries of the two regions on chromosome 19 are shown in Table 2. The information content across the genome was nearly universally > 0.9, indicating that the majority of meiotic information was extracted using this subset of SNPs.

3.3 | Non-Parametric Analysis With the Broad Phenotype: Mood Disorders

Linkage analysis was also performed using a broader phenotype definition, in which 8 subjects diagnosed with a mood disorder were classified as affected. Again, the trends for the NPLpairs and NPLall scoring functions were similar and only NPLall will be discussed. Similar to the analysis of the narrow phenotype, the only noteworthy signals were on chromosome 19, with a maximum score of 3.89 and a smaller peak with a maximum score of 2.87 (Figure 5). No other chromosome had a score > 2. The maximum score of 3.89 equaled the theoretical maximum for this phenotype in this family. Compared to the analysis of the narrow phenotype, the peak with the lower score has the same boundaries, and the peak with the larger score covered a slightly smaller region.

3.4 | Consolidated Regions and Genetic Variants

The regions identified in any of the linkage analyses (non-parametric, dominant or recessive models, broad or narrow phenotype) were consolidated. Overlapping regions were merged into the largest continuous region (e.g., regions spanning 16–22 cM and 16–29 cM were merged into a single region from 16–29 cM). When the first or last analyzed variant on a chromosome was included in the linkage region, the first or last known position on the chromosome was substituted, based on the values provided in the UCSC Genome Browser (hg38). Regions identified in the nonparametric analyses were contained within the regions identified by parametric analyses. The consolidated regions covered 100.3 Mb. The region that overlaps among these different analyses and thus has the strongest support for linkage is 19q13.43-19q13.44 (53.7-58 Mb

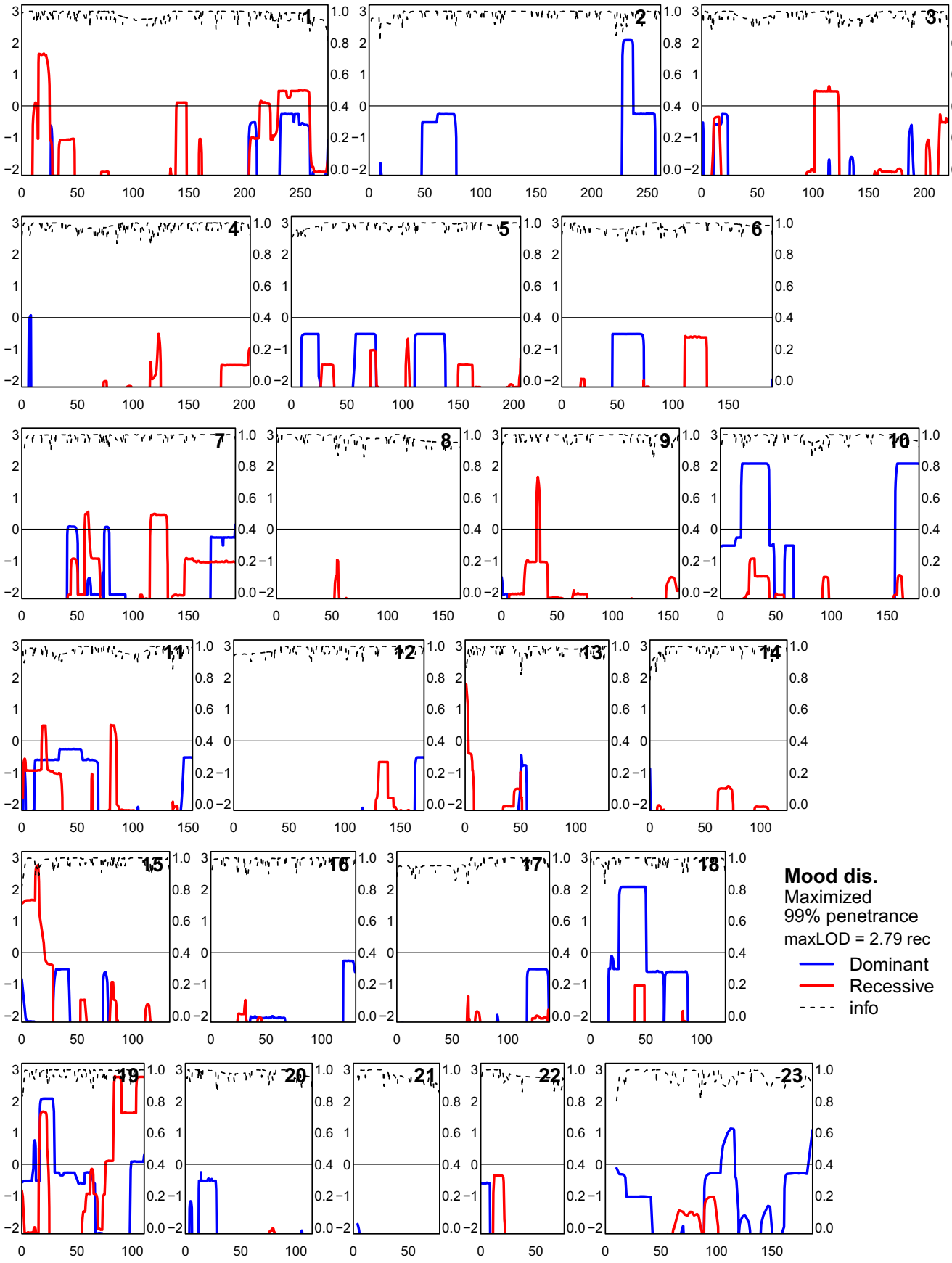


FIGURE 3 | LOD scores for the broader bipolar spectrum phenotype definition, solid lines, and information content, dashed lines. LOD scores use the axes on the left, and information scores use the axes on the right. LOD scores below −2 are not shown.

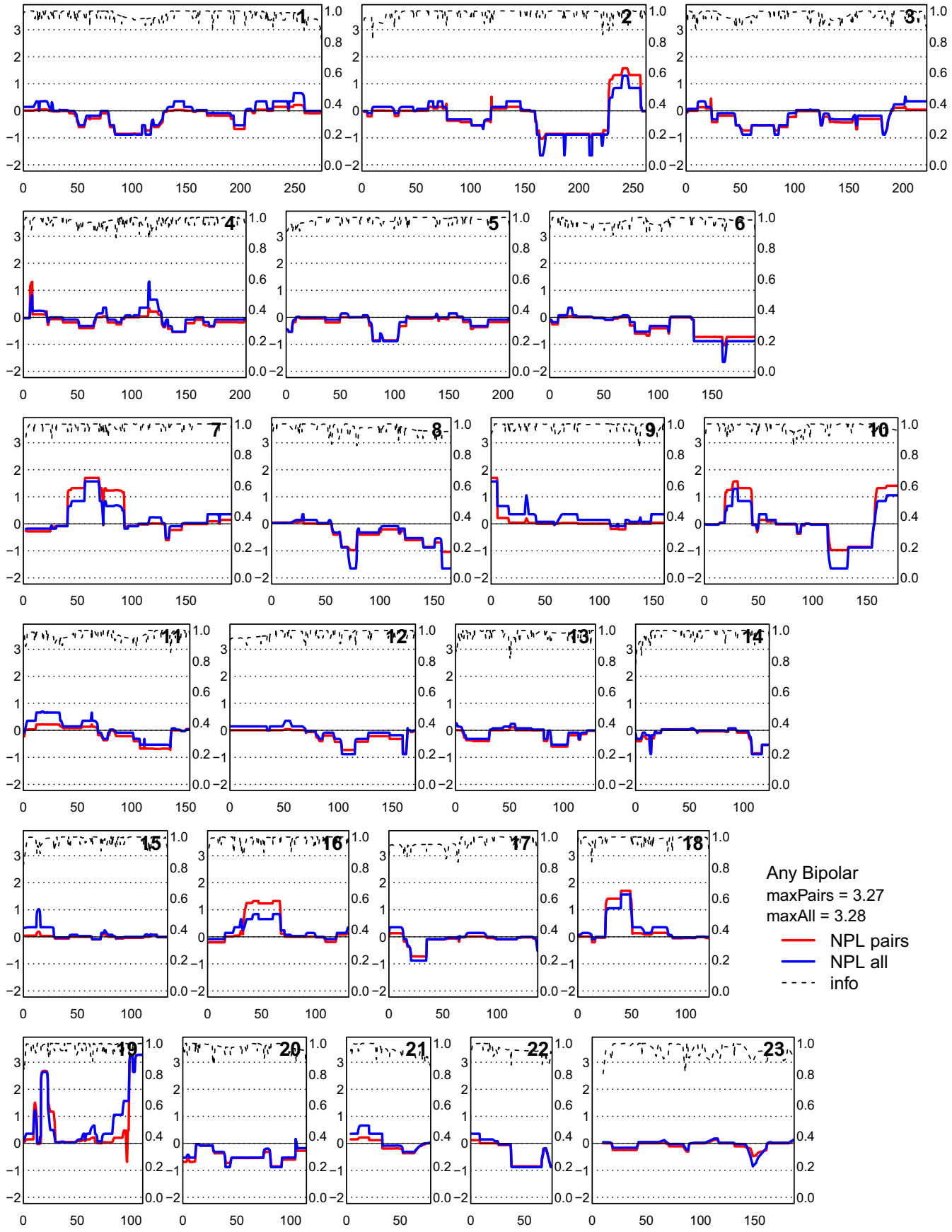


FIGURE 4 | NPL scores for the narrow bipolar phenotype definition, solid lines, and information content (dashed lines). NPL scores use the axes on the left, and information scores use the axes on the right.

TABLE 2 | Summary of linkage regions identified by both parametric and non-parametric analysis, under both narrow (bipolar I and II disorder) and broader phenotype (bipolar and major depressive episode) definitions, under both recessive and dominant modes of inheritance in the parametric analysis.

Chr	Start Bp (hg38)	End Bp	Length Mb
2	223,774,745	239,913,783	16.14
9	14,694,213	16,395,521	1.70
10	6,147,399	19,191,367	13.04
10	126,450,084	133,797,500	7.35
15	0	27,934,457	27.93
18	7,465,951	27,393,258	19.93
19	2,949,953	3,755,964	0.81
19	4,783,012	9,871,658	5.09
19	50,303,066	58,617,700	8.31

region). The non-parametric broad phenotype region (54.8-58 Mb) is within the region that overlaps among these different analyses (53.7-58 Mb region). The overlap region is within the parametric broad phenotype recessive model linkage region (50-58 Mb).

Isolated regions were further refined using the narrow phenotype using simple logic algorithms on Excel (v16.7) within the sample. Extrapolated regions were superimposed on the NCBI genome viewer (Assembly: GRCh38.p14) to highlight the expressed genes, which were then filtered by known function and potential involvement in psychiatric disorders. Figure 6 shows the distribution of structural variants by chromosome within the affected sample. Figure S3 shows the relative location of structural variants present in affected subjects within the linkage regions, and Table S1 enumerates the structural variants within the linkage regions.

4 | Discussion

We sequenced the genomes of twelve family members, seven who have lithium-responsive bipolar disorder, and identified the distal region of the long arm of chromosome 19 as having the most significant linkage. Non-parametric linkage analysis with both narrow and broader phenotypes identified similar regions on chromosome 19, as did parametric linkage analysis under a recessive inheritance model. The LOD scores for both phenotypes are in the range of 2.4–3.89, which is the theoretical maximum for these particular data from this family. The SESTD1 SNP rs116323614 on chromosome 2q31.2 has previously been reported to have a strong association with lithium-responsive BD. All sequenced members of the family in the current study are homozygous for the reference allele (G) at this location, so this variant does not appear to affect susceptibility in this family.

There are a number of previous genetic studies in which variants or regions of chromosome 19 have been associated or linked to bipolar disorder. Two genetic markers on chromosome 19 were associated with bipolar disorder in the large consortium GWAS, but these were not the most significant associations and did not reach genome-wide significance. Rs11085829 ($p = 6.96 \times 10^{-5}$; OR 0.92) is on the short arm of chromosome 19 where the gene *NFIX* is located at position 13,035,312, outside of the 4.7–9.9 Mb region identified with parametric linkage analysis with a dominant model of inheritance of the broad phenotype. The other GWAS SNP on chromosome 19, rs2287921 ($p = 3.08 \times 10^{-6}$; OR 1.10) is at 48,725,015 bp, outside the 19q13.43-19q13.44 (53.7-58 Mb region) linked with bipolar disorder in our study [13]. A CNV-weighted linkage analysis of 46 families identified a different region in chromosome 19q13.2-19q13.31 (chr19:48066441–48,114,839 and chr19:48114839–48,157,656) that exceeds the threshold for genome-wide significance and survived correction for multiple comparisons (empirical $p = 0.033$) [40]. Finally, a population-based linkage analysis of three case-control samples, two with schizophrenia subjects and one with bipolar disorder, reported significant linkage ($p = 0.0000026$) in another chromosome 19q13 region from 46.5 to 48.8 Mb [41]. These linkage regions are near to but do not overlap with the region in which we report linkage (19q13.42–13.43; 53.7-58 Mb region).

A more recent study combined family-based exome sequencing of many bipolar subjects in 8 multiplex families with genetic association testing in a large case–control meta-analysis of the rareBLISS sample of 3541 bipolar disorder cases and 4774 controls [10]. No single gene survived correction for multiple testing, but loci previously implicated as *de novo* variants associated with autism, schizophrenia, and fragile X syndrome (OMIM#300624) were identified. Another study sequenced many members of a group of families with many members diagnosed with bipolar disorder, focusing on identifying rare variants. They hypothesized that rare variants of moderate effect on bipolar disorder susceptibility could cluster together in shared pathophysiological pathways [11]. A number of genetic variants were found to segregate primarily with bipolar disorder in the families studied, although only one variant in one family was present exclusively in the members with bipolar disorder. None of the variants highlighted were found on chromosome 19. The exome constitutes only 1%–2% of the genomes, and therefore exome sequencing could miss causal non-coding variants.

There are examples in which linkage analysis of a single well-characterized family or small homogeneous population has been successful in finding new genetic causes of disease. The *Disrupted-In-Schizophrenia 1* (*DISC1*) gene was identified by linkage analysis of a unique Scottish family carrying a chromosomal translocation that interrupted the gene. Although that translocation is not present in the general population, study of *DISC1* function has significantly increased knowledge of neurodevelopmental mechanisms in schizophrenia and has led to the discovery of novel treatment targets [42]. Another example is the *amyloid-β precursor protein* (*APP*) and *presenilin* genes, discovered by studying early onset familial Alzheimer disease [43]. A rare but strongly protective coding variation in *APP* was discovered in the isolated population of Iceland [44], and though this variant is not applicable to other

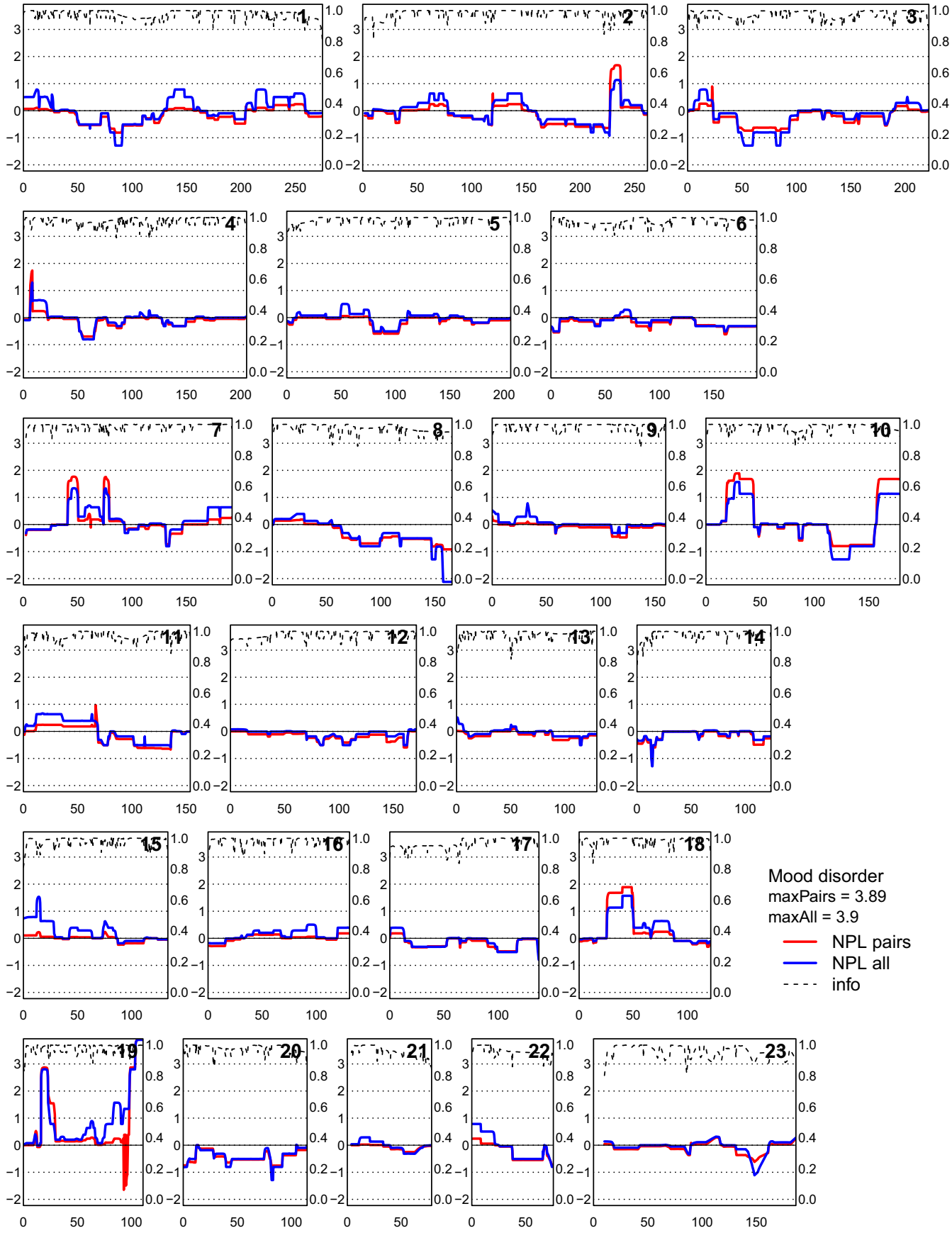
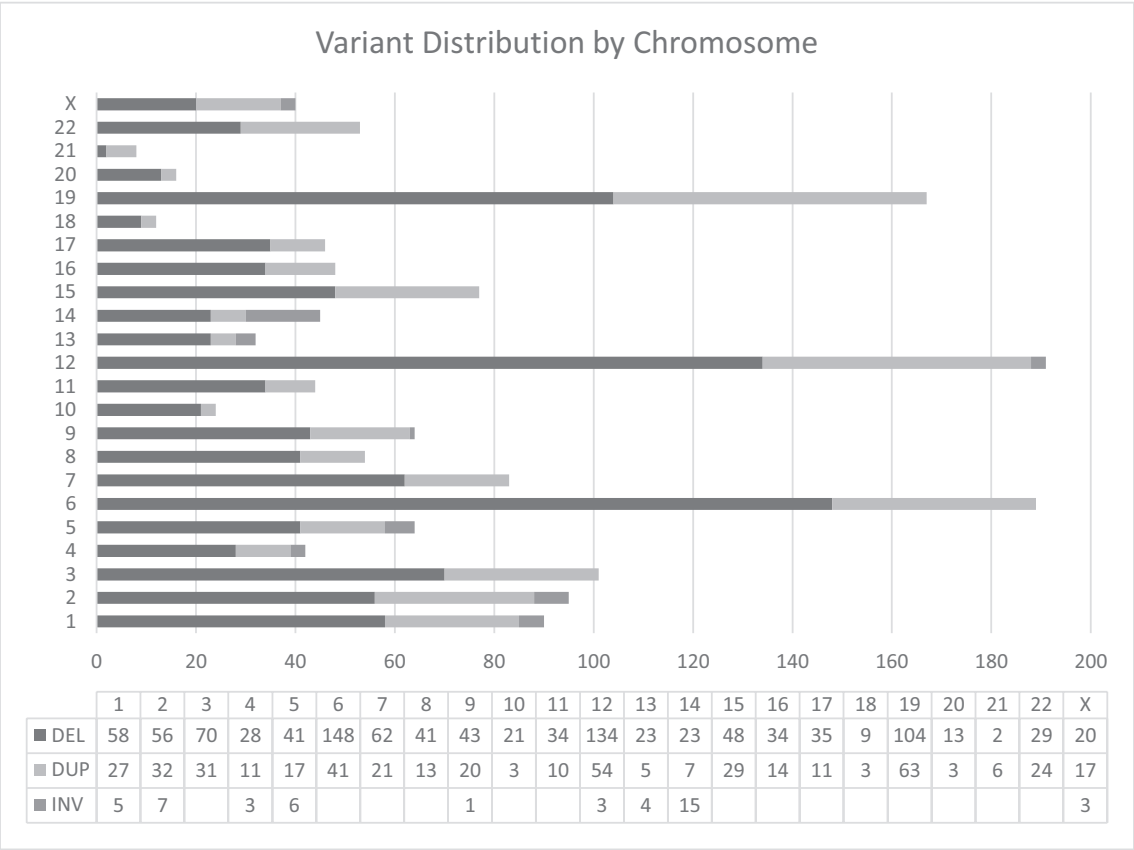


FIGURE 5 | NPL scores for the broader mood disorder phenotype definition, solid lines, and information content (dashed lines). NPL scores use the axes on the left, and information scores use the axes on the right.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
DEL	58	56	70	28	41	148	62	41	43	21	34	134	23	23	48	34	35	9	104	13	2	29	20
DUP	27	32	31	11	17	41	21	13	20	3	10	54	5	7	29	14	11	3	63	3	6	24	17
INV	5	7		3	6				1			3	4	15									3

FIGURE 6 | Distribution of Structural Variants by chromosome within the affected sample. Variant Types DEL: Deletion, DUP: Duplication, INV: Inversion.

populations, it has provided valuable insight into Alzheimer disease pathobiology.

4.1 | Candidate Genes and Relevance to Bipolar Disorder

A number of candidate genes and pathways in the linkage region have potential biological relevance to bipolar disorder. Table S3 shows the position of variants segregating with bipolar disorder within the selected candidate genes discussed below. Disruptions in neurodevelopment have been associated with bipolar disorder, and several genes in our linkage regions are related to neurodevelopment. PTPRS codes for receptor-type tyrosine-protein phosphatase sigma, which is involved in axon growth and guidance, and neuronal migration, and it has previously been associated with bipolar disorder [45]. TUBB4A is important for myelination during neurodevelopment, and there is evidence for white matter and myelination deficits in bipolar disorder [46]. Histone Deacetylase 4 (HDAC4) is a member of a class of enzymes that removes acetyl groups from histones and can modulate gene expression [47]. HDAC4 has been implicated in a number of psychiatric diagnoses including depression, post-traumatic stress disorder (PTSD), and substance use [48]. FRAS1-related extracellular matrix 1 (FREM1) has a splice variant Toll-like/interleukin-1 receptor regulator (TILRR) that could be involved in the neuroinflammation associated with bipolar disorder [49]. Other immune system-related genes in the linkage region include the leukocyte associated immunoglobulin-like receptors (LAIR) encoded by genes within the LILR region on chromosome 19. Leukocyte immunoglobulin-like receptors (LILR) regulate immune response, and their highly conserved genes are expressed in myeloid and lymphoid cells, including monocytes, B and T lymphocytes, and natural killer (NK) cells [50]. LAIR-1 inhibits T-cell signaling through C-terminal Src kinase (CSK) and can suppress inflammation [51]. It would be interesting to examine behaviors relevant to bipolar disorder in LAIR-1 deficient mice [52].

These findings represent the starting point for further work to identify novel molecular genetic causes and disease mechanisms for the lithium-responsive subtype of bipolar disorder. An important question is whether other families with high rates of lithium-responsive bipolar disorder also show genetic linkage to the same chromosomal region as the family we studied. An obvious extension of this work would be to perform the same genetic analysis with other families from the same research cohort.

Acknowledgements

The prospective phenotypic characterization and genetic study of bipolar disorder in lithium-responsive families has been supported with funding from the Canadian Institutes of Health Research, grant number PJT 152976.

Funding

This work was supported by Canadian Institutes of Health Research, PJT 152976.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

References

1. L. L. Judd, P. J. Schettler, D. A. Solomon, et al., "Psychosocial Disability and Work Role Function Compared Across the Long-Term Course of Bipolar I, Bipolar II and Unipolar Major Depressive Disorders," *Journal of Affective Disorders* 108, no. 1–2 (2008): 49–58.

2. P. Grof, A. Duffy, M. Alda, and T. Hajek, "Lithium Response Across Generations," *Acta Psychiatrica Scandinavica* 120, no. 5 (2009): 378–385.

3. P. McGuffin, F. Rijsdijk, M. Andrew, P. Sham, R. Katz, and A. Cardno, "The Heritability of Bipolar Affective Disorder and the Genetic Relationship to Unipolar Depression," *Archives of General Psychiatry* 60, no. 5 (2003): 497–502.

4. J. Edvardsen, S. Torgersen, E. Røysamb, et al., "Heritability of Bipolar Spectrum Disorders. Unity or Heterogeneity?," *Journal of Affective Disorders* 106, no. 3 (2008): 229–240.

5. P. F. Sullivan, M. J. Daly, and M. O'Donovan, "Genetic Architectures of Psychiatric Disorders: The Emerging Picture and Its Implications," *Nature Reviews Genetics* 13, no. 8 (2012): 537–551.

6. J. Ott, J. Wang, and S. M. Leal, "Genetic Linkage Analysis in the Age of Whole-Genome Sequencing," *Nature Reviews Genetics* 16, no. 5 (2015): 275–284.

7. R. L. Kember and M. Bucan, "Promising 2-Pronged Approach to Genetic Basis of Bipolar Disorder," *JAMA Psychiatry* 73, no. 6 (2016): 553–554.

8. B. Tang, T. Thornton-Wells, and K. D. Askland, "Comparative Linkage Meta-Analysis Reveals Regionally-Distinct, Disparate Genetic Architectures: Application to Bipolar Disorder and Schizophrenia," *PLoS One* 6, no. 4 (2011): e19073.

9. R. Segurado, S. D. Detera-Wadleigh, D. F. Levinson, et al., "Genome Scan Meta-Analysis of Schizophrenia and Bipolar Disorder, Part III: Bipolar Disorder," *American Journal of Human Genetics* 73, no. 1 (2003): 49–62.

10. F. S. Goes, M. Pirooznia, J. S. Parla, et al., "Exome Sequencing of Familial Bipolar Disorder," *JAMA Psychiatry* 73, no. 6 (2016): 590–597.

11. C. Cruceanu, A. Ambalavanan, D. Spiegelman, et al., "Family-Based Exome-Sequencing Approach Identifies Rare Susceptibility Variants for Lithium-Responsive Bipolar Disorder," *Genome* 56, no. 10 (2013): 634–640.

12. L. Hou, T. Blank, T. Blank, T. Blank, and T. Blank, "Genetic Variants Associated With Response to Lithium Treatment in Bipolar Disorder: A Genome-Wide Association Study," *Lancet (London, England)* 387, no. 10023 (2016): 1085–1093.

13. Psychiatric GWAS Consortium Bipolar Disorder Working Group, "Large-Scale Genome-Wide Association Analysis of Bipolar Disorder Identifies a New Susceptibility Locus Near ODZ4," *Nature Genetics* 43, no. 10 (2011): 977–983.

14. J. Song, S. E. Bergen, A. di Florio, et al., "Genome-Wide Association Study Identifies SESTD1 as a Novel Risk Gene for Lithium-Responsive Bipolar Disorder," *Molecular Psychiatry* 21, no. 9 (2016): 1290–1297.

15. A. Duffy and P. Grof, "Longitudinal Studies of Bipolar Patients and Their Families: Translating Findings to Advance Individualized Risk Prediction, Treatment and Research," *International Journal of Bipolar Disorders* 12, no. 1 (2024): 12.

16. P. Grof, M. Alda, E. Grof, P. Zvolsky, and M. Walsh, "Lithium Response and Genetics of Affective Disorders," *Journal of Affective Disorders* 32, no. 2 (1994): 85–95.

17. M. Manchia, M. Adli, N. Akula, et al., "Assessment of Response to Lithium Maintenance Treatment in Bipolar Disorder: A Consortium on Lithium Genetics (ConLiGen) Report," *PLoS One* 8, no. 6 (2013): e65636.

18. J. Mertens, Q. W. Wang, Y. Kim, et al., "Differential Responses to Lithium in Hyperexcitable Neurons From Patients With Bipolar Disorder," *Nature* 527, no. 7576 (2015): 95–99.

19. J. C. Soares, A. G. Mallinger, C. S. Dippold, E. Frank, and D. J. Kupfer, "Platelet Membrane Phospholipids in Euthymic Bipolar Disorder Patients: Are They Affected by Lithium Treatment?," *Biological Psychiatry* 45, no. 4 (1999): 453–457.

20. A. Duffy, "Toward a Comprehensive Clinical Staging Model for Bipolar Disorder: Integrating the Evidence," *Canadian Journal of Psychiatry / Revue Canadienne de Psychiatrie* 59, no. 12 (2014): 659–666.

21. A. Duffy, S. Goodday, C. Keown-Stoneman, and P. Grof, "The Emergent Course of Bipolar Disorder: Observations Over Two Decades From the Canadian High-Risk Offspring Cohort," *American Journal of Psychiatry* 176, no. 9 (2019): 720–729.

22. J. Endicott and R. L. Spitzer, "A Diagnostic Interview: The Schedule for Affective Disorders and Schizophrenia," *Archives of General Psychiatry* 35, no. 7 (1978): 837–844.

23. P. Grof, A. Duffy, P. Cavazzoni, et al., "Is Response to Prophylactic Lithium a Familial Trait?," *Journal of Clinical Psychiatry* 63, no. 10 (2002): 942–947.

24. H. Jiang, R. Lei, S. W. Ding, and S. Zhu, "Skewer: A Fast and Accurate Adapter Trimmer for Next-Generation Sequencing Paired-End Reads," *BMC Bioinformatics* 15 (2014): 182.

25. G. A. Van der Auwera, M. O. Carneiro, C. Hartl, et al., "From FastQ Data to High Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline," *Current Protocols in Bioinformatics* 43, no. 1110 (2013): 11.10.1–11.10.33.

26. "Picard Tools-GitHub Release 2208 [Computer Program]," (2019).

27. P. Cingolani, A. Platts, L. Wang le, et al., "A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of *Drosophila melanogaster* Strain w1118; Iso-2; Iso-3," *Fly (Austin)* 6, no. 2 (2012): 80–92.

28. U. Paila, B. A. Chapman, R. Kirchner, and A. R. Quinlan, "GEMINI: Integrative Exploration of Genetic Variation and Genome Annotations," *PLoS Computational Biology* 9, no. 7 (2013): e1003153.

29. S. Chen, L. C. Francioli, J. K. Goodrich, et al., "A Genomic Mutational Constraint Map Using Variation in 76,156 Human Genomes," *Nature* 625, no. 7993 (2024): 92–100.

30. A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, and W. M. Chen, "Robust Relationship Inference in Genome-Wide Association Studies," *Bioinformatics (Oxford, England)* 26, no. 22 (2010): 2867–2873.

31. K. J. Galinsky, P. R. Loh, S. Mallick, N. J. Patterson, and A. L. Price, "Population Structure of UK Biobank and Ancient Eurasians Reveals Adaptation at Genes Influencing Blood Pressure," *American Journal of Human Genetics* 99, no. 5 (2016): 1130–1139.

32. K. J. Galinsky, G. Bhatia, P. R. Loh, et al., "Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia," *American Journal of Human Genetics* 98, no. 3 (2016): 456–472.

33. S. Gazal, M. Sahbatou, M. C. Babron, E. Génin, and A. L. Leutenegger, "FSuite: Exploiting Inbreeding in Dense SNP Chip and Exome Data," *Bioinformatics (Oxford, England)* 30, no. 13 (2014): 1940–1941.

34. S. E. Hodge, P. C. Abreu, and D. A. Greenberg, "Magnitude of Type I Error When Single-Locus Linkage Analysis Is Maximized Over Models: A Simulation Study," *American Journal of Human Genetics* 60, no. 1 (1997): 217–227.

35. D. A. Greenberg, P. Abreu, and S. E. Hodge, "The Power to Detect Linkage in Complex Disease by Means of Simple LOD-Score Analyses," *American Journal of Human Genetics* 63, no. 3 (1998): 870–879.

36. A. A. Schäffer, M. Lemire, J. Ott, G. M. Lathrop, and D. E. Weeks, "Coordinated Conditional Simulation With SLINK and SUP of Many Markers Linked or Associated to a Trait in Large Pedigrees," *Human Heredity* 71, no. 2 (2011): 126–134.

37. R. V. Baron, C. Kollar, N. Mukhopadhyay, and D. E. Weeks, "Mega2: Validated Data-Reformatting for Linkage and Association Analyses," *Source Code for Biology and Medicine* 9, no. 1 (2014): 26.

38. N. Mukhopadhyay, L. Almasy, M. Schroeder, W. P. Mulvihill, and D. E. Weeks, "Mega2: Data-Handling for Facilitating Genetic Linkage and Association Analyses," *Bioinformatics (Oxford, England)* 21, no. 10 (2005): 2556–2557.

39. A. Kong and N. J. Cox, "Allele-Sharing Models: LOD Scores and Accurate Linkage Tests," *American Journal of Human Genetics* 61, no. 5 (1997): 1179–1188.

40. M. Lekman, R. Karlsson, L. Graae, O. Hössjer, and I. Kockum, "A Significant Risk Locus on 19q13 for Bipolar Disorder Identified Using a Combined Genome-Wide Linkage and Copy Number Variation Analysis," *Biodata Mining* 8 (2015): 42.

41. C. Francks, F. Tozzi, A. Farmer, et al., "Population-Based Linkage Analysis of Schizophrenia and Bipolar Case-Control Cohorts Identifies a Potential Susceptibility Locus on 19q13," *Molecular Psychiatry* 15, no. 3 (2010): 319–325.

42. D. J. Porteous, J. K. Millar, N. J. Brandon, and A. Sawa, "DISC1 at 10: Connecting Psychiatric Genetics and Neuroscience," *Trends in Molecular Medicine* 17, no. 12 (2011): 699–706.

43. M. C. Chartier-Harlin, F. Crawford, H. Houlden, et al., "Early-Onset Alzheimer's Disease Caused by Mutations at Codon 717 of the Beta-Amyloid Precursor Protein Gene," *Nature* 353, no. 6347 (1991): 844–846.

44. T. Jonsson, J. K. Atwal, S. Steinberg, et al., "A Mutation in APP Protects Against Alzheimer's Disease and Age-Related Cognitive Decline," *Nature* 488, no. 7409 (2012): 96–99.

45. Y. Wang, Y. Yang, X. Jia, et al., "Identifying Pleiotropic Genes for Major Psychiatric Disorders With GWAS Summary Statistics Using Multivariate Adaptive Association Tests," *Journal of Psychiatric Research* 155 (2022): 471–482.

46. S. Kloiber, J. D. Rosenblat, M. I. Husain, et al., "Neurodevelopmental Pathways in Bipolar Disorder," *Neuroscience and Biobehavioral Reviews* 112 (2020): 213–226.

47. G. Milazzo, D. Mercatelli, G. di Muzio, et al., "Histone Deacetylases (HDACs): Evolution, Specificity, Role in Transcriptional Complexes, and Pharmacological Actionability," *Genes (Basel)* 11, no. 5 (2020): 556.

48. A. Sarkar, P. Chachra, P. Kennedy, et al., "Hippocampal HDAC4 Contributes to Postnatal Fluoxetine-Evoked Depression-Like Behavior," *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology* 39, no. 9 (2014): 2221–2232.

49. M. A. Kashem, X. Y. Yuan, L. Li, J. Kimani, F. Plummer, and M. Luo, "TILRR (Toll-Like Interleukin-1 Receptor Regulator), an Important Modulator of Inflammatory Responsive Genes, Is Circulating in the Blood," *Journal of Inflammation Research* 14 (2021): 4927–4943.

50. L. Storm, J. Bruijnesteijn, N. G. de Groot, and R. E. Bontrop, "The Genomic Organization of the LILR Region Remained Largely Conserved Throughout Primate Evolution: Implications for Health and Disease," *Frontiers in Immunology* 12 (2021): 716289.

51. J. E. Park, D. D. Brand, E. F. Rosloniec, et al., "Leukocyte-Associated Immunoglobulin-Like Receptor 1 Inhibits T-Cell Signaling by Decreasing Protein Phosphorylation in the T-Cell Signaling Pathway," *Journal of Biological Chemistry* 295, no. 8 (2020): 2239–2247.

52. X. Tang, L. Tian, G. Esteso, et al., "Leukocyte-Associated Ig-Like Receptor-1-Deficient Mice Have an Altered Immune Cell Phenotype," *Journal of Immunology* 188, no. 2 (2012): 548–558.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Sample quality analysis for the twelve DNA samples. Histograms show the proportion of heterozygous autosomal genotypes (left) and transition/transversion ratio (right) per sample. **Figure S2:** Scatterplot of the first two principal components (PC) calculated using Eigenstrat. Individuals 05 and 08 from the family (black solid circles) were included with samples of known ancestry from the 1000 Genomes Project (open circles, color coded by ancestry). AFR = African, AMR = American, EAS = East Asian, EUR = European, SAS = South Asian. **Figure S3:** Karyotype plot of genomic variants and linkage regions. Vertical lines indicate the positions of rare variants identified in the affected individuals, while blue horizontal bars surrounding the chromosomes denote linkage regions associated with the broad phenotype in the family. **Table S1:**: DNA sample quality summary. **Table S2:**: Summary of Variants in Affected Sample. **Table S3:**: Overlapping variant positions within genes of interest identified from VCF data, based on GRCh38/hg38 coordinates. USC Genome Viewer (hg38).