A linkage and exome study of multiplex families with bipolar disorder implicates rare coding variants of \textit{ANK3} and additional rare alleles at 10q11-q21

Claudio Toma, PhD; Alex D. Shaw, PhD; Anna Heath BSc; Kerrie D. Pierce, BSc; Philip B. Mitchell, MBBS, MD; Peter R. Schofield, PhD, DSc; Janice M. Fullerton, PhD

\textbf{Background:} Bipolar disorder is a highly heritable psychiatric condition for which specific genetic factors remain largely unknown. In the present study, we used combined whole-exome sequencing and linkage analysis to identify risk loci and dissect the contribution of common and rare variants in families with a high density of illness. \textbf{Methods:} Overall, 117 participants from 15 Australian extended families with bipolar disorder (72 with affective disorder, including 50 with bipolar disorder type I or II, 13 with schizoaffective disorder–manic type and 9 with recurrent unipolar disorder) underwent whole-exome sequencing. We performed genome-wide linkage analysis using MERLIN and conditional linkage analysis using LAMP. We assessed the contribution of potentially functional rare variants using a gene-based segregation test. \textbf{Results:} We identified a significant linkage peak on chromosome 10q11-q21 (maximal single nucleotide polymorphism = rs10761725; exponential logarithm of the odds [LOD exp] = 3.03; empirical \( p = 0.046 \)). The linkage interval spanned 36 protein-coding genes, including a gene associated with bipolar disorder, ankyrin 3 (\textit{ANK3}). Conditional linkage analysis showed that common \textit{ANK3} risk variants previously identified in genome-wide association studies — or variants in linkage disequilibrium with those variants — did not explain the linkage signal (rs10994397 LOD = 0.63; rs9804190 LOD = 0.04). A family-based segregation test with 34 rare variants from 14 genes under the linkage interval suggested rare variant contributions of 3 brain-expressed genes: \textit{NRBF2} (\( p = 0.005 \)), \textit{PCDH15} (\( p = 0.002 \)) and \textit{ANK3} (\( p = 0.014 \)). \textbf{Limitations:} We did not examine non-coding variants, but they may explain the remaining linkage signal. \textbf{Conclusion:} Combining family-based linkage analysis with next-generation sequencing data is effective for identifying putative disease genes and specific risk variants in complex disorders. We identified rare missense variants in \textit{ANK3}, \textit{PCDH15} and \textit{NRBF2} that could confer disease risk, providing valuable targets for functional characterization.

\textbf{Introduction}

Bipolar disorder is a complex psychiatric disorder with a population prevalence of approximately 1%\textsuperscript{1}. It is typically characterized by cyclic manic or hypomanic and depressive affect\textsuperscript{2}. Genetic factors are strongly implicated in the disease etiology, with heritability estimates of approximately 70%\textsuperscript{3}. Genome-wide association studies (GWAS) indicate that 20% to 30% of the genetic liability for bipolar disorder is due to common genetic variants, but a significant proportion of the genetic risk factors remain unknown. Overall, GWAS have been an effective approach for identifying susceptibility genes in bipolar disorder, and common susceptibility variants in \textit{CACNA1C}, \textit{ANK3}, \textit{ODZ4} and other genes have been implicated\textsuperscript{4-9}. Recent sequencing studies also suggest a role for rare variants in disease pathogenesis\textsuperscript{10-12}, thus, the genetic model in bipolar disorder is likely to comprise both common and rare variants, the latter with potential for higher penetrance effects\textsuperscript{12}. Despite recent advances in high-throughput technologies, many individual risk alleles remain unknown, but they are potentially tractable in large multiplex families with multiple affected relatives\textsuperscript{13}. Linkage studies have been systematically applied to map disease genes for Mendelian diseases, and they represent a powerful and successful method when a single gene carries etiologic mutations. However, linkage approaches for complex disorders have not achieved similar success: replication of linkage signals among independent studies has typically failed, and follow-up association studies have been largely unable to reveal susceptibility variants in linkage regions, because of substantial heterogeneity both at the allelic and locus levels and across cohorts.
More recently, whole-exome sequencing (WES) or whole-genome sequencing (WGS) have become essential tools for the identification of rare pathogenic variants in multiplex families with psychiatric diseases. In bipolar disorder, WGS or WES in multiplex or extended families have implicated pathogenic single nucleotide variants and rare truncating mutations. However, because of diverse family structures and the relatively small sample sizes thus far employed, results from these individual studies are not currently converging onto the same pool of genes.

Although large family studies have been mostly abandoned in favor of GWAS over the past decade, they have re-emerged as a powerful alternative approach for gene discovery, particularly for mapping rare variants with relatively large effect sizes. Recently, studies in recessive or dominant Mendelian diseases have adopted a linkage analysis using WES- or WGS-derived genotypes, followed by rare variant analysis and the identification of the etiologic variant. This approach has only recently been applied to some psychiatric disorders, in which mostly single families have been examined for pedigree-specific linkage (although a study in attention-deficit/hyperactivity disorder also employed a linkage approach with WES in 3 families comprising 10 relatives); thus, the strength of this strategy for unravelling missing heritability is yet to be clarified. While a number of studies have performed rare variant analysis across the genome, rare variant analyses in specific linkage regions has been performed in only 3 independent studies of bipolar disorder, each examining a single extended family: (1) a large Old Order Amish pedigree with single nucleotide polymorphism (SNP) genotypes for linkage analysis of 388 family members and WGS of 50 people with bipolar disorder; (2) a family of 6 relatives with bipolar disorder with WES and SNP data; and (3) a 4-generation family with 11 patients with bipolar disorder examined using linkage and WES. These studies identified significant linkage regions that bore potential rare etiologic variants, although they did not identify highly penetrant rare variants that would entirely explain the linkage peaks, suggesting that polygenic inheritance also plays a role in these extended families.

In the present study, we aimed to extend our previously published work, which focused on rare-variant sharing in families with bipolar disorder. We performed a genome-wide linkage analysis across 15 extended families (with an average of 4.2 relatives affected with bipolar disorder type I or schizoaffective disorder–manic type) to identify major susceptibility loci and simultaneously map common and rare pathogenic variants from WES that contribute to susceptibility to bipolar disorder.

**Methods**

**Ethics approval**

All experiments were carried out in accordance with regulations and guidelines approved by the University of New South Wales Human Research Ethics Committee (initial approval HREC04144; extensions HREC10078, HC15503, HC16347). We obtained written informed consent from all participants.

**Participants**

Fifteen families with a high density of affective illness over 2 to 4 generations were selected for WES (Appendix 1 Figure S1, available at jpn.ca/200083-a1). The sequenced sample comprised 117 participants: 44 were diagnosed with bipolar disorder I (31 female), 6 with bipolar disorder II (5 female), 13 with schizoaffective disorder–manic type (7 female) and 9 with recurrent unipolar depression (8 female); 45 were unaffected or their status was unknown (25 female; Appendix 1, Table S1). All diagnoses were ascertained using the Family Interview for Genetic Studies and the Diagnostic Interview for Genetic Studies. Additional clinical and demographic information is as previously described.

**WES-derived genotypes and linkage analysis procedures**

Exome enrichment, template sequencing and variant calling were performed at the Lotterywest State Biomedical Facility (Perth, Australia). AmpliSeq exome enrichment and Ion Proton sequencing were performed as previously described: 92% of bases captured were covered by more than 10 reads, and the mean read depth was 112×. Genotypes were called from aligned WES reads using SAMTools pileup (https://github.com/samtools/samtools). The WES-derived genotypes were filtered to identify independent haplotype-informative markers (HapMap CEU population) using LINKDATAGEN for input into linkage analysis, which employed 4044 WES-derived SNPs from 22 autosomes and the X chromosome across all 15 families. Family relationships were confirmed with identity-by-descent analysis using independent WES-derived genotypes in PLINK. Nonparametric linkage analyses were performed using the “all” statistic in MERLIN, under the Kong and Cox linear and exponential models.

Because of the high frequency of depression in the general population and to reduce genetic heterogeneity, for linkage analysis we differentiated between transmitting and nontransmitting parents with recurrent unipolar depression. People with transmitting recurrent unipolar depression were those who reached the threshold criteria for a diagnosis of recurrent unipolar depression and who had 1 offspring or more with bipolar disorder I or schizoaffective disorder–manic type. Those who reached the threshold criteria for a diagnosis of recurrent unipolar depression but who had no descendants, or whose descendants were not diagnosed with bipolar disorder I or schizoaffective disorder–manic type, were considered “nontransmitting” and were conservatively coded as “unknown” rather than “affected.” Thus, we coded 60 participants as “affected”: those with a diagnosis of severe bipolar disorder (bipolar disorder I or schizoaffective disorder–manic type; n = 57), or parents with recurrent unipolar depression who had affected offspring (n = 3; Appendix 1, Table S1). Participants with all other diagnoses were coded as “unknown,” and those who
did not reach any diagnostic criteria were coded as “unaffected” (Appendix 1, Figure S1).

**Refining intervals and empirical significance of linkage**

To reduce the inter-marker interval to less than 1 cM across the maximal linkage peak at chromosome 10q11-q21, we selected 5 additional SNPs with high heterozygosity in Caucasian Europeans between rs2271128 and rs2273773 (85.4–90.8 cM): rs7073837 (86.217 cM), rs1061259 (87.263 cM), rs2105702 (89.187 cM), rs4548513 (89.5 cM) and rs7896005 (90.717 cM). These markers were not in linkage disequilibrium (LD; \( r^2 < 0.03 \)), based on the genotypes of 405 non-Finnish Europeans (www.internationalgenome.org).

After linkage interval refinement, we computed the relative family contribution to the overall linkage using the -perFamily option in MERLIN.

We calculated the empirical significance of linkage using the -simulate option in MERLIN, in which input genotype data were replaced with simulated chromosomes conditional on family structure and phenotypes, generating 10000 simulated data sets. The maximum logarithm of the odds (LOD) for linkage from each simulation was identified, and the empirical significance was defined by the number of times the observed maximum LOD exceeded the simulated maximum LOD, divided by the total number of simulations. A significance threshold was set at empirical \( p < 0.05 \).

**Conditional linkage analysis and conditional SNP selection**

Because significant associations have been reported between bipolar disorder and a number of common SNP variants in the ankyrin 3 (ANK3) gene,\(^5\)–\(^9\),\(^35\),\(^36\) we performed conditional linkage analysis to assess whether variants commonly associated with bipolar disorder were responsible for the 10q11-q21 linkage peak. We performed an assessment of LD between previously reported GWAS-associated variants using genotype data from Caucasian Europeans (www.internationalgenome.org) to identify independent risk variants (Appendix 1, Figure S2). We selected 2 independent risk variants of ANK3 for conditional linkage analysis: rs10994397 (intron 2) and rs9884190 (intron 36).\(^5\),\(^6\),\(^36\) These non-coding variants were not covered by WES amplicons, so we imputed genotypes from Illumina PsychChip data obtained for all 117 relatives. The imputation procedure followed ENIGMA2 genetics protocols (http://enigma.ini.usc.edu/) and used the Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html). Imputed genotypes had high-quality metrics (rs9804190 Rsq = 0.811; rs10994397 Rsq = 0.999), and resulting allele frequencies were expected for non-Finnish Europeans (rs10994397 minor allele frequency [MAF] = 0.061, non-Finnish Europeans = 0.054; rs9804190 MAF = 0.235, non-Finnish Europeans = 0.225); genotypes were in Hardy–Weinberg equilibrium. Imputed genotypes were 100% concordant with direct TaqMan SNP genotyping (Applied Biosystems; rs10994397 assay ID C___31344967_10) for all 117 participants, confirming imputation accuracy.

We performed conditional linkage using LAMP software\(^37\) to evaluate whether the observed linkage signal was explained by GWAS risk alleles. The LAMP analysis framework is based on a parametric single-locus penetrance model for the trait; likelihood is maximized over a range of penetrance models and family structures, so it is not dependent on a specific model,\(^37\) unlike the Kong and Cox single parameter framework, which is based on identity-by-descent sharing statistics. Further, LAMP quantifies the degree of LD between the candidate SNP and putative disease locus through joint modelling of linkage and association to determine whether the SNP and the disease locus are (1) in linkage equilibrium (so that the SNP plays no causal role in the linkage signal) or (2) in complete LD (so that the SNP or a marker in complete LD with it may account fully for the linkage signal). LAMP provides 3 likelihood ratio tests of the null hypothesis, and LOD scores are generated as follows: (1) a test for linkage; (2) a test for association in presence of linkage (i.e., that a candidate variant is in complete LD with the disease variant); and (3) a test for other linked variants. We performed conditional linkage analysis considering 17 SNPs, which spanned a 1-LOD drop interval after fine-mapping (listed in Appendix 1, Table S2).

**Family-based association analysis for rare variants across 10q11-q21**

We extracted rare variants for all coding genes within the 95% confidence interval of the chromosome 10q11-q21 linkage peak for all individuals across the 15 pedigrees. Only variants with a MAF less than 5% were included, after examination of allele frequencies from the Exome Variant Server (http://evs.gs.washington.edu/EVS/), ExAC (http://exac.broadinstitute.org), dbSNP138 (www.ncbi.nlm.nih.gov/projects/SNP) and the 1000 Genomes Phase I integrated call set (www.internationalgenome.org). All variants were checked for adequate read coverage (> 10×) and consistency with phased haplotypes in each family to exclude potential artifacts. We predicted variant pathogenicity using Variant Effect Predictor annotation tool software (www.ensembl.org/Tools/VEP), which incorporates both PolyPhen and SIFT.\(^38\),\(^39\) However, because the detected exonic variants could be in LD with a causal variant but be nonpathogenic themselves, we examined all rare variants regardless of predicted pathogenicity. This selection identified 34 rare variants, including all missense, indels or stop mutations.

Next, we performed a family-based association test, using the gene-based segregation test (GESE) package implemented in R (https://cran.r-project.org/web/packages/GESE/) and further described in Appendix 1. The parameters for GESE analysis were as follows: segregation of rare variants was assessed in people diagnosed with bipolar disorder I and schizoaffective disorder–manic type (bipolar disorder II and recurrent unipolar depression were considered unknown, as per nonparametric linkage parameters); the population frequency of each of the 34 rare variants was included in an allele-frequency weighting; \( p \) values were calculated after 100000 simulations; and per-family weights were included based on a combination of the ratio of severely
affected patients (bipolar disorder I and schizoaffective disorder–manic type) per family and diagnosis age at onset, to include a family burden of illness variable.

Genes emerging from the association test were assessed for evidence of brain expression, defined from RNA sequencing data from the developmental transcriptomics project (www.brainspan.org/rnaseq/search/index.html), where genes with reads per kilobase per million (RPKM) greater than 1 at any age and in any brain region were defined as brain-expressed (35% of all genes).

Finally, we identified all rare variants in ANK3 from WES regardless of variant type or pathogenicity (intronic, indels, synonymous or missense with MAF < 5%) and performed a segregation study for each variant in the corresponding family. All rare variants in ANK3 reported here were validated by Sanger sequencing.

Results

Linkage analysis

We performed a genome-wide nonparametric linkage analysis using the WES-derived genotype data of 117 people from 15 extended bipolar disorder pedigrees, including 60 severely affected individuals with diagnoses of bipolar disorder I or schizoaffective disorder–manic type, plus their transmitting parents with recurrent unipolar depression (Appendix 1, Figure S1). The highest peak LOD score was on chromosome 10q11-q21, with an exponential LOD of 2.863 ($p = 1.347 \times 10^{-4}$) and a linear LOD of 2.424 ($p = 4.163 \times 10^{-4}$) at rs10761725 (87.25 cM; Figure 1). With the addition of 5 independent WES-derived markers to reduce inter-marker intervals, the evidence for linkage at 10q11-q21 increased (exponential LOD =

Fig. 1: Results of the genome-wide linkage analysis in 15 extended families with bipolar disorder. The results are shown for nonparametric linkage under linear (black line) and exponential (red line) models. Each chromosome is represented in a plot, including the X chromosome. A linkage signal with LOD greater than 2 was observed at chromosome 10q11-q21, with a maximum linear LOD score of 2.424 at marker rs10761725 (uncorrected $p = 0.00042$) and a maximum exponential LOD score of 2.863 (uncorrected $p = 0.00014$). Additional markers were subsequently added to refine the 10q11-q21 locus. Chr = chromosome; LOD = logarithm of the odds (for linkage).
3.103, \( p = 7.84 \times 10^{-5} \); linear LOD = 2.591, \( p = 2.76 \times 10^{-4} \) and rs10761725 remained the peak marker (Figure 2A; Appendix 1, Table S2). To assess the empirical genome-wide significance of the 10q11-q21 linkage, we permuted phenotypes across 10000 simulations, obtaining significance levels of \( p = 0.045 \) and \( p = 0.046 \) for the observed linkage under linear and exponential models, respectively.

The linkage interval, as defined by a 1-LOD drop interval, spanned a genetic distance of 15.88 cM between markers rs12571253 and rs2105702 (73.3 cM to 89.19 cM) and a physical interval of 16.1 Mb (hg19/chr10: 51589149 to 67726514). Per-family linkage analysis indicated that 11 families were contributing positively to the LOD score at 10q11-q21 (Appendix 1, Table S3). This 95% confidence interval contained...
36 protein-coding genes (Figure 2A), including the ANK3 gene, previously implicated in GWAS studies of bipolar disorder.

**Conditional linkage analysis and ANK3 SNP selection**

Significant or suggestive association with a number of SNPs within the ANK3 locus have been reported in GWAS studies of bipolar disorder: rs10994336, rs10994337, rs10994339, rs10994344, rs10994347, rs10994338, rs4948418, rs10994359, rs10994415 and rs9804190. Therefore, we examined whether common variants associated with bipolar disorder were partly or entirely responsible for the observed linkage signal at 10q11-q21 by conditional linkage analysis.

First, we performed an LD study to assess the independence of these 8 risk alleles associated with bipolar disorder. Using 1000 Genomes genotype data from 405 non-Finnish Europeans, we observed high to moderate LD among the low-frequency SNPs associated with bipolar disorder in the 5’ regulatory region of ANK3 (Appendix 1, Figure S2; hg19/chr10: 62076628 to 62322034 bp; high LD among rs10994299, rs10994336, rs10994338, rs4948418, rs10994359 and rs10994397; moderate LD with rs10994415; D’ > 0.65 and R2 > 0.38; average MAF = 0.054). We observed an independent signal with the more common SNP variant rs9804190 (MAF = 0.225) within ANK3 intron 36 (Figure 2B). Thus, conditional linkage with 1 SNP from the 5’ LD block would index the same risk effect, and in conjunction with rs9804190, would capture most of the common variant association signal from ANK3. Therefore, we selected rs10994397 (with the highest imputation quality metric; Rsq = 0.999) as the tagging SNP for the 5’ LD block, which maps within intron 2 of ANK3 (Figure 2B).

Conditional linkage analysis using LAMP revealed that the risk variants rs10994397 and rs9804190 only partially explained the 10q11-q21 linkage signal (rs10994397 LOD = 0.63; rs9804190 LOD = 0.04), providing evidence that those variants contributed minimally to the overall LAMP linkage signal of a LOD greater than 3 (Appendix 1, Table S4). Furthermore, the test for other linked variants suggested that variants other than rs10994397 and rs9804190 (with “other linked variant” LODs of 3.4 and 2.91, respectively) were likely responsible for the observed linkage signal at 10q11-q21.

**Family-based association analysis for variants across 10q11-q21**

Next, we explored the possibility that rare alleles — in ANK3 or in other genes in the region and not indexed by SNPs identified by GWAS (typically with MAF > 5%) — explained the linkage peak at 10q11-q21. We extracted all rare variants (MAF < 5% in Exome Variant Server) from WES data in protein-coding genes under the linkage peak interval (Figure 2A); after quality-control filters, we selected 34 variants from 14 genes for segregation analysis (Appendix 1, Table S5).

We performed a family gene-based segregation test of the 34 rare variants using the GESE package, with no selection for the predicted pathogenicity of variants or the total count of rare alleles among affected or unaffected relatives in the 15 pedigrees. Allele-frequency weighted segregation analysis revealed significant rare variant segregation with bipolar disorder in 4 genes: NRBFB2 (p = 0.005), PCDH15 (p = 0.002), ASAH2 (p = 0.004) and ANK3 (p = 0.014; Table 1). The only truncating variant (nonsense) was found in the ASAH2 gene, and although sphingolipid metabolism is important for neuronal differentiation and signalling, this gene appears to show extremely low or nondetectable expression in the human brain over development, so its potential contribution to bipolar disorder is unlikely. Interestingly, the 4 rare variants that segregated with bipolar disorder in the genes NRBFB2, PCDH15, ANK3 and NRBFB2, with scores of greater than 20 and greater than 0.99, respectively, placing them in the top 1% to 0.01% of deleterious variants in the human genome.

**Further exploration of variation in ANK3**

Given the established role of ANK3 common variants in risk for bipolar disorder, we explored the potential involvement of rare variants in this gene by considering coding and non-coding alleles in all bipolar disorder subphenotypes (bipolar disorder I, bipolar disorder II, schizoaffective disorder–manic type and recurrent unipolar depression) with any segregation pattern among affected or unaffected participants (Table 3). We also included rare intronic or rare synonymous variants, given their potential effect on splicing or exon splicing enhancers, which are typically discarded during the WES or WGS pathogenic variant selection process.

We observed a single missense variant (rs780899852) in 1 unaffected parent with segregation to all 4 affected offspring and none of the 2 unaffected offspring of pedigree 18 (Tables 2; Table 3). This was predicted to be damaging in 3 ANK3 transcripts, including the longest canonical

| Gene | Rare variants/ segregating rare variants, n | Observed probability | p value | Weighted p value |
|------|-------------------------------------------|---------------------|--------|-----------------|
| NRBFB2 | 1/1 | 0.00015 | 0.00051 | 0.00538 |
| PCDH15 | 10/2 | 0.00017 | 0.00774 | 0.00220 |
| ASAH2 | 1/1 | 0.00141 | 0.01275 | 0.00408 |
| ANK3 | 5/1 | 0.00454 | 0.04415 | 0.01403 |
| JMJD1C | 7/2 | 0.02067 | 0.24728 | 0.09346 |

The number of rare variants per gene relative to the number of rare variants segregating with bipolar disorder status across the 15 families with bipolar disorder (n = 117 individuals). The observed segregating probability for each gene is provided, as is the basic gene-based p value for association after simulations and the gene-based p value after per-family weight corrections. Suggestive p values (0.1) are reported.

*Genes not expressed in the brain (defined by reads per kilobase per million [RPKM] < 1 in developmental transcriptomics RNA sequence data, representing 65% of mapped genes).
Table 2: Most significantly associated rare variants in the family-based association test and their predicted pathogenicity*

| Chr:BP (hg19) | dbSNP identifier | Gene | Reference allele/alternative allele | Variant type | CADD Phred score | DANN score | Alternative allele count† | Pedigree (per-family LOD) |
|---------------|------------------|------|------------------------------------|--------------|-----------------|------------|------------------------|--------------------------|
| 10:52005095   | rs116049719       | ASAH2‡ | G/A                                 | Nonsense     | 35              | 0.9974     | 9                      | 131 (0.31)               |
| 10:55566719   | rs570828018       | PCDH15 | G/A                                 | Missense     | 19.18           | 0.9984     | 5                      | 138 (0.43)               |
| 10:55581787   | Novel            | PCDH15 | T/A                                 | Benign       | 23              | 0.9664     | 5                      | 18 (0.42)§               |
| 10:64967951   | rs139722368       | JMJD1C | CTAAAC/−                            | Indel        | 19.05           | —          | 15                     | 129 (0.24)               |
| 10:64974380   | rs41274074       | JMJD1C | G/C                                 | Missense     | 15.64           | 0.8880     | 15                     | 129 (0.24)               |
| 10:61815652   | rs780899852       | ANK3  | G/C                                 | Benign       | 24.3            | 0.9927     | 5                      | 18 (0.42)§               |
| 10:64913602   | rs68223680       | NRB2/F2 | A/G                                | Missense     | 28.3            | 0.9980     | 6                      | 17 (0.57)§               |

Table 3: Inventory of coding and non-coding rare variants identified in ANK3, which were identified in 8 of the 15 families

| Chr:BP (hg19) | dbSNP identifier | Reference allele/alternative allele | Alternative allele count* | Pedigree† | Affected, any diagnosis (alternative/total, n)‡ | Unaffected (alternative/total, n)§ | Protein impact | PolyPhen | CADD Phred score | DANN score |
|---------------|------------------|------------------------------------|-------------------------|-----------|-----------------------------------------------|-----------------------------------|----------------|----------|-----------------|------------|
| 10:61815652   | rs78099852       | G/C                                | 5                       | 1/8       | 2**/5                                         | 1/4                               | Q4277E         | Q1761E†   | 24.3            | 0.9927     |
| 10:61819543   | rs139092048       | C/A                                | 7                       | 19        | 3/5                                           | 0/2                               | D1654Y†        | D1654Y†   | 20.1            | —         |
| 10:61824133   | rs71495624       | G/A                                | 6                       | 121       | 3/4                                           | 3/4                               | Intronic       | —         | 0.294           | —         |
| 10:61827731   | rs748863853       | A/G                                | 1                       | 17        | 1/5                                           | 0/2                               | M4094T         | M1475T†   | 26.3            | 0.9839     |
| 10:61828561   | rs200650397       | G/A                                | 2                       | 131       | 2/6                                           | 0/2                               | S4026          | —         | 9.294           | —         |
| 10:61831651   | rs41274672        | C/G                                | 3                       | 74        | 1/5                                           | 0/3                               | Q2996H         | Benign    | 0.49            | 0.9712     |
| 10:61834861–3 | rs764632652       | CTC/−                              | 5                       | 131       | 3/6                                           | 2/5                               | E1926/−        | —         | 22              | —         |
| 10:61946557   | rs146438080       | G/A                                | 4                       | 131       | 3/6                                           | 1/2                               | S661†          | —         | 13.29           | —         |

CADD = Combined Annotation Dependent Depletion; Chr:BP (hg19) = the chromosome (chr) and base pair (BP) position of each variant, based on Genome Reference Consortium Human Build 37 (GRCh37/hg19); DANN = Deleterious Annotation of genetic variants using Neural Networks; dbSNP ID = SNP identifier; PolyPhen = polymorphism phenotyping, functional prediction of SNP impact on protein product; SNP = single nucleotide polymorphism.

Pathogenicity prediction is presented for missense variants using PolyPhen, the CADD Phred scaled score and DANN scores. A CADD Phred scaled score of 20 means that the variant is among the top 1% of deleterious variants in the human genome, and a score of 30 means that the variant is in the top 0.1%. For DANN, a score of 0.96 identifies 92.1% of pathogenic variations from ClinVar (www.ncbi.nlm.nih.gov/clinvar) and 18.1% of false-positive benign variations; values greater than 0.98 predict protein-disrupting or -altering variants. Positive values for the per-family LOD score indicate a positive contribution to the linkage signal from each family carrying that variant.

*Pathogenicity prediction is presented for missense variants using PolyPhen, the CADD Phred scaled score and DANN scores. The reference sequence for protein is NP_066267 (ANK3, which were identified in 8 of the 15 families out of the total number of family members with any diagnosis).
†Number of alternative allele carriers in family members affected with any mood disorder diagnosis (i.e., bipolar disorder I, bipolar disorder II, schizoaffective disorder manic type or recurrent unipolar depression; see Appendix 1, Table S1) out of the total number of family members with any diagnosis.
‡Number of alternative allele carriers in unaffected family members out of the total number of family members with any diagnosis.
§Number of alternative allele carriers in unaffected family members out of the total number of family members with any diagnosis.
¶This variant was identified in homozygous form in the unaffected spouse of a person with bipolar disorder I; the 2 people were the parents of a person with bipolar disorder I.
**Reference sequence for ANK3 protein: NP_001191332 (ENST00000373827, isoform-3).
Reference allele/alternative allele: Reference allele/alternative allele is indicated.
Affected, any diagnosis (alternative/total, n): Affected, any diagnosis (alternative/total, n) is indicated.
Unaffected (alternative/total, n): Unaffected (alternative/total, n) is indicated.
Protein impact: Protein impact is indicated.
PolyPhen: PolyPhen is indicated.
CADD Phred score: CADD Phred score is indicated.
DANN score: DANN score is indicated.

isoform-1 (ENST00000280772, Q4277E; NP_066267), isoform-3 (ENST00000373827, Q1761E; NP_001191332) and the short most brain-abundant transcript (ENST00000373820, Q359E; H0Y3A4). We found no evidence for full segregation of any of the observed synonymous or intronic variants of ANK3 (rs71495624, rs20065397, rs146438080) with
bipolar disorder in these families. Other missense or indel variants (rs139092048, rs748883853, rs41274672, rs764632652) were observed in a few affected relatives or were predicted to be benign (Table 3).

Discussion

The combination of next-generation sequencing with linkage analysis has exponentially increased the identification of genes implicated in rare Mendelian diseases. In complex disorders, linkage studies can be promising when combined with whole-exome or genome sequence data, where genetic loci can be simultaneously inspected for rare variant segregation with putative higher penetrance effects in large families. This combined strategy has been employed in several large individual families with a range of psychiatric conditions, and also in a joint analysis of 3 families affected by attention-deficit/hyperactivity disorder (including a total of 10 cases with WES). Here, we expanded this approach by performing a linkage study across 15 unrelated Australian-European families comprising 4 or more relatives per family with an affective disorder (including 117 participants with WES), with the aim of identifying genes that carried rare pathogenic variants through family-based segregation tests. Linkage analysis across the 15 extended families revealed a significant linkage peak at chromosome 10q11-q21, encompassing 16.13 Mb and 36 protein-coding genes, including ANK3.

Several large GWAS studies have reported significant associations with bipolar disorder at variants in ANK3, and the majority of these SNPs are in high LD, suggesting that they are largely indexing a similar genetic effect. A conditional linkage analysis based on independent GWAS-significant SNPs rs10994397 and rs9804190 showed that common bipolar disorder risk alleles in ANK3 contributed only marginally to the observed linkage signal at 10q11-q21, although this did not exclude the involvement of other common risk variants not indexed by rs10994397 and rs9804190.

To uncover the possible effects of rare variants with higher penetrance, we conducted a gene-based segregation analysis that revealed significant segregation in individual families for rare alleles in the protein-coding regions of the brain-expressed genes ANK3, PCDH15 and NRBF2. The conditional analysis and family-based segregation results suggest that common variants play a minor role in explaining the linkage to 10q11-q21, which may be more adequately explained by rare genetic variants. However, segregating rare variants in ANK3, PCDH15 and NRBF2 were observed only in some individual families, and not all families who contributed positively to the linkage signal had segregating rare variants in any of the 36 protein-coding genes examined, suggesting that other variants not captured by WES (such as those in regulatory regions or non-protein-coding genes) also contribute to the genetic risk of bipolar disorder and remain to be elucidated.

A compelling body of evidence makes the ANK3 gene the most plausible candidate gene to emerge from our study. ANK3 belongs to the family of postsynaptic density genes and encodes the ankyrin-G protein, which acts as scaffolding adaptor between ion channels and the β-spectrin–actin cytoskeleton. There are 3 main ankyrin-G isoforms: the canonical 190 kDa isoform is expressed in most tissues and acts as a perisynaptic scaffold in dendritic spine heads and necks; the 270 and 480 kDa isoforms are found predominantly in the brain and nervous system, acting as master organizers of ion channels. These latter isoforms are involved in signal propagation at the axon initial segment and formation of the nodes of Ranvier.

Mutations in ANK3 are reported in several psychiatric diseases, suggesting that this gene has a pleiotropic effect. Several studies also suggest a correlation between the severity of ANK3 mutations and the severity of the phenotype: homozygous truncating mutations in an ANK3 isoform were associated with a neurodevelopmental phenotype characterized by intellectual disability, speech delay and behavioural abnormalities; a balanced translocation in ANK3 affecting all isoforms was reported in a patient with autism spectrum disorder, intellectual disability and muscular hypotonia; heterozygous missense mutations were found in autism spectrum disorder with mild developmental delay; and common variant associations have been reported in bipolar disorder. These studies suggest that highly disrupting loss-of-function coding mutations that segregate in bipolar disorder are unlikely, although it is plausible that not only missense variants but also non-coding variants in ANK3 may explain the observed linkage. Thus, common and rare non-coding variants of ANK3 may be implicated in bipolar disorder and act in multiple processes. For instance, the complex alternative splicing of this gene, which includes up to 51 exons and 6 different start exons, is regulated in a temporal and tissue-specific manner. A rare variant leading to a novel alternative-spliced exon in a specific ANK3 isoform has been recently associated with bipolar disorder and schizophrenia, and common intronic variants may also influence ANK3 splicing. Second, genetic variation in ANK3 may also modulate epigenetic mechanisms through abolishing or creating CpG islands. Methylation occurs primarily in the first intron and the 5’ untranslated region (UTR) of ANK3, with a specific temporal profile that has been shown to affect gene expression. Third, post-transcriptional regulation through specific sites at the 3’-UTR may mediate binding with miRNAs. Rare variants that modify consensus regions at the 3’-UTR recognized by miRNA seed regions may affect gene function, as is the case for miR-34a that regulate ANK3 transcripts. The rare variants reported here may provide valuable targets to take forward into functional analyses examining the molecular mechanisms through which ANK3 increases risk to psychiatric disorders.

However, other genes also appear to contribute to genetic risk for bipolar disorder in the 10q11-q21 locus. From our study, PCDH15 and NRBF2 carried rare variants that segregated with bipolar disorder. Protocadherin-related 15 (PCDH15) is part of the large family of cadherin genes involved in neuronal differentiation and synapse formation. Several members of this superfamily have been associated with psychiatric disorders. PCDH15 is responsible for Usher syndrome type 1F, a recessive Mendelian condition characterized by congenital deafness, retinitis pigmentosa and...
vestibular dysfunction; of those with this syndrome, approxi-
ately 20% manifest mental or behavioural disorders, in-
cluding schizophrenia-like psychosis, bipolar disorder and
atypical autism.38 Structural variants across PCDH15 have
also been reported in patients with bipolar disorder,57 and
common variants influence the personality trait extraversion,
which is associated with bipolar disorder. A recent func-
tional study of neurons derived from induced pluripo-
tent stem cells of 2 patients with bipolar disorder carrying
PCDH15 deletions indicated dendrite shortening and de-
creasing synapse numbers.59 However, the contribution of
rare variants in PCDH15 to psychiatric disorder is unclear,
and although evidence of pleiotropy with respect to other
psychiatric risk genes is compelling,60–62 studies on PCDH15
have been too small to draw definitive conclusions.

The final potential candidate gene emerging from our
study was the nuclear receptor binding factor 2 (NRF2B)
gene that positively regulates autophagy.63 Autophagic pro-
cesses are essential for neuronal survival and function, pre-
serving the balance between synthesis, degradation and recy-
cling of cellular components in response to stress. Cellular
autophagy has been implicated in the risk of both schizo-
phrenia and bipolar disorder.64 Lithium, a first-line medicinal
treatment for bipolar disorder, has been shown to induce
autophagy,65 providing neuroprotection against mitochon-
drial dysfunction and the accumulation of reactive oxygen
species.66 Thus, NRF2B warrants further attention as a poten-
tial contributor to risk of bipolar disorder.

Limitations

Although the conditional SNPs chosen to represent ANK3
common risk variants appeared to index the majority of asso-
ciated SNPs reported thus far, we could not exclude the in-
volvement of other common risk variants not indexed by
rs10994397 or rs9804190. As well, it is plausible that other
non-coding variants in ANK3 or other genes in the linkage
region may explain the remaining linkage we observed. Non-
coding variants in UTRs and cis-regulatory elements may
play a significant role in psychiatric disorders.10,67 However,
tergenic, promoter and intronic sequences were not cap-
tured by the WES employed in this study, and UTRs had lim-
ited coverage by the amplicons used for sequence library
preparation.

Conclusion

Linkage to chromosome 10q11.23-q21.3 in 15 extended fam-
ilies was not explained by common variations in ANK3 asso-
ciated with bipolar disorder, but it was consistent with the
contribution of multiple coding rare variants from multiple
genes, including ANK3, PCDH15 and NRF2B. Rare coding
variants at the 10q11.23-q21.3 locus explain part of the gen-
etic risk for bipolar disorder, and further studies are war-
ranted to identify additional functional alleles, with particu-
lar focus on rare non-coding variants that may affect the
splicing, methylation and post-translational regulation of
ANK3 and additional candidate genes at this locus.

Acknowledgments: We are grateful to all participants and their fam-
ilies, as well as clinical collaborators who were originally involved in
collecting and phenotyping these families. We would also thank
Richard J. N. Alcock from the University of Western Australia (Aus-
tralia) for WES library preparation and analytic support. This study
was funded by the Australian National Medical and Health Research
Council (NHMRC) project grants 1063960 and 1066177, and program
grant 1037196. We gratefully acknowledge the Janette Mary O’Neil
Research Fellowship (to JMF) and Mrs. Betty Lynch (dec.) for sup-
porting this work and our team. DNA was extracted by Genetic Re-
source Identification Australia, an Enabling Facility that was supported
by NHMRC Enabling Grant 401184. Samples were sequenced at the
Lot-
tery State Biomedical Genomics Facility, University of Western Aus-
tralia, and genotyped on Illumina PsychChip at the Australian
genome Research Facility. This research was undertaken with the as-
sistance of resources from the National Computational Infrastructure
(NCI), which is supported by the Australian Government.

Affiliations: From Neuroscience Research Australia, Sydney, Aus-
tralia (Toma, Shaw, Heath, Pierce, Schofield); the School of Medical Sci-
ces, University of New South Wales, Sydney, Australia (Toma, Shaw,
Schofield, Fullerton); the Centro de Biología Molecular ‘Severo
Ochoa’, Universidad Autónoma de Madrid/CSIC, Madrid, Spain
(Toma); the School of Psychiatry, University of New South Wales,
Sydney, Australia (Mitchell); and the Black Dog Institute, Prince of
Wales Hospital, Sydney, Australia (Mitchell).

Competing interests: P. Mitchell reports personal fees from Sanofi
(Hangzhou) and Janssen-Cilag, outside the submitted work.

Contributors: C. Toma, A. Shaw, P. Mitchell, P. Schofield and J.
Fullerton designed the study. All authors acquired the data, which
C. Toma, A. Shaw, P. Mitchell, P. Schofield and J. Fullerton analyzed.
C. Toma and J. Fullerton wrote the article, which all authors re-
viewed. All authors approved the final version to be published and
can certify that no other individuals not listed as authors or in the ac-
knowledgments have made substantial contributions to the paper.

Content licence: This is an Open Access article distributed in ac-
cordance with the terms of the Creative Commons Attribution
(CC BY-NC-ND 4.0) licence, which permits use, distribution and
reproduction in any medium, provided that the original publica-
tion is properly cited, the use is noncommercial (i.e., research or
educational use), and no modifications or adaptations are made.
See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Merikangas KR, Jin R, He JP, et al. Prevalence and correlates of
bipolar spectrum disorder in the world mental health survey ini-
tiative. Arch Gen Psychiatry 2011;68:241-51.
2. Mitchell PB, Johnston AK, Frankland A, et al. Bipolar disorder in
a national survey using the World Mental Health Version of the
Composite International Diagnostic Interview: the impact of dif-
fering diagnostic algorithms. Acta Psychiatr Scand 2013;127:891-93.
3. McGuffin P, Rijsdijk F, Andrew M, et al. The heritability of bipolar
affective disorder and the genetic relationship to unipolar depres-
sion. Arch Gen Psychiatry 2003;60:497-502.
4. Stahl EA, Breen G, Forstner AJ, et al. Genome-wide association
study identifies 30 loci associated with bipolar disorder. Nat Genet
2015;47:793-803.
5. Ferreira MA, O’Donovan MC, Meng YA, et al. Collaborative
genome-wide association analysis supports a role for ANK3 and
CACNA1C in bipolar disorder. Nat Genet 2008;40:1056-8.
6. Psychiatric GWAS Consortium Bipolar Disorder Working Group.
Large-scale genome-wide association analysis of bipolar disorder
identifies a new susceptibility locus near ODZ4. Nat Genet
2011;43:977-83.
7. Charney AW, Ruderfer DM, Stahl EA, et al. Evidence for genetic
heterogeneity between clinical subtypes of bipolar disorder. Transl
Psychiatry 2017;7:e993.
8. Cross-Disorder Group of the Psychiatric Genomics Consor-
tium. Identification of risk loci with shared effects on five major

J Psychiatry Neurosci 2021;46(2)
E255
psychiatric disorders: a genome-wide analysis. *Lancet* 2013; 381:1371-9.

9. Mühleisen TW, Leber M, Schulze TG, et al. Genome-wide association study reveals two new risk loci for bipolar disorder. *Nat Commun* 2014;5:3339.

10. Ament SA, Szelinger S, Glusman G, et al. Rare variants in neuronal excitability genes influence risk for bipolar disorder. *Proc Natl Acad Sci U S A* 2015;112:3576-81.

11. Goes FS, Pirooznia M, Parla JS, et al. Exome sequencing of familial bipolar disorder. *JAMA Psychiatry* 2016;73:590-7.

12. Toma C, Shaw AD, Alcock RN, et al. An examination of multiple classes of rare variants in extended families with bipolar disorder. *Transl Psychiatry* 2018;8:65.

13. Glahn DC, Nimmozion VL, Raventos H, et al. Rediscovering the value of families for psychiatric genetics research. *Mol Psychiatry* 2019;24:523-35.

14. Cruceanu C, Ambalavanan A, Spiegelman D, et al. Family-based exome-sequencing approach identifies rare susceptibility variants for lithium-responsive bipolar disorder. *Genome* 2013;56:634-40.

15. Georgi B, Craig D, Kember RL, et al. Genomic view of bipolar disorder revealed by whole genome sequencing in a genetic isolate. *PloS Genet* 2014;10:e1004229.

16. Strauss KA, Marks S, Georgi B, et al. A population-based study of KCNHI7 p.Arg394His and bipolar spectrum disorder. *Hum Mol Genet* 2014;23:6395-406.

17. Rao AR, Yourswh M, Christensen B, et al. Rare deleterious mutations are associated with disease in bipolar disorder families. *Mol Psychiatry* 2017;22:1089-14.

18. Salvoro C, Bortoluzzi S, Coppe A, et al. Rare risk variants identification by identity-by-descent mapping and whole-exome sequencing implicates neuronal development pathways in schizophrenia and bipolar disorder. *Mol Neurobiol* 2018;55:7366-76.

19. Zhang T, Hou L, Chen DT, et al. Exome sequencing of a large family identifies potential candidate genes contributing risk to bipolar disorder. *Gene* 2018;645:119-23.

20. Domanico D, Fragiotta S, Cutini A, et al. The role of cadherin genes in five isolated agenesis of the corpus callosum with incomplete penetration. *Mol Psychiatry* 2014;19:603-12.

21. Anjanappa RM, Nayak S, Moity NS, et al. A linkage and exome study implicates rare variants of KANK4 and CAP2 in bipolar disorder in a multiplex family. *Bipolar Disord* 2020;22:70-8.

22. John J, Sharma A, Kukhal P, et al. Rare variants in tissue inhibitor of metalloproteinase 2 as a risk factor for schizophrenia: evidence from familial and cohort analysis. *Schizophr Bull* 2019;45:256-63.

23. John J, Bhattacharyya U, Yadav N, et al. Multiple rare inherited variants in a four generation schizophrenia family offer leads for complex mode of disease inheritance. *Schizophr Res* 2020;216:288-94.

24. Corominas J, Klein M, Zayats T, et al. Identification of ADH risk genes in extended pedigrees by combining linkage analysis and whole-exome sequencing. *Mol Psychiatry* 2020;25:2047-57.

25. NIMH Genetics Initiative: Family Interviews for Genetic Studies (FIGS). Rockville (MD): National Institute of Mental Health; 1992.

26. Nurnberger JJ Jr, Bhrar MC, Kaufmann CA, et al. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch Gen Psychiatry* 1994;51:849-59, discussion 63-4.

27. Smith KR, Bromhead CJ, Hildebrand MS, et al. Reducing the exome search space for Mendelian diseases using genetic linkage analysis of some genotypes. *Genome Biol* 2011;12:R85.
58. Lo MT, Hinds DA, Tung JY, et al. Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. Nat Genet 2017;49:152-6.
59. Ishii T, Ishikawa M, Fujimori K, et al. In vitro modeling of the bipolar disorder and schizophrenia using patient-derived induced pluripotent stem cells with copy number variations of PCDH15 and RELN. eNeuro 2019;6:ENEURO.0403-18.2019.
60. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell 2019;179:1469-82.e11.
61. Torrico B, Anton-Galindo E, Fernandez-Castillo N, et al. Involvement of the 14-3-3 gene family in autism spectrum disorder and schizophrenia: genetics, transcriptomics and functional analyses. J Clin Med 2020;9:1851.
62. Torrico B, Shaw AD, Mosca R, et al. Truncating variant burden in high-functioning autism and pleiotropic effects of LRP1 across psychiatric phenotypes. J Psychiatry Neurosci 2019;44:350-9.
63. Zhong Y, Morris DH, Jin L, et al. Nrbf2 protein suppresses autophagy by modulating Atg14L protein-containing Beclin 1-Vps34 complex architecture and reducing intracellular phosphatidylinositol-3 phosphate levels. J Biol Chem 2014;289:26021-37.
64. Kim Y, Santos R, Gage FH, et al. Molecular mechanisms of bipolar disorder: progress made and future challenges. Front Cell Neurosci 2017;11:30.
65. Sarkar S, Rubinsztein DC. Inositol and IP3 levels regulate autophagy: biology and therapeutic speculations. Autophagy 2006;2:132-4.
66. Hou L, Heilbronner U, Degenhardt F, et al. Genetic variants associated with response to lithium treatment in bipolar disorder: a genome-wide association study. Lancet 2016;387:1085-93.
67. Brandler WM, Antaki D, Gujral M, et al. Paternally inherited cis-regulatory structural variants are associated with autism. Science 2018;360:327-31.