

# De novo mutations in bipolar disorder implicate neurodevelopment, immunity, and synaptic signaling

Takafumi Shimada (✉ [t.shimada.tz@juntendo.ac.jp](mailto:t.shimada.tz@juntendo.ac.jp))

Juntendo University Graduate School of Medicine <https://orcid.org/0000-0002-8765-1943>

Rachel Smith

<https://orcid.org/0009-0001-5318-219X>

Heejong Sung

Emily Besancon

Fabiana Lopez

Layla Kassem

National Institute of Mental Health

Nirmala Akula

Francis McMahon

National Institute of Mental Health Intramural Research Program; National Institutes of Health

<https://orcid.org/0000-0002-9469-305X>

---

## Article

### Keywords:

**Posted Date:** October 5th, 2023

**DOI:** <https://doi.org/10.21203/rs.3.rs-3381851/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Bipolar disorder (BD) is a debilitating disorder affecting ~ 1% of the world's population. Although many common and some rare alleles are associated with this complex disorder, little is known about the role of de novo variation. For the first time, this study investigates de novo mutations (DNMs) in families ascertained from genetically isolated populations. Exomes of approximately 1200 individuals consisting of 214 trios were quality controlled and analyzed using Genome Analysis Toolkit (GATK). DNMs were called using HAIL, followed by stringent sample and variant filters. Genes carrying deleterious DNMs (dDNMs) in affected participants were annotated for biological functions and associated brain co-expression modules. A total of 42 loss of function or damaging missense DNMs in 42 genes, including *NRXN1*, *SHANK3*, and *SPECC1*, were detected among individuals with BD and related disorders. Additionally, five genes, *XKR6*, *MRC2*, *SUGP2*, *DICER1*, *PLEC* showed recurrent dDNMs, of which *XKR6* and *MRC2* were previously reported. These genes were significantly enriched for functions related to learning, post-synaptic organization, nervous system development, and calcium ion transport. These genes also significantly overlapped with brain co-expression modules associated with neurogenesis and immunity and significantly enriched in genes expressed in excitatory neurons, endothelial cells, and microglia. These findings support a role for DNM in BD and shed light on its neurobiology. If replicated, genes with significant burdens of DNMs are good candidates for functional genomic studies.

## Introduction

Bipolar disorder (BD) is a complex mental disorder characterized by recurrent manic and depressive episodes. The lifetime prevalence of this disorder is approximately 2%<sup>1,2</sup>. Genetic and epidemiological studies provide compelling evidence that BD is a multifactorial disorder and that genetic and environmental factors contribute to its pathogenesis<sup>34</sup>. A meta-analysis of 24 twin studies estimates the broad-sense heritability of BD to be about 67%<sup>4</sup>. Variants across the allele frequency spectrum are implicated in BD: common single nucleotide polymorphisms account for about 20% of the heritability<sup>5,6</sup>. Association studies have also implicated rare CNVs such as 16p11.2 duplication<sup>7</sup>. The Bipolar Exome (BipEx) collaboration recently identified an excess of inherited ultra-rare loss of function (lof) single nucleotide variants (SNVs) in patients with BD among genes under strong evolutionary constraint in major BD subtypes, compared to controls<sup>8</sup>.

With respect to de novo variation, germline<sup>9-12</sup>, postzygotic mosaic<sup>11,13</sup>, and mitochondrial heteroplasmic variants<sup>13</sup> have been proposed as risk candidates contributing to BD, however their sample sizes are still small, and no genes yet meet conventional criteria as de novo "hits" in BD. Most of the existing evidence of de novo mutation (DNM) in BD comes from the study of simplex families. About 15% of first-degree relatives of people with a BD diagnosis develop BD themselves<sup>14,15</sup>, even though they share ~ 50% of inherited genetic risk factors. This suggests a role for (non-inherited) de novo variation in risk or penetrance. Thus, multiplex families may be highly informative for DNM studies. However, the

contribution of DNMs to BD in multiplex families is largely unexplored, with only 1 study reporting DNMs in 18 multiplex families<sup>12</sup>.

To better understand the contribution of rare de novo single-nucleotide variants (SNVs) and insertion/deletion variants in multiplex BD families, we used deeply clinically phenotyped and whole exome sequenced families from Amish and Mennonite (Anabaptist) communities in North and South America who were ascertained through probands with BD or related illnesses as part of the Amish Mennonite Bipolar Genetics Study (AMBiGen; Detera-Wadleigh et al in press). Communities such as these provide access to large families, mostly with both parents available for study, psychiatric diagnoses largely uncomplicated by comorbid substance abuse<sup>16</sup>, and a relatively homogeneous genetic background owing to founder effects and low rates of introgression from non-Anabaptist populations<sup>17</sup>.

The present study comprised all 199 complete trios from the AMBiGen cohort who have been exome sequenced to date. Unaffected siblings were used as controls when available. DNMs were identified through a rigorous quality control (QC) procedure. Genes carrying deleterious DNMs in affected participants were annotated for biological functions and associated brain co-expression modules. The results highlight neurodevelopmental, immunological, and synaptic contributions to bipolar disorder.

## Methods

### Study cohort

All complete trios were drawn from among 1 179 individuals ascertained, assessed, and whole-exome sequenced as part of the Amish Mennonite Bipolar Genetics Study (AMBiGen). AMBiGen consists of families ascertained through probands with bipolar disorder and related conditions (Detera-Wadleigh et al in press). DNA was extracted from whole blood (n = 771), lymphoblastoid cell lines (n = 403), or saliva (n = 4) using Qiagen DNeasy Blood & Tissue or OraGene saliva kits. One sample had an unknown DNA source. Supplementary Table S1 provides details of the cohort.

### Definition of Phenotypes

AMBiGen recruited probands with bipolar disorder and their family members, some of whom also have a psychiatric disorder. All diagnoses were based on the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM 5). The offspring of 199 trios that passed filtering included individuals with a variety of major psychiatric disorders: 107 cases (49 bipolar I [BD-I], 11 bipolar II [BD-II], 6 schizophrenia [Scz], 6 schizoaffective disorder [SczAD], 11 recurrent [MDD-R] and 12 single episode [MDD-S] cases of major depressive disorder, 2 social anxiety disorder [SAD], 10 unspecified psychiatric disorder. The 92 trios with offspring unaffected by any psychiatric disorder were used as controls. Since a variety of psychiatric disorders show familial aggregation with BD<sup>18</sup> and there is strong genetic overlap among psychiatric disorders across the allele frequency spectrum<sup>3,8</sup>, we divided the trios into four groups based on offspring phenotype: 1) Narrow phenotype (severe bipolar subtypes: BD-I and SczAD, n = 55); 2) Broad

phenotype (the narrow group plus BD-II, Scz, and MDD-R, n = 83); 3) All cases (the broad group plus MDD-S, SAD, and unspecified major psychiatric disorder, n = 107); and 4) Controls (none of the listed disorders, n = 92). (Supplementary Table S2)

## Whole-exome sequencing (WES)

WES was performed by the Regeneron Genetics Center (RGC) (RGC, Tarrytown, NY, USA). Library capture and sequencing has been described in detail previously<sup>19</sup>. Briefly, the IDT xGen Exome Research Panel v1.0 (Integrated DNA Technologies, Coralville, IA, USA) capture was used and 75bp paired-end read sequencing was performed on the Illumina NovaSeq 6000 platform (Illumina San Diego, CA, USA). All samples were randomized before library preparation and sequencing.

## Read alignment and variant calling

Reads were aligned to the human genome build 38 (GRCh38) reference genome provided by UCSC using BWA-mem2 version 2.2.1<sup>20</sup>. We used Genome analysis Tool Kit (GATK) version 4.2.4.1<sup>21</sup> for variant calling based on the GATK4 best practices workflow<sup>22</sup>. Single nucleotide variants (SNVs) and insertions/deletions were called jointly across all 1,179 samples using the GATK HaplotypeCaller package to produce a version 4.2 variant callset file (VCF). Variant call accuracy was estimated using the GATK variant quality score recalibration (VQSR) approach<sup>23</sup>.

## Dataset QC

The VCF file, containing 1 179 samples, was loaded into Hail 0.2 (<https://hail.is/>) to perform basic QC steps. Multi-allelic sites were split into bi-allelic sites using HAIL 0.2. A total of 5 438 676 variants in 19 396 genes were included in the VCF file. An overview of the QC and data cleaning process is presented in Supplementary Figure S1.

## Initial variant filtering

Low-complexity regions defined by RepeatMasker (downloaded from the UCSC Table browser: <http://genome.ucsc.edu/cgi-bin/hgTables>) were removed, as were SNVs and Indels that failed VQSR (tranche filter level of 99 for both SNVs and Indels).

## Genotype filtering

Samples with mismatched genotyped gender were excluded from the downstream analysis. Variants with < 10 reads, homozygous reference calls with a genotype quality (GQ) < 25; homozygous variant calls with < 0.9 of the read depth supporting the alternate allele or with a Phred-scaled likelihood (PL) of being homozygous reference of < 25 were excluded. Additionally heterozygous calls with variant call rate < 0.9 ((Reference allele depth + alternative allele depth) divided by total depth < 0.9), with a PL of being homozygous for the reference allele < 25, or with < 0.25 of the read depth supporting the alternate allele (i.e., an allele balance of < 0.25) were excluded. Heterozygous calls in the X or Y non-pseudoautosomal regions in males were excluded.

## Sample QC

We removed samples with estimated contamination levels using FREEMIX  $> 2.0\%$ <sup>24</sup> or chimeric reads  $> 8.5\%$ . We also removed low quality samples with call rate  $< 95\%$ . To check the accuracy of the reported pedigree information, relatedness was calculated between each pair of samples using Hail's King function and sex was imputed for each sample using Hail's impute\_sex function. Combined with the imputed sex, these inferred pedigrees were compared to reported pedigrees and checked for discrepancies. We defined duplicate and 1st-degree relative samples using a KING<sup>25</sup> kinship value of greater than 0.354 and 0.117, respectively. No duplicate samples were identified. As a result of the above QC steps, a total of 23 samples in the dataset were excluded, leaving 1 156 samples in the analysis.

## Final variant filtering

For final variant filtering, variants with call rate  $< 90\%$  or a Hardy-Weinberg equilibrium p-value less than  $1 \times 10^{-6}$  were excluded, leaving 1 156 samples and 1 082 271 unique variants. This dataset was then used as the starting point for the de novo workflow.

## Variant annotation

We used the Variant Effect Predictor (VEP<sup>26</sup>) version 104 to annotate variants against GRCh38. VEP assigns properties such as gene name, consequence, and pathogenicity inference by Combined Annotation Dependent Depletion (CADD) version 1.6 to each variant<sup>27</sup>. In addition, we annotated with allele frequencies in the Genome Aggregation Database (gnomAD) r2.1.1 in non-neuro samples<sup>28</sup> (<https://gnomad.broadinstitute.org/>), and allele frequencies in Anabaptist populations (from the Anabaptist Variant Server (AVS), <https://edn.som.umaryland.edu/Anabaptist/query.htm>) after lifting the genome coordinates over to GRCh38. Finally, we annotated constraint matrix with probability of being loss-of-function intolerant (pLI) scores, loss-of-function observed/expected upper bound fraction (LOEUF) scores, using the gnomAD loss-of-function metrics table from release 2.1.1<sup>28</sup>.

We processed the VEP annotated consequences, and we defined variant-specific consequences and gene annotations as the most severe consequence of a canonical transcript in which that variant lies. We then assigned variants to four distinct consequence classes: lof, missense, synonymous and noncoding. We subdivided missense variants into 'missense damaging' (misD) if the CADD Phred-scaled scores is greater than 15. The threshold for the CADD Phred-scaled scores was preset according to previous studies<sup>10,11</sup>. Lof or misD DNMs were referred to as deleterious or dDNMs. We defined evolutionarily-constrained genes as those with LOEUF scores  $< 0.35$ , as recommended by the gnomAD team.

## Detection of DNMs

De novo variants were called using the de novo function of Hail 0.2, developed by Samocha et al.<sup>29</sup> ([https://github.com/ksamocha/de\\_novo\\_scripts](https://github.com/ksamocha/de_novo_scripts)). Population allele frequencies for variants were obtained from the non-neuro subset of gnomAD<sup>28</sup> and these frequencies were used as the input priors.

As additional parameters, parents' homozygous reference genotypes were required to have no more than 3% of reads supporting the alternate allele, offspring's heterozygous calls were required to have at least 30% of reads supporting the alternate allele, and the ratio of offspring read depth to parental read depth was required to be at least 0.3.

## DNM filtering and Sample QC

This process identified 4 415 putative de novo variants at 3 729 distinct genomic locations in the 208 offspring in this dataset. For QC on the de novo variants, we retained variants if they were high confidence as indicated by the calling algorithm (Hail 0.2), medium confidence and a singleton in the dataset (N = 1 156). To remove variants stemming from cell line artifacts, an allele balance of at least 0.4 was required for the 104 offspring whose data were generated from lymphoblastoid cell line DNA. Since true de novo variants should be rare, variants were removed if they had an allele frequency > 0.1% in the non-neuro subset of gnomAD, or > 1% in Anabaptist populations (based on the AVS). Variants were excluded if they appeared more than twice in the remaining list of putative DNMs and were then limited to one variant per person per gene, retaining variants with the most severe consequences. For sample QC, samples whose DNA source was whole-blood or saliva were excluded if they had more than seven protein-coding putative de novo variants. Samples whose DNA source was from cell lines were dropped if they had more than five protein-coding putative de novo variants. We subsequently performed manual inspection with IGV-2.11.8<sup>30</sup> and excluded remaining DNMs with either of the following criteria as in the previous study<sup>11</sup>: (1) supported by less than two reads in IGV visualization, (2) coinciding with other two or more variant positions in the same read (likely due to misalignment), and (3) with two or more reads supporting the variant in the parent(s) (suggestive of transmission or systematic errors). For the final sample QC, a total of five samples with more than seven DNMs were excluded from the remaining samples.

On average, the 199 offspring in the final dataset had 0.97 [range: 0.91–0.99] of the exome target meeting 15× sequencing depth, 0.0027 [range: 0.0019–0.0179] of free-mix contamination, and 0.02 [range: 0.003–0.083] of chimeric read percentage. Sample information for the 199 trios passing QC is available in Supplementary Table S3.

## DNM validation

We validated a subset of the DNMs with Sanger sequencing prioritizing dDNMs that were lof, misD in the narrow phenotype (BD-I and SczAD). Sequencing primers were designed using NCBI Primer-Blast (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) and synthesized by Integrated DNA Technologies (IDT Inc, Coralville, Iowa, USA). Forward and reverse primer sequences are shown in Supplementary Table S6. Sanger sequencing was performed by Psomagen Inc. (Gaitthersburg, Maryland, USA).

## Incorporation of published DNMs in controls

To increase statistical power of our case-control comparisons, we incorporated DNMs from published control trios comprising the unaffected siblings of ASD probands in SPARK<sup>31</sup>. We chose SPARK trios

because their DNMs were detected in a similar method to ours: They were exome-sequenced at RGC, aligned to GRCh38, genotype-called using the GATK4 pipeline, and DNM-called using Hail0.2. Variant and sample filtering methods are also similar to what we used. However, due to data availability, we only used autosomal DNMs for our case-control comparison. After re-annotation with our procedures (described above), 3 583 exonic DNMs from 3 032 control trios were combined with the autosomal exonic DNMs from our 92 control trios for the downstream analysis.

### **Statistical analysis of the patterns of dDNM enrichment and genes recurrently hit by dDNMs by phenotype**

All statistical tests were performed using R (<http://www.r-project.org/>). To test for overall rates and recurrence of DNMs, we fit the data to Poisson-distributed models. For comparisons against control DNMs, we used a one-tailed two-sample exact Poisson test. For comparisons against mutation model expectations, we used *denovolyzeR*<sup>32</sup> that estimates the number of expected DNMs by incorporating the triplet context, indel rates, the gene length, and null expectation based on macaque–human gene comparisons. We used a previously developed mutability model<sup>29</sup> to compute a mutability table containing the expected number of DNMs per gene per variant class. The mutation rate for damaging missense DNMs (CADD score  $\geq 15$ ) in a gene was used from the mutability table generated by Dong et al<sup>33,34</sup>. We did not include inframe indels in this analysis because mutation rates for inframe indels are not evaluated within the statistical framework. The observed versus expected number of DNMs for each variant class were compared using a one-tailed Poisson test. The mutation rates were adjusted with the overall rate of synonymous DNMs in SPARK control trios, assuming that the rates of synonymous DNMs are not greatly different across case and control groups from different ancestries, based on the results in the lossifov et al.<sup>35</sup> and Howrigan et al.<sup>36</sup> studies.

Statistical significance for the observed numbers of dDNMs in a gene was assessed by using the *denovolyzeByGene* function in *DenovolyzeR*<sup>32</sup>. As above, we used the mutation rate in a gene from the mutability table generated by Samocha et al.<sup>29</sup> and Dong et al.<sup>33,34</sup>, and excluded inframe indels from the analysis. The exome-wide significance threshold was defined as  $P = 2.74 \times 10^{-6}$  based on the number of genes with available mutation rates in Samocha et al. ( $n = 18\,271$ )<sup>29</sup>.

All the analyses except for overall DNM rate comparison against the mutation model expectation were restricted to autosomal DNMs since Chr X DNMs are hard to interpret and none were validated.

## **Incorporation of published DNMs in Bipolar disorder**

We incorporated published DNMs in BD to further refine their effect on BD risk. We collected results from independent exome-sequenced BD trios from three previous studies<sup>9–11</sup>. Specific publications and descriptive data are listed Supplementary Table S4. In total, we assessed 354 published BD trios. DNMs extracted from Supplementary Data 1 of Nishioaka et al<sup>11</sup> which included Fromer et al<sup>9</sup> and Goes et al<sup>10</sup>. These were re-annotated using the same procedures as the present study and combined with our list of

DNMs. The combined DNM list in cases includes DNMs from the 355 narrow, 437 broad (including BD-NOS), and 461 total case phenotype groups.

## Functional enrichment analysis

To explore functional enrichment of genes carrying dDNMs we employed g:Profiler (<https://biit.cs.ut.ee/gprofiler/gost>)<sup>37</sup>, which provides up-to-date information from numerous databases, including gene ontology (GO).

## Co-expression analysis of DNMS

In the co-expression analysis, we considered variants based on their phenotype and function. We used six independent gene lists extracted from the DNM list combined with previous studies: 1) genes with lof variants from narrow group; 2) genes with misD variants from narrow group; 3) genes with lof variants from broad group; 4) genes with misD variants from broad group; 5) lof variants from control group; 6) genes with misD variants from control group.

### Reference module overlap

To characterize which functional pathways are associated with the DNMs from six gene lists, we used published co-expression modules. Weighted gene co-expression network analysis (WGCNA) identifies clusters, or 'modules', of genes that are highly correlated due to similar expression patterns<sup>38</sup>. The reference modules used were identified in healthy controls across multiple brain regions in Gandal et al.<sup>39</sup> and Hartl et al.<sup>40</sup>. A hypergeometric test was run on the gene overlap between each of DNM gene lists with each reference module from Gandal et al.<sup>39</sup> & Hartl et al.<sup>40</sup>. The background genes were set as the 19 396 unique protein-coding genes compiled from the list of the IDT xGen Exome Research Panel v1.0 (Integrated DNA Technologies, Coralville, IA, USA) capture.

To determine if the gene list overlap is greater than what would be expected by random chance, a permutation test was also performed on each overlap greater in size than 1 gene and not with the grey modules. To do this, a set of genes equal in size to the list was selected from the AMBIGen background. The gene set selected was random but corrected for gene length, i.e., selected from a similar distribution of gene lengths as the original AMBIGen list (with small defined as < 13 Mbp, medium 13–46 Mbp, and large > 46 Mbp; these groupings were selected by creating 3 groups of approximately equal size among the union of the AMBIGen gene lists). The overlap of the permuted gene list with each reference module was calculated; this process was repeated 1000 times for each overlap. Modules that significantly overlapped with the control lists were removed from consideration, so that the reported modules are exclusive to phenotype-associated DNM lists. All p-values were adjusted using the Benjamini-Hochberg (BH) method.

## Characterization of reference modules

### Functional enrichment



Functional enrichment for each of the reference modules was performed using the R package topGO<sup>41</sup>, which accounts for the hierarchical structure of the GO database by penalizing 'parent' pathways with enriched 'children'. Pathways are scored by the number of genes in that pathway also found in DNM list ('significant') vs the number of total genes in that pathway from the background ('annotated'). Using the expected number of significant genes given the size of the list and background, Fischer's exact test is performed. All p-values were corrected for multiple comparisons using the BH method. We used the same background genes (n = 19 396) as in the co-expression analysis.

To further synthesize and extract meaning from these GO results, topic modeling was employed on all significant GO terms across models. Briefly, topic modeling is a text mining method for unsupervised classification of text. The algorithm identifies natural groups of co-occurring words in significant GO terms, or "topics". The algorithm can then quantify the mixture of words associated within each topic, while also determining the mixture of topics that describes each grouping (in this case – reference module). Here, Frequently in and Exclusively (FREX) words are used to characterize each topic. The strength of association of each word with each topic is described by the parameter beta. The strength of association of each gene list module with each topic is described by the parameter gamma, which is the estimated proportion of words from the GO terms in that module that are generated from the respective topic. Thus, each module is summarized by general biological function across all pathway results. Number of topics (K = 4) was chosen based on optimal exclusivity and semantic coherence, biological knowledge, strength of association between gene lists and topics.

### **Cell type enrichment**

Cell-type enrichment for each module significantly overlapped with at least one DNM gene list was calculated as the hypergeometric overlap between the module genes list and the list of genes associated with each cell type<sup>42</sup>. All p-values were corrected for multiple comparisons using the BH method.

### **Developmental trajectories**

In order to explore the expression of DNMs during development, we performed a developmental trajectory analysis on each reference module that significantly overlapped with at least one DNM gene list. Neocortical gene expression values across different windows of life from 421 samples from 41 human brains were accessed from Li et al.<sup>43</sup>. For each module, genes were averaged across samples at each time window and plotted as a smooth curve to visualize periods of average higher and lower expression.

## **Results**

### **DNMs in AMBiGen**

WES data passed all QC procedures in 199 trios, comprising 107 affected offspring (including 55 narrow or 83 broad cases), 92 unaffected offspring, 110 fathers, and 110 mothers. After variant-level QC filtering,

the DNM rate was 1.69 per affected and 1.66 per unaffected offspring, following the expected Poisson distribution (Supplementally Figure S2), similar to the DNM rate in a recent study <sup>44</sup>.

A total of 334 rare DNMs were observed (Supplementary Table S5), including 170 DNMs in protein-coding exons among 59/107 (55%) of affected and 52/92 (57%) of unaffected offspring. In trios where the offspring was affected with a “broad” phenotype, we observed 42 rare dDNMs in 42 distinct genes; no gene carried more than one DNM (Table 1). All 10 DNMs selected for validation were successfully validated by Sanger sequencing (Supplementary Table S6).

## Patterns of DNM enrichment

As expected, there were no significant differences overall in the rates of dDNMs between affected and unaffected offspring (Table 2). The lof DNM rate among offspring with a narrow diagnosis was slightly above the DNM model expectation (rate ratio = 1.47, uncorrected P = 0.18) and nominally increased over controls (rate ratio = 2.27, uncorrected P = 0.03). The lof DNM rate in lof-intolerant genes was similar between affected and unaffected offspring, as were the rates of synonymous DNMs. There was a trend toward enrichment of lof in offspring with BD-I (Supplementary Table S7, rate ratio = 2.23, uncorrected P = 0.04), while misD DNMs were more slightly prevalent in offspring with MDD-R (ratio = 2.17, uncorrected P = 0.03).

## Functional enrichment analysis of genes with dDNMs in affected offspring

Genes hit by dDNM in affected offspring represented a non-random subset of all genes tested. Functional enrichment analysis demonstrated that genes with a dDNM among offspring affected with a narrow diagnosis were significantly associated with neuron projection, nervous system development, and calcium ion transmembrane activity (Fig. 1a). Genes with dDNMs among offspring with a broad diagnosis were significantly enriched for functions associated with learning and postsynaptic organization (Fig. 1b). To test whether the observed dDNMs were agnostic to phenotype, we removed from the list the 44–58 genes that carried dDNMs in unaffected offspring and repeated the enrichment analysis. The results were essentially unchanged.

## Genes hit by recurrent dDNMs

The occurrence of multiple de novo events in a single gene, in a cohort of individuals with a common phenotype, may implicate that gene in the pathogenesis of the condition under study. To test this, we combined our results with those of three previously published studies of DNM in bipolar disorder <sup>9–11</sup> and compared them altogether. Five genes, *XKR6*, *MRC2*, *SUGP2*, *DICER1*, and *PLEC* showed recurrent dDNMs in the broad phenotype group. *XKR6* and *MRC2* were previously reported as recurrent hits by Nishioka et al. <sup>11</sup> The numbers of dDNMs per gene were compared to the expected values calculated by

denovolyzeByGene. Nominally significant excesses of dDNMs were detected in *SUGP2* (uncorrected  $P = 2.86 \times 10^{-4}$ ), *DICER1* (uncorrected  $P = 8.72 \times 10^{-4}$ ), and *PLEC* (uncorrected  $P = 0.0015$ ).

## Co-expression analysis of dDNMs

Supplementary Table S8 shows the six lists of genes used in the co-expression analysis: 1) genes with lof or 2) misD variants in “narrow” cases; 3) genes with lof or 4) misD variants in “broad” cases; 5) genes with lof or misD variants in controls.

**Hypergeometric overlap with reference modules.** The broad and narrow misD gene lists were found to significantly overlap with the PsychENCODE modules geneM11, geneM15, geneM18, geneM25, geneM29, geneM32, and geneM34<sup>39</sup>, above what would be expected by random chance per 1000 permutation tests (Fig. 2a, Table 3). The broad misD gene list also significantly overlapped with geneM15, while the narrow LoF gene list overlapped gene M5<sup>39</sup>. The broad and narrow misD and LoF gene lists all significantly overlapped with Hartl module BRNACC.M2<sup>40</sup> (Fig. 2a, Table 3). Notably, the module Hartl\_BRNACC.M2 is by far the most enriched for genes hit by lof DNMs (OR = 6.1–6.8), while PsychENCODE\_geneM34 is most enriched for misD (OR = 4.9–6.1) in the both phenotype groups (Table 3).

**Functional enrichment of overlapped modules.** GO functional enrichment followed by topic modeling revealed that PsychENCODE modules geneM5 and geneM32 are related to immune and inflammatory response, geneM11 and geneM25 to transcription and epigenetic modification, and geneM29, geneM34<sup>39</sup>, and Hartl BRNACC.M2<sup>40</sup> to transcription related to synaptic signaling (Fig. 2b & Fig. 2c). These results suggest a broad range of implicated biological functions, with an emphasis on immune response, transcription, and synaptic signaling.

**Cell type enrichment.** Reference modules geneM5, geneM15, and geneM32 were significantly enriched for endothelial cells, geneM15 was enriched for excitatory neurons, and geneM5 was enriched for microglia (Fig. 3a). The other reference modules were not found to be significantly enriched for one of the Lake et al. cell type categories<sup>42</sup>.

**Developmental trajectories.** The reference modules described above followed three general developmental expression trajectories: (1) High levels of prenatal expression and decreased expression throughout the lifespan (PsychENCODE geneM11, geneM29, geneM34, and Hartl BRNACC.M2); (2) Low levels of expression that increase during embryonic windows, decreasing during late pregnancy and early-life years, then increasing again from childhood through adulthood (geneM11 and geneM25); (3) Expression that increases through pregnancy into early-life years, then decreases through adulthood (geneM5 and geneM32). These developmental trajectories are shown in Fig. 3b. Together, these expression patterns implicate genes highly expressed prenatally and in infancy.

## Discussion

To our knowledge, this is the first study that investigated DNMs in coding regions in multiplex BD families from a genetic isolate, adding substantially to the limited set of BD trios that have been subjected to de novo analysis to date. The results support a role for DNM in BD and related disorders. This is also the first de novo study to investigate gene overlap with gene co-expression modules in human brain, thus implicating immune-related genes in the etiology of BD. This finding lends strong support to the previously proposed role for immune dysfunction in BD <sup>45</sup>.

The enrichment patterns of dDNMs showed a tendency for enrichment of lof DNMs in offspring with more severe phenotypes (BD, SczAD, and Scz) and enrichment of misD in offspring with milder phenotypes (MDD). These findings agree with previous studies of DNMs in BD <sup>11</sup> and other psychiatric disorders <sup>35,36</sup>.

Gene set enrichment analysis demonstrated that genes hit by dDNMs in BD and related conditions are enriched for functions related to synapses, learning, post-synaptic organization, nervous system development, and calcium ion transport. The pathway results were robust to removal of genes that carried dDNMs in unaffected offspring. Synapse and calcium ion channel genes are associated consistently with BD, as previously discussed <sup>11</sup>. Learning involves neuroplasticity, the promotion of which by lithium may be related to its therapeutic effects in BD. These results also highlight genes that are expressed in early life and involved in neurodevelopment, suggesting that at least some cases of BD and related disorders have a neurodevelopmental origin.

Genes recurrently hit by dDNM in this and in previous studies deserve further scrutiny. While we found no genes reaching the exome-wide significance threshold, we observed enrichment of dDNMs in *SUGP2*, *DICER1*, and *PLEC*, which have also been implicated in BD by other lines of evidence.

*SUGP2* was detected in BD-I subjects in both our study and the study of Goes et al. *SUGP2* encodes a member of the arginine/serine-rich family of splicing factors. The encoded protein functions in mRNA processing. In TWAS using TWAS hub (<http://twas-hub.org/>)<sup>46</sup>, which measures an association between gene expression and a complex phenotype using GWAS summary-level data, the model trained using transcriptome data of GTEx Brain Cerebellum showed significant associations between *SUGP2* and bipolar disorder and schizophrenia with a Z scores of 4.4 or higher. Open Targets Genetics (<https://genetics.opentargets.org/>)<sup>47,48</sup> that can highlight functionally involved genes by integrating functional and biological data from multiple disparate sources into GWAS summary data, *SUGP2* was significantly associated with BD-I.

*DICER1* was detected in SczAD subjects in both our study and the study of Fromer et al<sup>9</sup>. *DICER1* synthesizes DICER, a member of the ribonuclease III protein family that is involved in the generation of microRNAs (miRNAs), which regulate gene expression at the posttranscriptional level. MiRNAs are 22-nt-long RNAs generated from longer precursor RNAs. In general, miRNAs repress translation, but they can also acquire other functions after binding to their target RNA. Notably, many studies have implicated miRNAs in the development of psychotic disorders. DICER has an important role in the development and

function of the immune<sup>49</sup> and central nervous systems<sup>50</sup>. *DICER1* is upregulated in the dorsolateral prefrontal cortex<sup>50,51</sup>, and lymphoblastoid cell lines of schizophrenia cases<sup>52</sup>. In addition, *DICER1* single-nucleotide polymorphisms<sup>53</sup> and copy-number variations<sup>54</sup> are associated with schizophrenia. Interestingly, valproic acid, a mood stabilizer used to treat bipolar disorder, induces DICER degradation<sup>55</sup>. Similarly, lithium affects expression of let-7e, which is thought to target DICER mRNA<sup>56</sup>.

*PLEC* was detected in an MDD-R subject in our study and a BD-I subject in the study of Nishioka et al<sup>11</sup>. *PLEC* encodes plectin that is a prominent member of an important family of structurally and functionally related proteins, termed plakins or cytolinkers, that are capable of interlinking different elements of the cytoskeleton. Plakins, with their multi-domain structure and enormous size, not only play crucial roles in maintaining cell and tissue integrity and orchestrating dynamic changes in cytoarchitecture and cell shape, but also serve as scaffolding platforms for the assembly, positioning, and regulation of signaling complexes. Interestingly, *PLEC* lies within one of the most significant BD GWAS loci<sup>5</sup>, and *PLEC* was identified as one of the 56 candidate antidepressant response genes that are associated with electroconvulsive stimuli-induced recovery in a mouse model of depression<sup>57</sup>.

Co-expression analysis identified clusters, or modules, of genes hit by lof or misD DNMs with similar expression patterns in brain from donors with BD and related disorders. PsychENCODE modules gene M25 and geneM11, enriched for genes hit by misD DNMs in the narrow phenotype in our study, were associated with transcription and epigenetic modification. These modules were not enriched in GWAS of BD or Scz<sup>39</sup>. This suggests a distinction between genes hit by dDNMs and those regulated by common variants. Module geneM5, enriched for genes hit by lof DNMs in the narrow phenotype, and M32, enriched for genes hit by misD DNMs in the narrow and broad phenotypes, are immune-related modules that increase neuroinflammatory processes and are broadly expressed in signaling pathways. These modules were associated with BD, Scz, and ASD in PsychENCODE<sup>39</sup>. These findings support a role for inflammatory processes in BD as in other major psychiatric disorders<sup>58</sup>. Genes hit by dDNMs in the narrow or broad phenotypes were enriched in nucleus accumbens, a brain region that has been suggested to be associated with BD and Scz<sup>59</sup> (Fig. 2a, Table 3).

The present study has several limitations. First, the sample size is modest for a DNM study, limiting statistical power to detect exome-wide significant associations, particularly with individual genes. However, since this is a family sample, we were able to include relatives with SczAD, Scz, and MDD-R, thus increasing sample size and statistical power. Other potential limitations include a proportion of sequenced DNA extracted from LCLs. However, we did not observe a substantially higher rate of DNM in DNA from LCLs. While selected DNMs all validated by Sanger sequencing, we could not validate all dDNMs due to limited DNA availability. In addition, some probands had incomplete phenotype information and controls were limited to unaffected siblings who could develop mood disorders later in life. However, we augmented some of the case-control comparisons by use of unaffected siblings from an ASD study that were processed under a similar pipeline. We do not know how DNMs contribute to

familial BD; we speculate that DNMs add to existing polygenic and other inherited risk factors to modify penetrance, symptom severity, or associated impairment.

We have identified several rare, dDNMs among cases of BD and related conditions in this family sample, supporting an etiological role for both neurodevelopment and immunity in BD and related disorders. The results suggest that DNMs may be a genetic contributor to BD even in multiplex families. While further studies are needed, genes with recurrent dDNMs are good targets for functional genomic investigation.

## Declarations

### Acknowledgements

This work was supported in part by the NIMH Intramural Research Program (ZIA MH002843). RLS is funded by the NIH Oxford Cambridge Scholars Program. We appreciate the generous contributions of the participating families to this project. We thank Alan Shuldiner and the Regeneron Genetics Center for carrying out the exome sequencing. We thank Infinity BiologiX and Rutgers University for biobanking services. This work utilized the computational resources of the NIH HPC Biowulf cluster.

(<http://hpc.nih.gov>)

### Conflict of Interest

The authors declare no conflict of interests.

### Supplementary Information

The online version contains supplementary material available at MP's website.

### Availability of Data

All exome and phenotype data used in this study has been submitted to dbGAP (phs000899) and is available upon request to dbGAP by qualified researchers.

## References

1. Merikangas KR, Akiskal HS, Angst J, Greenberg PE, Hirschfeld RM, Petukhova M *et al.* Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Arch Gen Psychiatry* 2007; 64(5): 543–552.
2. Merikangas KR, Jin R, He JP, Kessler RC, Lee S, Sampson NA *et al.* Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Arch Gen Psychiatry* 2011; 68(3): 241–251.
3. Gordovez FJA, McMahon FJ. The genetics of bipolar disorder. *Mol Psychiatry* 2020; 25(3): 544–559.
4. Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* 2015;

- 47(7): 702–709.
5. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat Genet* 2021; 53(6): 817–829.
  6. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetsky V *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* 2019; 51(5): 793–803.
  7. Green EK, Rees E, Walters JT, Smith KG, Forty L, Grozeva D *et al.* Copy number variation in bipolar disorder. *Mol Psychiatry* 2016; 21(1): 89–93.
  8. Palmer DS, Howrigan DP, Chapman SB, Adolfsson R, Bass N, Blackwood D *et al.* Exome sequencing in bipolar disorder identifies AKAP11 as a risk gene shared with schizophrenia. *Nat Genet* 2022; 54(5): 541–547.
  9. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* 2014; 506(7487): 179–184.
  10. Goes FS, Pirooznia M, Tehan M, Zandi PP, McGrath J, Wolyniec P *et al.* De novo variation in bipolar disorder. *Mol Psychiatry* 2021; 26(8): 4127–4136.
  11. Nishioka M, Kazuno AA, Nakamura T, Sakai N, Hayama T, Fujii K *et al.* Systematic analysis of exonic germline and postzygotic de novo mutations in bipolar disorder. *Nat Commun* 2021; 12(1): 3750.
  12. Toma C, Shaw AD, Overs BJ, Mitchell PB, Schofield PR, Cooper AA *et al.* De Novo Gene Variants and Familial Bipolar Disorder. *JAMA Netw Open* 2020; 3(5): e203382.
  13. Nishioka M, Takayama J, Sakai N, Kazuno AA, Ishiwata M, Ueda J *et al.* Deep exome sequencing identifies enrichment of deleterious mosaic variants in neurodevelopmental disorder genes and mitochondrial tRNA regions in bipolar disorder. *Mol Psychiatry* 2023.
  14. Lau P, Hawes DJ, Hunt C, Frankland A, Roberts G, Mitchell PB. Prevalence of psychopathology in bipolar high-risk offspring and siblings: a meta-analysis. *Eur Child Adolesc Psychiatry* 2018; 27(7): 823–837.
  15. Vandeleur CL, Merikangas KR, Strippoli MP, Castelao E, Preisig M. Specificity of psychosis, mania and major depression in a contemporary family study. *Mol Psychiatry* 2014; 19(2): 209–213.
  16. Gill KE, Cardenas SA, Kassem L, Schulze TG, McMahon FJ. Symptom profiles and illness course among Anabaptist and Non-Anabaptist adults with major mood disorders. *Int J Bipolar Disord* 2016; 4(1): 21.
  17. Strauss KA, Puffenberger EG. Genetics, medicine, and the Plain people. *Annu Rev Genomics Hum Genet* 2009; 10: 513–536.
  18. Rasic D, Hajek T, Alda M, Uher R. Risk of mental illness in offspring of parents with schizophrenia, bipolar disorder, and major depressive disorder: a meta-analysis of family high-risk studies. *Schizophr Bull* 2014; 40(1): 28–38.
  19. Van Hout CV, Tachmazidou I, Backman JD, Hoffman JD, Liu D, Pandey AK *et al.* Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* 2020; 586(7831): 749–756.

20. Vasimuddin Md SM, Heng Li, Srinivas Aluru. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *IEEE Parallel and Distributed Processing Symposium (IPDPS)*. IEEE: Rio de Janeiro, Brazil, 2019.
21. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A *et al*. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013; 43(1110): 11 10 11–11 10 33.
22. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A *et al*. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; 20(9): 1297–1303.
23. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C *et al*. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; 43(5): 491–498.
24. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR *et al*. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* 2012; 91(5): 839–848.
25. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010; 26(22): 2867–2873.
26. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A *et al*. The Ensembl Variant Effect Predictor. *Genome Biol* 2016; 17(1): 122.
27. Rentzsch P, Schubach M, Shendure J, Kircher M. CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med* 2021; 13(1): 31.
28. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q *et al*. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020; 581(7809): 434–443.
29. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM *et al*. A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 2014; 46(9): 944–950.
30. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G *et al*. Integrative genomics viewer. *Nat Biotechnol* 2011; 29(1): 24–26.
31. Fu JM, Satterstrom FK, Peng M, Brand H, Collins RL, Dong S *et al*. Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat Genet* 2022; 54(9): 1320–1331.
32. Ware JS, Samocha KE, Homsy J, Daly MJ. Interpreting de novo Variation in Human Disease Using denovolyzeR. *Curr Protoc Hum Genet* 2015; 87: 7 25 21–27 25 15.
33. Dong W, Jin SC, Allocco A, Zeng X, Sheth AH, Panchagnula S *et al*. Exome Sequencing Implicates Impaired GABA Signaling and Neuronal Ion Transport in Trigeminal Neuralgia. *iScience* 2020; 23(10): 101552.
34. Diab NS, King S, Dong W, Allington G, Sheth A, Peters ST *et al*. Analysis workflow to assess de novo genetic variants from human whole-exome sequencing. *STAR Protoc* 2021; 2(1): 100383.



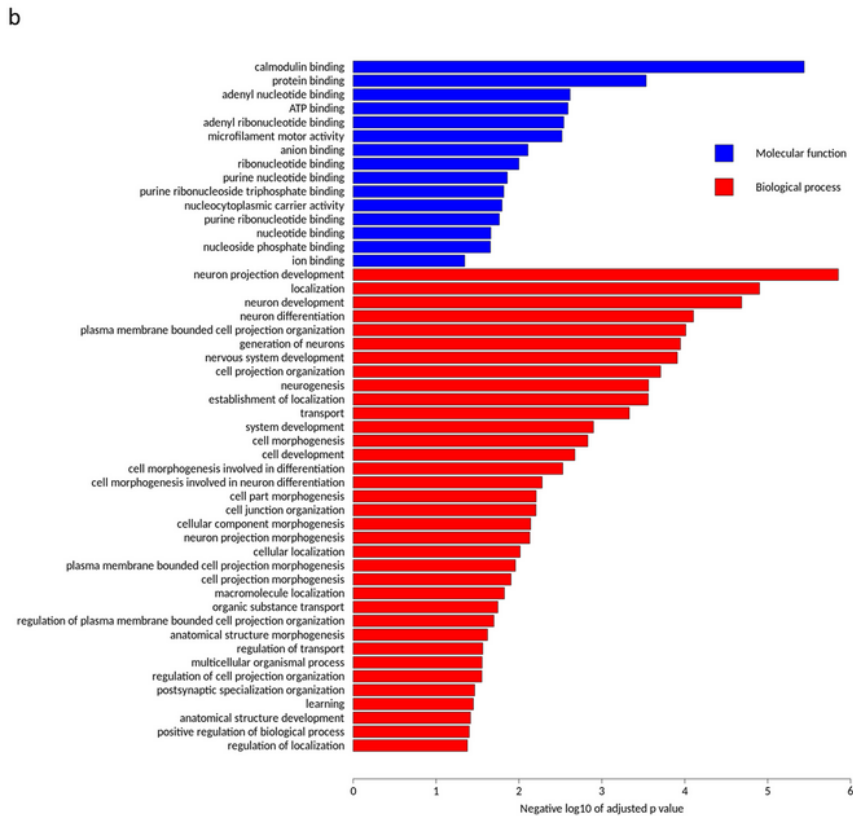
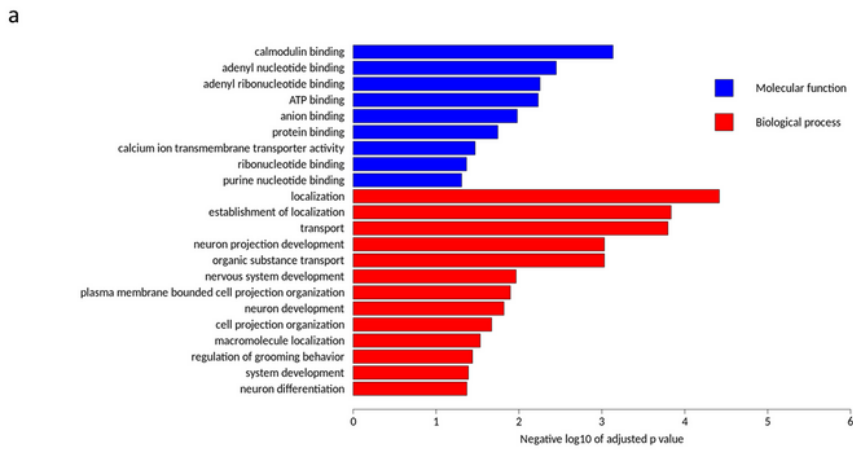
35. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; 515(7526): 216–221.
36. Howrigan DP, Rose SA, Samocha KE, Fromer M, Cerrato F, Chen WJ *et al.* Exome sequencing in schizophrenia-affected parent-offspring trios reveals risk conferred by protein-coding de novo mutations. *Nat Neurosci* 2020; 23(2): 185–193.
37. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* 2019; 47(W1): W191-W198.
38. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005; 4: Article17.
39. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* 2018; 362(6420).
40. Hartl CL, Ramaswami G, Pembroke WG, Muller S, Pintacuda G, Saha A *et al.* Coexpression network architecture reveals the brain-wide and multiregional basis of disease susceptibility. *Nat Neurosci* 2021; 24(9): 1313–1323.
41. Alexa AR, J. topGO: Enrichment Analysis for Gene Ontology. 2023.
42. Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol* 2018; 36(1): 70–80.
43. Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S *et al.* Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 2018; 362(6420).
44. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* 2020; 180(3): 568–584 e523.
45. Rosenblat JD, McIntyre RS. Bipolar Disorder and Immune Dysfunction: Epidemiological Findings, Proposed Pathophysiology and Clinical Implications. *Brain Sci* 2017; 7(11).
46. Mancuso N, Shi H, Goddard P, Kichaev G, Gusev A, Pasaniuc B. Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. *Am J Hum Genet* 2017; 100(3): 473–487.
47. Ghousaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A *et al.* Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res* 2021; 49(D1): D1311-D1320.
48. Mountjoy E, Schmidt EM, Carmona M, Schwartzentruber J, Peat G, Miranda A *et al.* An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat Genet* 2021; 53(11): 1527–1533.
49. Devasthanam AS, Tomasi TB. Dicer in immune cell development and function. *Immunol Invest* 2014; 43(2): 182–195.

50. Santarelli DM, Beveridge NJ, Tooney PA, Cairns MJ. Upregulation of dicer and microRNA expression in the dorsolateral prefrontal cortex Brodmann area 46 in schizophrenia. *Biol Psychiatry* 2011; 69(2): 180–187.
51. Beveridge NJ, Gardiner E, Carroll AP, Tooney PA, Cairns MJ. Schizophrenia is associated with an increase in cortical microRNA biogenesis. *Mol Psychiatry* 2010; 15(12): 1176–1189.
52. Sanders AR, Goring HH, Duan J, Drigalenko EI, Moy W, Freda J *et al*. Transcriptome study of differential expression in schizophrenia. *Hum Mol Genet* 2013; 22(24): 5001–5014.
53. Zhou Y, Wang J, Lu X, Song X, Ye Y, Zhou J *et al*. Evaluation of six SNPs of MicroRNA machinery genes and risk of schizophrenia. *J Mol Neurosci* 2013; 49(3): 594–599.
54. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 2008; 40(7): 880–885.
55. Zhang Z, Convertini P, Shen M, Xu X, Lemoine F, de la Grange P *et al*. Valproic acid causes proteasomal degradation of DICER and influences miRNA expression. *PLoS One* 2013; 8(12): e82895.
56. Hunsberger JG, Chibane FL, Elkahloun AG, Henderson R, Singh R, Lawson J *et al*. Novel integrative genomic tool for interrogating lithium response in bipolar disorder. *Transl Psychiatry* 2015; 5(2): e504.
57. Rooney AG, Kilpatrick AM, Ffrench-Constant C. Electroconvulsive stimuli reverse neuro-inflammation and behavioral deficits in a mouse model of depression. *bioRxiv* 2023.
58. Bechter K. The Challenge of Assessing Mild Neuroinflammation in Severe Mental Disorders. *Front Psychiatry* 2020; 11: 773.
59. Bayassi-Jakowicka M, Lietzau G, Czuba E, Patrone C, Kowianski P. More than Addiction-The Nucleus Accumbens Contribution to Development of Mental Disorders and Neurodegenerative Diseases. *Int J Mol Sci* 2022; 23(5).

## Tables

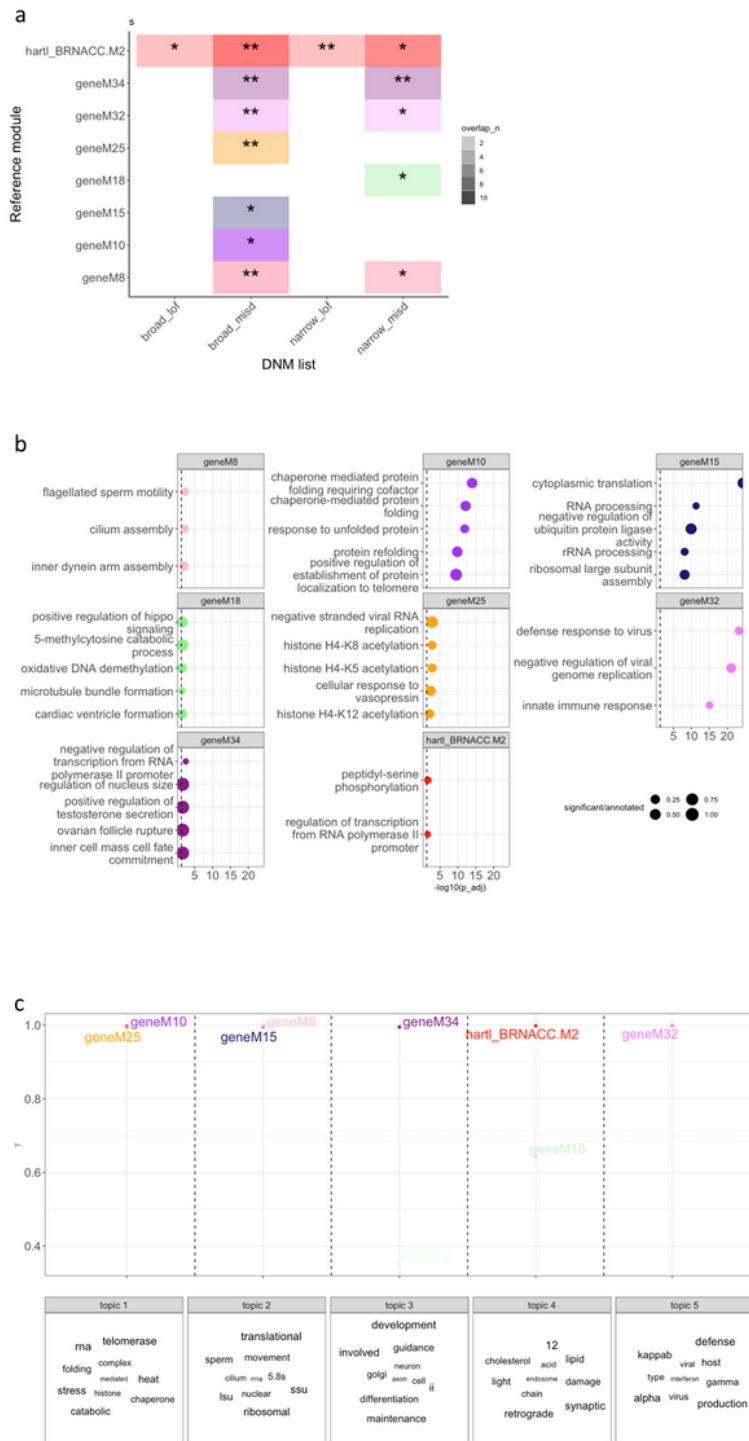
Tables 1-3 is available in the Supplementary Files section.

## Figures



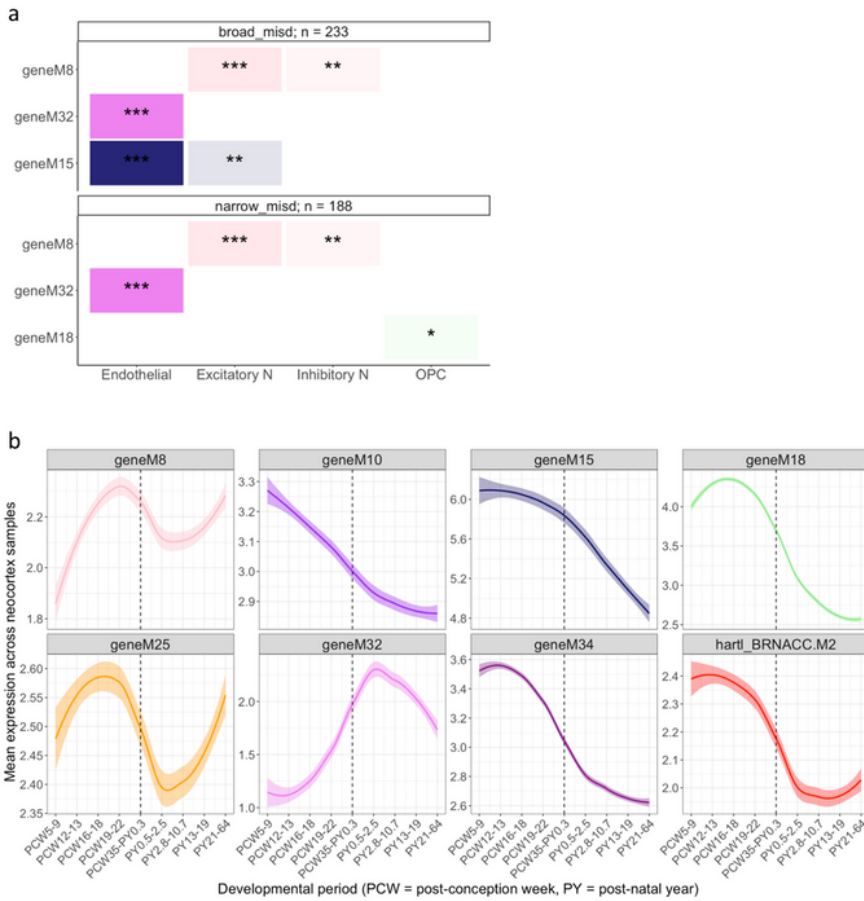
**Figure 1**

**g:Profiler analysis for the genes hit by dDNMs. a** Significantly enriched GO terms of the genes hit by dDNMs from narrow group in molecular function (MF) and biological process (BP). **b** Significantly enriched GO terms of the genes hit by dDNMs from broad group in MF and BP.



**Figure 2**

**Co-expression analysis for the genes hit by dDNMs. a** Significant overlaps between DNM lists and reference module by permutation test. **b** Gene ontology (GO) functional enrichment of reference modules. **c** Reference module strength of association with each topic.



**Figure 3**

**Characterization of reference modules that significantly overlap with the genes hit by dDNMs. a** Cell-type enrichment of reference modules that significantly overlap with gene lists. **b** Average gene expression per module across development.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS1.tif](#)
- [FigS2.tif](#)
- [Table1.xlsx](#)
- [Table2.xlsx](#)
- [Table3.xlsx](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)
- [TableS7.xlsx](#)
- [TableS8.xlsx](#)