

## **Coding and non-coding RNA dysregulation in bipolar disorder**

Jurjen J. Luykx<sup>\*1,2,3#</sup>, F. Giuliani<sup>\*1</sup>, G. Giuliani<sup>1</sup>, J.H. Veldink<sup>1,4</sup>

<sup>1</sup> Department of Translational Neuroscience, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

<sup>2</sup> Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht (UMCU), Utrecht University, Utrecht, the Netherlands

<sup>3</sup> Department of Psychiatry, ZNA Hospitals, Antwerp, Belgium

<sup>4</sup> Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht (UMCU), Utrecht University, Utrecht, the Netherlands

\* These authors equally contributed to this work.

# To whom correspondence should be addressed at:

Jurjen J Luykx | Brain Center Rudolf Magnus | University Medical Center Utrecht, Universiteitsweg 100, office 4.127 (Stratenum), HP 4.205 | 3584 CG Utrecht | The Netherlands | +31 (0)88 75 68638 | [j.luykx@umcutrecht.nl](mailto:j.luykx@umcutrecht.nl)

Word count abstract: 192

Word count manuscript: 3146

Number of figures: 3

Number of tables: 1

Running title: Coding and non-coding RNAs in bipolar disorder.

Keywords: bipolar disorder, circular RNA, long non-coding RNA, alternative splicing, cEPHA3, histone H3-K4 demethylation.

## Abstract

The molecular mechanisms underlying bipolar disorder (BPD) have remained largely unknown. Postmortem brain tissue studies comparing BPD patients with healthy controls have produced a heterogeneous array of potentially implicated protein-coding RNAs. We hypothesized that dysregulation of not only coding, but multiple classes of RNA (coding RNA, long non-coding (lnc) RNA, circular (circ) RNA, and/or alternative splicing) underlie the pathogenesis of BPD. Using non-polyadenylated libraries we performed RNA sequencing in postmortem human medial frontal gyrus tissue from BPD patients and healthy controls. We found twenty genes, some of which not previously implicated in BPD, differentially expressed. PCR validation and replication confirmed the implication of these DE genes. Functional analyses identified enrichment of angiogenesis, vascular system development and histone H3-K4 demethylation. In addition, ten lncRNA transcripts were differentially expressed. Furthermore, we detected an overall increased number of alternative splicing events in BPD, as well as an increase in the number of genes carrying alternative splicing events. Finally, we report altered levels of two circular transcripts, *cNEBL* and *cEPA3*. In conclusion, our non-coding RNA findings demonstrate that RNA dysregulation in BPD is not limited to coding regions, opening avenues for future pharmacological investigations and biomarker research.

## Introduction

Bipolar disorder (BPD) is characterized by one or more manic episodes alternating with euthymia or depressive episodes. Bipolar spectrum disorders carry a lifetime prevalence of up to 2% and are estimated to be the seventeenth leading cause of disability worldwide <sup>1</sup>. They entail increased risks of comorbid conditions and premature death (in part owing to suicide) <sup>2,3</sup>. The diagnosis of BPD is often challenging due to its clinical heterogeneity and, consequently, diagnostic delays are common. Moreover, response to currently available treatment modalities is seldom complete and burdensome adverse reactions often emerge. Such diagnostic pitfalls and imperfect treatments may all benefit from neurobiological studies aimed at increasing our understanding of its pathophysiology.

To date, the biological mechanisms underlying BPD remain largely elusive, although research hints that both genetic and environmental factors contribute. For example, twin-based heritability estimates vary between 70 and 90% <sup>4</sup>. Recent studies aimed at elucidating the molecular basis of BPD have mostly relied on genetic approaches, including genome-wide association (GWA) <sup>5,6</sup> and whole-exome sequencing studies <sup>7-9</sup>, to identify both inherited and *de novo* variation contributing to BPD. Such genome-wide studies have not provided the biological insight as has been the case for schizophrenia <sup>10</sup>. Moreover, three out of five top BPD GWA hits harbor non-coding RNA loci <sup>6</sup>. Therefore, gene expression analysis of the relevant brain regions constitutes a primordial step to help identify the molecular pathways altered in BPD.

In one of the first comprehensive gene expression analyses in BPD, peripheral blood cells for microarray-based transcriptome analysis were used to identify changes in levels of transcripts involved in G-protein signalling <sup>11</sup>. More recently, next generation sequencing (NGS) technologies have been used to survey the brain transcriptome. RNA sequencing of hippocampus, the anterior cingulate gyrus, the dorsolateral prefrontal cortex, and the dorsal striatum of BPD postmortem tissue has identified genes linked to G-protein coupled receptors, circadian rhythm, the immune system, inflammatory response and metabolic pathways <sup>12-17</sup>. However, these RNAseq experiments were designed to capture polyadenylated RNA transcripts - which include protein-coding mRNAs and a number of non-coding RNAs -, while most RNAs (>90% of the transcriptome) do not carry a polyadenylated tail. On a similar note, NGS experiments indicate that less than 5% of transcription across the human genome results in protein-coding genes, while the remaining pool is associated with non-protein coding transcripts <sup>18</sup>, approximately 60% of which belong to the class of long non-coding RNAs (lncRNAs) <sup>19</sup>. To date, only a few lncRNAs have been characterized at the molecular or functional level but their dysregulation is being increasingly reported in cancer and in numerous neurological, cardiovascular, and developmental diseases <sup>20-24</sup>. Furthermore, although the

dorsolateral prefrontal gyrus has been targeted <sup>12</sup>, other sections of the (pre)frontal gyrus have been left unused in RNA sequencing studies of postmortem brain tissue in BPD patients. This absence of frontal gyrus RNA sequencing studies in BPD is in sharp contrast to the currently available impressive body of literature hinting at the implication of particularly the medial frontal cortex in BPD. For example, meta-analytic evidence points to medial frontal gray matter reductions in BPD compared to controls <sup>25</sup>, resting-state connectivity aberrations in the medial frontal cortex <sup>26</sup>, and altered activity in this area in BPD based on fMRI studies <sup>27</sup>. Thus, because the medial part of the frontal gyrus has been particularly implicated in BPD by a range of studies we set out to obtain frozen sections of this brain region for RNA sequencing.

Sequencing experiments of non-polyadenylated transcriptomes have led to the discovery of new RNA classes, such as circular RNAs (circRNAs), a category of lncRNAs produced by back-splicing reactions that covalently link the 3' end of an exon to the 5' end of an upstream exon <sup>28-30</sup>. circRNAs have been implicated in gene regulation, by functioning as molecular sponges to regulate gene expression of microRNAs, sequestering RNA binding proteins and competing with other lncRNAs <sup>31-33</sup>. Recent studies have shown that circRNAs and other lncRNAs also play pivotal roles in brain development and neuronal integrity <sup>34-40</sup>. Non-polyadenylated RNAseq libraries also allow probing of alternative splicing, a process that not only generates protein diversity, but also constitutes a means to regulate gene expression post-transcriptionally. Aberrant splicing may lead to the production of transcripts that could encode potentially deleterious proteins. However relevant non-coding RNAs may be in disease, these have to the best of our knowledge not been comprehensively examined in BPD brain tissue.

To comprehensively probe the implication of numerous RNA classes in BPD, we performed the first multi-class RNA sequencing experiment in the frontal gyrus of BPD patients, regardless of polyadenylation status. We thus compared the relative abundances of both protein-coding and non-coding RNAs in BPD with healthy controls. We find several differentially expressed genes (DEGs) and long lncRNA transcripts.



## Results

### Baseline characteristics

Baseline characteristics were similar for BPD patients and controls (Table 1 and Supplemental Table, sheets 1 and 2). The average number of reads was ~63 million with 71% of all reads mapping to the human genome. To rule out that other variables than disease status (e.g. RNA integrity, age, cell type composition, and postmortem interval) may be driving case-control differences, we first carefully matched cases and controls, then performed gene expression principal components analysis (PCA, Supplemental results, figure 1) and finally validation and replication experiments.

### Coding RNA: differential expression, network and enrichment analyses

At  $FDR < 0.05$ , our differential expression analysis revealed 20 significantly differentially expressed genes (DEGs; Supplemental Table, sheet 4). At  $FDR < 0.1$ , we found 36 DEGs, all of which were upregulated in BPD (Figure 1A). RT-PCR and RT-qPCR (Supplemental Methods) were then used to 1) validate our differentially expressed gene (DEG) findings within the discovery cohort; and 2) replicate these findings in an independent replication cohort. To that end, we selected primers for the top three DEGs and two randomly chosen DEGs. We thus successfully validated this entire predefined subset of DEGs (*CD93*, *SOCS3*, *BCL6B*, *PODXL* and *ABCA1*; Figure 1B and Supplemental Table, sheet 5). We also successfully replicated four out of five of these genes (*CD93*, *SOCS3*, *PODXL* and *ABCA1*) in our independent replication cohort (Figure 1C and Supplemental Table, sheet 5). Although *BCL6B* did not statistically replicate in this cohort, its expression level aligned with the RNAseq data (Figure 1C). In addition, we found a significant correlation in log<sub>2</sub>-fold change values between all genes in our dataset and both published datasets<sup>12</sup> ( $r=0.39$  and  $r=0.29$ ; both p-values  $< 2.2 \times 10^{-16}$ ; Supplemental Results, figure 2).

We then proceeded to assess cell-type specificity of the top five differentially expressed genes and found all but one of those genes having more than 10-fold higher expression levels in endothelial tissue than any other cell type (Supplemental Results, figures 3A-E).

Using weighted gene co-expression network analyses (WGCNA) we were unable to identify any significantly correlated modules. Gene ontology (GO) term enrichment analysis, however, showed significant enrichments of biological processes related to angiogenesis and vascular system development (Supplemental Results, figure 4A and Supplemental Table, sheet 6). In our gene-set enrichment analysis (GSEA), we identified enrichment in the GO term GO:0034720 "Histone H3-K4 demethylation" (FDR = 0.00017; Supplemental Results, figure 4B and Supplemental Table, sheet 6): the expression levels of genes associated with this GO term were upregulated in BPD compared to controls.

### **Alternative splicing (AS) events**

The total number of AS events was higher in BPD subjects than in controls (137,017 in BPD versus 129,294 in controls,  $p = 1.23 \times 10^{-50}$ ; Supplemental Results, figure 5 and Supplemental Table, sheet 7). Many of these AS events occur in DEGs (Supplemental Table, sheet 8) and in BPD-susceptibility genes (Supplemental Table, sheet 9; derived from whole-exome sequencing studies<sup>7-9</sup>). It is currently unknown whether DEGs may sometimes be explained by AS events. We also observed significant increases in single (ESI) and multiple (MESI) exon skipping, intron retention (IRI), and alternative first (A5) and last exon (A3) usage in BPD brains of 4.4%, 5.9%, 6.3%, 8.3% and 8.9% (all five adjusted p-values  $\leq 6.5 \times 10^{-6}$ ; Supplemental Table, sheet 7), respectively. The number of mutually exclusive exon (MEE) events was similar for BPD patients and controls. The total number of genes carrying AS events was also increased in BPD patients (adjusted  $p = 1.3 \times 10^{-13}$ ; Supplemental Table, sheet 7).

The functional analysis of BPD-specific AS events indicated significant GO term enrichment for categories associated with cell-cycle, post-translational protein modification, protein transport/localization, RNA splicing and processing for the ESI, MESI and IRI classes; in addition, A5 BPD-specific events were significantly overrepresented in the mitochondrial respiratory chain category (Supplemental Results, Figures 6A-E).

### **Differentially expressed long non-coding RNAs in BPD**

We detected expression from 27,348 lncRNA transcripts in both patients and controls. These detected lncRNAs exhibited an overall lower expression than protein-coding genes ( $t = 42.27$ ;  $p = 1.08 \times 10^{-9}$ ; Supplemental Results, figure 7A-B). The differential expression analysis identified a total number of ten differentially expressed lncRNAs (six upregulated and four downregulated lncRNAs) in BPD patients at  $FDR < 0.05$ , while at  $FDR < 0.1$  these numbers were 11 and four, respectively (Figure 2 and Supplemental table, sheet 10).

### **Circular RNA transcripts in healthy and affected brain tissue**

We identified a reservoir of 22,530 circRNAs with a minimum of two reads spanning the back-splice junction in the medial frontal gyrus. Of the 22,530 circRNAs identified, 21,019 come from 5,906 annotated loci (3,532 of which give rise to multiple circles), while 1,511 circRNAs map outside the genomic region of known genes (Figure 3 A-B; Supplemental Table, sheet 11). When applying a log fold change cutoff of  $\geq 1.5$  or  $\leq -1.5$  irrespective of statistical significance, 689 circRNAs were either upregulated or downregulated in BPD (Supplemental Table, sheets 12 and 13), one of which (*ZDHHC11*) overlapped with the DEGs. Three additional circRNA transcripts encompass this locus (Supplemental Table, sheet 14). At the 104 BPD risk genes based on the whole-exome sequencing

studies, 5 circRNA transcripts are upregulated at a log fold change cutoff of  $\geq 1.5$  and 1 is downregulated at  $\leq -1.5$  (Supplemental Table, sheet 15).

Differential expression analysis of the predicted circRNA transcripts between patients and controls identified one circRNA at  $FDR < 0.05$  (*cNEBL*) and one additional circRNA at  $FDR < 0.1$  (*cEPHA3*; Figure 3C). As we did not find significant differences in expression of linear *NEBL* and *cEPHA3* (Supplemental Table, sheet 16), the differential expression of these transcripts between cases and controls is unlikely to be driven by their linear counterparts.

## Discussion

We here demonstrate differences in both coding and non-coding RNA segments between bipolar disorder (BPD) and healthy controls. By applying the bioinformatics pipeline that we developed to total RNA libraries we first signal several differentially expressed genes and then validate and replicate our findings by PCR. We then demonstrate that alternative splicing events are increased in BPD and that expression levels of several lncRNAs differ between cases and controls. Finally, we describe a reservoir of circular RNAs in human medial frontal cortex and altered levels of circular RNA transcripts in BPD.

Functional annotation analysis of the differentially expressed genes (DEGs) showed significant enrichments of biological processes related to angiogenesis and vascular system development. Possibly, these DEGs would not have been picked up with a polyadenylated library. The results of our cell-type specificity analyses pointing to endothelial tissue agree with vascular tissue being implicated in bipolar disorder. This alignment between pathway and cell-type analyses is intriguing since, to our knowledge, no other study had reported a role for genes involved in angiogenesis and vascular development in BPD. Meta-analyses of fMRI studies in BPD point to frontal hypoactivation<sup>41,42</sup>. Based on our findings, one may postulate that impaired angiogenesis may be at the root of such hypoactive areas in the BPD brain; on the other hand, increased angiogenesis may constitute a compensatory mechanism for such hypoactivity. Whether microvascular changes have a role in the *neurobiology* of BPD remains to be unraveled. An alternative explanation for this pathway finding relates to medication use in BPD as lithium and valproic acid promote angiogenesis<sup>43,44</sup>, even in infarcted brain areas<sup>45</sup>. A potential mechanism at play here may be that these agents increase expression levels of genes involved with angiogenesis. Since all study participants had at death or previously used one of these two mood stabilizers (Supplemental Table, sheet 2), our results may reflect a consequence of treatment rather than disease mechanisms. Future studies aimed at collecting postmortem brain tissue of medication-naïve BPD patients would be highly laborious and time-consuming, yet not unconceivable.

Using gene-set enrichment analysis (GSEA) we identified enrichment in histone H3-K4 demethylation: the expression levels of genes associated with this GO term were upregulated in BPD brains compared to controls. Numerous studies indicate that histone modifications have important roles in BPD<sup>46</sup>. One of the best characterized histone modifications is H3-K4 methylation, an epigenetic phenomenon highly enriched at transcriptional start sites (TSS) and associated with active transcription. The design of the current study did not allow us to investigate H3-K4 methylation status on the TSS of genes in BPD brains. Future studies integrating gene expression with methylation probes within the same cohort may enable such analyses.

CircRNAs are increasingly being discovered as a functionally important kind of ubiquitous, endogenous non-coding RNAs<sup>32,33</sup>. Recent evidence reveals that circRNAs function as miRNA sponges<sup>31</sup> and regulate parent gene expression, predisposing subjects to certain diseases<sup>32,33</sup>. Despite the importance of circRNAs to several types of cancer<sup>47</sup>, the contribution of this highly stable class of RNAs to (neuro)psychiatric disorders, including BPD, had not been explored. Here, we reveal two circRNAs to be upregulated in BPD. These circRNAs derive from the *NEBL* and *EPHA3* loci, respectively. Particularly *EPHA3* may be of relevance to BPD as it is implicated in developmental events, mostly in the central nervous system (CNS). The Eph receptor belongs to the ephrin receptor subfamily of the protein-tyrosine kinase family. Eph receptors and ephrins (their ligands) regulate key CNS processes, such as neurotransmitter release, postsynaptic glutamate receptor conductance, synaptic glutamate reuptake, and dendritic spine morphogenesis<sup>48</sup>. They are involved in memory formation<sup>49</sup> and anxiety<sup>50</sup>. Memory deficits (during episodes and interepisodic) and high anxiety levels (during episodes) are both often reported in BPD. It is therefore of interest that Ephs and ephrins (their ligands) may be pharmacologically targeted<sup>48,51,52</sup>: EphB2-blockade impedes physiological stress responses in wild-type mice<sup>50</sup>. In addition to serving as therapeutic targets, traces of RNAs may serve as markers to monitor disease progression, as has been demonstrated for cancer<sup>47</sup>. Although one previous whole-transcriptome study on BPD brain tissue also used an rRNA-depleted library, circRNAs and AS events were not reported, while lncRNA results did not exceed significance thresholds<sup>14</sup>. If future studies confirm our BPD circRNA findings in peripheral tissues, circRNA molecules could one day aid in the diagnostic workup of patients suspected of a bipolar spectrum disorder.

Previous RNAseq studies in BPD have focused on other brain regions, in particular hippocampus and dorsal striatum tissue where immune response pathways were found to be implicated<sup>15,16</sup>. A different study on hippocampus tissue detected aberrant miR-182 signaling<sup>13</sup>. Other authors deep sequenced dorsolateral prefrontal cortex and detected dysregulation of neuroplasticity, circadian rhythms and GTPase binding in BPD<sup>12</sup>. Finally, anterior cingulate sequencing revealed G protein-receptor dysregulation<sup>14</sup>. The current findings thus add to a quite heterogeneous array of biological pathways implicated in BPD based on whole transcriptomics studies, which in part may be owing to the disparaging brain regions targeted in these studies. Future studies incorporating extremely large sample sizes while comparing coding and non-coding RNAs in all possible brain regions of BPD patients with healthy controls may disentangle the state and region-specificity of such pathways to the disease.

When comparing the RNA classes for which we detected case-control differential expression, it came to our attention that the ratio of DE RNAs to all RNAs in a specific class is highest for genes, followed by lncRNAs and then circRNAs. We believe such differences cannot be accounted for by technical

characteristics of our experiments. Moreover, based on our results and the current knowledge of disease mechanisms at play in BPD, we cannot explain such differences.

Several limitations should be borne in mind when interpreting the results of the current study. First, the modest sample size may have impeded power to detect any significant modules in our WGCNA. Second, having chosen a non-polyadenylated library as opposed to a small RNA library by consequence precluded us from examining miRNAs, molecules of particular relevance to schizophrenia. Third, having focused on coding and non-coding RNA in rather unique and scant sections of postmortem tissue prevented us from designing analyses at the protein level, a limitation that future studies may be able to overcome. And finally, while our circRNA findings in both healthy and affected brain tissues are intriguing, this relatively unexplored domain of RNA molecules currently does not allow cross-validation in other datasets. These results consequently will have to await confirmation in other cohorts. Moreover, future studies may disentangle the specificity of our findings to BPD by incorporating other psychiatric disorders.

In conclusion, our study shows that RNA dysregulation in bipolar disorder is not limited to coding regions.

### **Conflict of interest:**

All authors declare they have no conflict of interest.

### **Funding:**

This project was supported by a personal grant from the Rudolf Magnus Young Talent Fellowship program of the University Medical Center Utrecht to Jurjen Luykx.

### **Acknowledgements:**

We are indebted to the Netherlands Brain Bank (NBB) and the individuals and their family members who conceded us the opportunity to make use of the participants' brain tissue and their phenotypic information.

## Materials and Methods

### Study population

All methods were carried out in accordance with relevant guidelines and regulations. Postmortem material from all BPD and healthy control subjects used in this study was obtained from the Netherlands Brain Bank (NBB), Netherlands Institute for Neuroscience, Amsterdam, the Netherlands (open access: [www.brainbank.nl](http://www.brainbank.nl)). All experimental protocols were approved by the NBB. All material has been collected from donors for or from whom written informed consent was obtained by the NBB. The informed consent includes permission for a brain autopsy and the use of the material and clinical information for research purposes. A first cohort that consisted of four cases and four controls -matched on gender, age, post-mortem interval (<9h), and pH (6.0 - 7.0)- was used for RNA sequencing experiments. A second cohort also consisting of four cases and four controls and matched on age, post-mortem interval (<9h), and pH (= 6.0 - 7.0) was used for replication experiments. Extensive phenotype information available from the NBB was used to verify diagnosis and healthy control status. An overview of detailed phenotypic information (including medication) is presented in the Supplemental Table, sheets 1 and 2.

### Human tissue and RNA processing

Frozen medial frontal gyrus tissue from each subject was powdered with the help of a grinder in a pre-chilled mortar to prevent thawing. Total RNA was then extracted from 20–40 mg of the powdered frozen samples using the RNeasy Plus Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and then DNase-treated to eliminate genomic DNA contamination. Finally, RNA integrity (RIN) was verified on an Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and only samples with RIN  $\geq$  5.0 were used for RNA sequencing. Detailed information about postmortem tissue conditions is provided in the Supplemental Table, sheet 3.

### RNAseq library preparation, sequencing and alignment

A total of 1  $\mu$ g of RNA per sample in a 5- $\mu$ l volume was rRNA depleted using the Ribo-Zero Magnetic Kit (human/mouse/rat; Epicentre) before library preparation. Stranded, paired-end sequencing libraries were prepared by the ServiceXS sequencing facility (BaseClear, Leiden, The Netherlands) and sequenced on the Illumina HiSeq 2500 platform (Illumina, San Diego, California). The reads (2x100 bp, mapping to both exons and introns) were de-multiplexed and converted to FASTQ format using CASAVA software from Illumina by the ServiceXS sequencing core facility. FASTQ files were mapped to the hg19/GCh37 reference human genome (iGenomes) with TopHat2 (version 2.0.13)<sup>53</sup>, using the 'fr-firststrand' option for strand orientation to generate BAM-formatted genomic coordinates. For lncRNA alignment, reads were aligned using Tophat2 aligner against the NONCODEv4<sup>54</sup> reference database that contains 54,073 human annotated lncRNA sequences.

### Differential Expression (DE) and Weighted Gene Co-expression Network Analyses (WGCNA)

Count data for genes and lncRNA transcripts were analyzed in R ([www.r-project.org](http://www.r-project.org)) using the Bioconductor package edgeR version 3.12.1<sup>55</sup> with the trimmed mean of M-



values (TMM) normalization method<sup>56</sup>. A generalized linear model was used to test the null hypothesis of absence of differential expression between the two groups. Gene expression levels were corrected for gender effect by including sex as a covariate in the model. P-values adjusted for multiple testing were calculated using Benjamini Hochberg false discovery rate (FDR) and only genes at FDR<0.05 were considered significantly DE. Volcano plots and expression MA scatter plots were generated in R using the ggplot2 library. To establish in which cell types the differentially expressed genes are expressed, we used an online tool that provides gene-based transcriptomics results parsed by cell type<sup>57</sup>. This tool outputs fragments per kilobase of transcript sequence per million mapped fragments (FPKM) in each possible brain cell type (astrocytes, neurons, oligodendrocyte precursor cells (OPCs), newly formed oligodendrocytes, myelinating oligodendrocytes, microglia and endothelial cells).

WGCNA analysis methods are described in the Supplemental Methods.

### **Validation and replication analyses**

Lower RIN values than 5 were allowed for validation and replication experiments, which is in line with previous findings demonstrating reliable expression data derived from tissues with fairly low RIN values in qt-PCR<sup>58</sup>. No data were available publicly or available in this cohort to account for cell type composition. The unpaired Welch t-test was used to examine case-control differences in gene expression levels in these validation and replication steps, with a 95% confidence cutoff.

Furthermore, to test the concordance between our DEG results and previous RNAseq findings in BPD, using Pearson correlation analyses in R ([www.r-project.org](http://www.r-project.org)) we compared the log<sub>2</sub>-fold change values of all genes in our dataset with two independent brain tissue RNAseq datasets<sup>12</sup>.

### **Gene Ontology (GO) term and Gene Set Enrichment Analyses (GSEA)**

DEGs at FDR<0.1 were entered into the GO-term R package goseq<sup>59</sup> to correct for bias due to transcript length. Gene set tests were conducted using the *fry* function<sup>60</sup> that runs an infinite number of rotations to test whether a set of genes is differentially expressed by assessing the entire set of genes as a whole.

### **Alternative splicing (AS) detection**

To identify alternative splicing events, we used the Bioconductor package spliceR<sup>61</sup> that uses the output from any full-length RNA-seq assembler to seek for single (ESI) or multiple (MESI) exon skipping, alternative donor and acceptor sites, intron retention (IRI), alternative first (A5) or last exon (A3) usage, and mutually exclusive exon (MEE) events. We ran the spliceR workflow using output from the RNAseq assembler Cufflinks<sup>62</sup> to detect such events. As spliceR does not generate summary statistics we performed a chi-squared test to assess whether the total number of AS events and the numbers of each of these classes differed between cases and controls. We also ran a chi-squared test to examine any possible differences between the number of genes carrying each class of AS events and the total number of genes potentially carrying AS events. We again adjusted the p-values for multiple testing by using Benjamini Hochberg FDR correction and only AS events with FDR<0.05 were considered significant. Enrichment analysis of associated transcripts at FDR<0.1 was performed with the Bioconductor package



GOstats <sup>63</sup> that uses the hypergeometric distribution to test for overrepresentation. To interpret those results reliably, the resulting GO terms and FDR adjusted p-values were analyzed with the GO interpretation tool REVIGO <sup>64</sup> to group the terms and reduce redundancy using “medium” as the size of the output list of terms and “Homo sapiens” as reference.

### **Detection and expression analysis of circular RNAs**

Detection of circRNA was performed using the python scripts of the find\_circ pipeline <sup>32</sup>, which is available from circBase at <http://circbase.org/cgi-bin/downloads.cgi> <sup>32</sup>. Briefly, after FASTQ alignment using Tophat2, the unmapped reads were used as inputs for the find\_circ tool to identify back-splice sites as described by the developers. Since find\_circ returns bed files, IDs were assigned by matching the genomic coordinates of every identified circle to those of its host gene by BEDTools <sup>65</sup>. For differential expression, count data were analyzed in R using the edgeR with TMM normalization in the same generalized linear model as explained above for gene and lncRNA differential expression analyses. To reduce the likelihood of type-I errors (although at the cost of increased type-II error risk), we set the FDR significance threshold for this first analysis of circRNAs in BPD at <0.1.

### **Code availability**

Codes used for the statistical analyses are available upon request.

## Table, figure and supplement legends:

Table 1 Demographic information for the subjects used in this study.

Abbreviations: BPD, bipolar disorder; F, female; M, male; PMI, postmortem interval.

Figure 1. Differentially expressed genes analyses.

Plot showing the 36 significant differentially expressed genes (DEGs) at  $FDR < 0.1$  in subjects with BPD versus healthy controls (A). Validation (B) and replication (C) of RNAseq gene expression analysis by RT-qPCR.

RNA level differences were calculated by unpaired Welch t-test (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.005$ ). All 5 DEGs were successfully validated by RT-PCR and had significant increased expression levels in BPD patients compared to controls (B). Four out of five DEGs were successfully replicated and had increased expression levels in BPD patients compared to controls; only BCL6B did not replicate significantly, although its expression level aligned with the sequencing data (C).

Figure 2. Differential regulation of lncRNAs in medial frontal gyrus in bipolar disorder.

This MA plot depicts the 15 differentially expressed lncRNA transcripts at  $FDR < 0.1$ . On the x-axis, the logCPM is the log-transformed average expression level (expressed in CPM, counts per million mapped reads) for each gene across the two groups; on the y-axis, the logFC represents the log of the ratio of expression levels for each gene between two experimental groups. In red are the upregulated transcripts; in blue the downregulated transcripts.

Figure 3. Differential expression of circRNAs in BPD brains.

Pie-chart representing the numbers and the percentages of the circRNAs from annotated loci and those from unknown loci, respectively (A). Pie-chart representing the numbers and percentages of loci producing only one circRNA and those giving rise to multiple circles, respectively (B). MA plot showing the two differentially expressed circular RNAs at  $FDR < 0.1$  in subjects with BPD versus healthy controls (C).

## Supplements:

Supplemental Table.

- Sheets 1 and 2: Detailed demographic information of the study population. Abbreviations: BPD, bipolar disorder; F, female; M, male; NA, not available; PMI, postmortem interval; Side, brain side; L, left; Mood stabilizers at death, use of mood stabilizers in the last 24 hours of life; Psychoactive drugs at death, use of psychotropic medications in the last

- 24 hours of life; Psychoactive drugs, ever-reported use of psychotropic medications; ECT, history of electroconvulsive therapy; 'Y' represents if there is a record of an individual having taken at least one medication/therapy and 'N' if there is no such record.
- Sheet 3: Postmortem brain tissue information. Outlined are the postmortem brain conditions per subject.
  - Sheet 4: Differential regulation of gene expression in medial frontal gyrus in bipolar disorder. Summary table of the differentially expressed genes in BPD medial frontal gyrus at  $FDR < 0.05$  (bold) and  $FDR < 0.1$ .
  - Sheet 5: Results of the validation and replication analyses. Shown are the summary statistics for the 5 genes in the validation and replication steps.
  - Sheet 6: Summary statistics for GO term enrichment analysis on the 36 differentially expressed genes at  $FDR < 0.1$  (A); and gene-set enrichment analysis (GSEA) for BPD-associated gene sets (B).
  - Sheet 7: Overview of chi-squared test results for each of the AS (alternative splicing) classes and the total number of AS events. Also shown are the numbers of genes carrying such AS events.
  - Sheet 8: Overlap between AS events and DEGs.
  - Sheet 9: AS events in BPD susceptibility genes.
  - Sheet 10: Summary table of the differentially expressed lncRNA transcripts.
  - Sheet 11: Overview of circular RNAs populating the medial frontal gyrus (first 2000 rows).
  - Sheet 12: When applying a log fold change cutoff of  $\geq 1.5$  or  $\leq -1.5$  irrespective of statistical significance, 689 circRNAs were either upregulated or downregulated in BPD
  - Sheet 13: This is the full list of circRNA toptags from the DE analysis, where all the logFC and the FDR of all circRNA transcripts included in the DE analysis can be found.
  - Sheet 14: Three additional circular transcripts generate the ZDHHC11 locus (whose expression in logFC is between 0,04 and 0,49).
  - Sheet 15: Of the 104 known BPD-associated loci described in ref7-9, 43 gave rise to a number of circular transcripts. Although all of them are not significant, five of these circRNAs exhibit high logFC (green = up; red = down).
  - Sheet 16: There is no significant difference in expression for linear NEBL and EPHA3 transcripts between the controls and BPD brains.

Supplemental Methods: This contains supplemental methods.

Supplemental Results: Overview of supplemental figures pertaining to results section.

## References

- 1 Global Burden of Disease Study, C. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **386**, 743-800, doi:10.1016/S0140-6736(15)60692-4 (2015).
- 2 Van Heeringen, C. & Marusic, A. Understanding the suicidal brain. *Br J Psychiatry* **183**, 282-284 (2003).
- 3 Nock, M. K., Hwang, I., Sampson, N. A. & Kessler, R. C. Mental disorders, comorbidity and suicidal behavior: results from the National Comorbidity Survey Replication. *Mol Psychiatry* **15**, 868-876, doi:10.1038/mp.2009.29 (2010).
- 4 Craddock, N. & Sklar, P. Genetics of bipolar disorder. *Lancet* **381**, 1654-1662, doi:10.1016/S0140-6736(13)60855-7 (2013).
- 5 Hou, L. *et al.* Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. *Hum Mol Genet*, doi:10.1093/hmg/ddw181 (2016).
- 6 Psychiatric, G. C. B. D. W. G. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* **43**, 977-983, doi:10.1038/ng.943 (2011).
- 7 Kataoka, M. *et al.* Exome sequencing for bipolar disorder points to roles of de novo loss-of-function and protein-altering mutations. *Mol Psychiatry* **21**, 885-893, doi:10.1038/mp.2016.69 (2016).
- 8 Cruceanu, C. *et al.* Family-based exome-sequencing approach identifies rare susceptibility variants for lithium-responsive bipolar disorder. *Genome* **56**, 634-640, doi:10.1139/gen-2013-0081 (2013).
- 9 Goes, F. S. *et al.* Exome Sequencing of Familial Bipolar Disorder. *JAMA Psychiatry* **73**, 590-597, doi:10.1001/jamapsychiatry.2016.0251 (2016).
- 10 Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427, doi:10.1038/nature13595 (2014).
- 11 Middleton, F. A. *et al.* Gene expression analysis of peripheral blood leukocytes from discordant sib-pairs with schizophrenia and bipolar disorder reveals points of convergence between genetic and functional genomic approaches. *Am J Med Genet B Neuropsychiatr Genet* **136B**, 12-25, doi:10.1002/ajmg.b.30171 (2005).
- 12 Akula, N. *et al.* RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. *Mol Psychiatry* **19**, 1179-1185, doi:10.1038/mp.2013.170 (2014).
- 13 Kohen, R., Dobra, A., Tracy, J. H. & Haugen, E. Transcriptome profiling of human hippocampus dentate gyrus granule cells in mental illness. *Transl Psychiatry* **4**, e366, doi:10.1038/tp.2014.9 (2014).
- 14 Cruceanu, C. *et al.* Transcriptome sequencing of the anterior cingulate in bipolar disorder: dysregulation of G protein-coupled receptors. *Am J Psychiatry* **172**, 1131-1140, doi:10.1176/appi.ajp.2015.14101279 (2015).
- 15 Kim, S., Hwang, Y., Webster, M. J. & Lee, D. Differential activation of immune/inflammatory response-related co-expression modules in the

hippocampus across the major psychiatric disorders. *Mol Psychiatry* **21**, 376-385, doi:10.1038/mp.2015.79 (2016).

16 Pacifico, R. & Davis, R. L. Transcriptome sequencing implicates dorsal striatum-specific gene network, immune response and energy metabolism pathways in bipolar disorder. *Mol Psychiatry*, doi:10.1038/mp.2016.94 (2016).

17 Hu, J. *et al.* Systematically characterizing dysfunctional long intergenic non-coding RNAs in multiple brain regions of major psychosis. *Oncotarget*, doi:10.18632/oncotarget.12122 (2016).

18 Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136**, 629-641, doi:10.1016/j.cell.2009.02.006 (2009).

19 Nagano, T. & Fraser, P. No-nonsense functions for long noncoding RNAs. *Cell* **145**, 178-181, doi:10.1016/j.cell.2011.03.014 (2011).

20 Brettschneider, J. *et al.* Stages of pTDP-43 pathology in amyotrophic lateral sclerosis. *Ann Neurol* **74**, 20-38, doi:10.1002/ana.23937 (2013).

21 Ziats, M. N. & Rennert, O. M. Aberrant expression of long noncoding RNAs in autistic brain. *J Mol Neurosci* **49**, 589-593, doi:10.1007/s12031-012-9880-8 (2013).

22 Faghihi, M. A. *et al.* Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med* **14**, 723-730, doi:10.1038/nm1784 (2008).

23 Nishimoto, Y. *et al.* The long non-coding RNA nuclear-enriched abundant transcript 1\_2 induces paraspeckle formation in the motor neuron during the early phase of amyotrophic lateral sclerosis. *Mol Brain* **6**, 31, doi:10.1186/1756-6606-6-31 (2013).

24 Soreq, L. *et al.* Long non-coding RNA and alternative splicing modulations in Parkinson's leukocytes identified by RNA sequencing. *PLoS Comput Biol* **10**, e1003517, doi:10.1371/journal.pcbi.1003517 (2014).

25 Ganzola, R. & Duchesne, S. Voxel-based morphometry meta-analysis of gray and white matter finds significant areas of differences in bipolar patients from healthy controls. *Bipolar Disord* **19**, 74-83, doi:10.1111/bdi.12488 (2017).

26 Vargas, C., Lopez-Jaramillo, C. & Vieta, E. A systematic literature review of resting state network--functional MRI in bipolar disorder. *J Affect Disord* **150**, 727-735, doi:10.1016/j.jad.2013.05.083 (2013).

27 Ozerdem, A., Ceylan, D. & Can, G. Neurobiology of Risk for Bipolar Disorder. *Curr Treat Options Psychiatry* **3**, 315-329, doi:10.1007/s40501-016-0093-6 (2016).

28 Ashwal-Fluss, R. *et al.* circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* **56**, 55-66, doi:10.1016/j.molcel.2014.08.019 (2014).

29 Zhang, X. O. *et al.* Complementary sequence-mediated exon circularization. *Cell* **159**, 134-147, doi:10.1016/j.cell.2014.09.001 (2014).

30 Starke, S. *et al.* Exon circularization requires canonical splice signals. *Cell Rep* **10**, 103-111, doi:10.1016/j.celrep.2014.12.002 (2015).

31 Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384-388, doi:10.1038/nature11993 (2013).

32 Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333-338, doi:10.1038/nature11928 (2013).

33 Lasda, E. & Parker, R. Circular RNAs: diversity of form and function. *RNA* **20**, 1829-1842, doi:10.1261/rna.047126.114 (2014).

- 34 Mercer, T. R. *et al.* Long noncoding RNAs in neuronal-glia fate specification and oligodendrocyte lineage maturation. *BMC Neurosci* **11**, 14, doi:10.1186/1471-2202-11-14 (2010).
- 35 Ramos, A. D. *et al.* Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny in vivo. *Cell Stem Cell* **12**, 616-628, doi:10.1016/j.stem.2013.03.003 (2013).
- 36 Liu, S. J. *et al.* Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol* **17**, 67, doi:10.1186/s13059-016-0932-1 (2016).
- 37 Rybak-Wolf, A. *et al.* Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol Cell* **58**, 870-885, doi:10.1016/j.molcel.2015.03.027 (2015).
- 38 Veno, M. T. *et al.* Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. *Genome Biol* **16**, 245, doi:10.1186/s13059-015-0801-3 (2015).
- 39 You, X. *et al.* Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat Neurosci* **18**, 603-610, doi:10.1038/nn.3975 (2015).
- 40 Chen, W. & Schuman, E. Circular RNAs in Brain and Other Tissues: A Functional Enigma. *Trends Neurosci* **39**, 597-604, doi:10.1016/j.tins.2016.06.006 (2016).
- 41 Chen, C. H., Suckling, J., Lennox, B. R., Ooi, C. & Bullmore, E. T. A quantitative meta-analysis of fMRI studies in bipolar disorder. *Bipolar Disord* **13**, 1-15, doi:10.1111/j.1399-5618.2011.00893.x (2011).
- 42 Delvecchio, G. *et al.* Common and distinct neural correlates of emotional processing in Bipolar Disorder and Major Depressive Disorder: a voxel-based meta-analysis of functional magnetic resonance imaging studies. *Eur Neuropsychopharmacol* **22**, 100-113, doi:10.1016/j.euroneuro.2011.07.003 (2012).
- 43 Dell'Osso, L., Del Grande, C., Gesi, C., Carmassi, C. & Musetti, L. A new look at an old drug: neuroprotective effects and therapeutic potentials of lithium salts. *Neuropsychiatr Dis Treat* **12**, 1687-1703, doi:10.2147/NDT.S106479 (2016).
- 44 Chiu, C. T., Wang, Z., Hunsberger, J. G. & Chuang, D. M. Therapeutic potential of mood stabilizers lithium and valproic acid: beyond bipolar disorder. *Pharmacol Rev* **65**, 105-142, doi:10.1124/pr.111.005512 (2013).
- 45 Tsai, L. K. *et al.* Mesenchymal stem cells primed with valproate and lithium robustly migrate to infarcted regions and facilitate recovery in a stroke model. *Stroke* **42**, 2932-2939, doi:10.1161/STROKEAHA.110.612788 (2011).
- 46 Ludwig, B. & Dwivedi, Y. Dissecting bipolar disorder complexity through epigenomic approach. *Mol Psychiatry* **21**, 1490-1498, doi:10.1038/mp.2016.123 (2016).
- 47 Wang, Y. *et al.* Circular RNAs in human cancer. *Mol Cancer* **16**, 25, doi:10.1186/s12943-017-0598-7 (2017).
- 48 Dines, M. & Lamprecht, R. The Role of Ephs and Ephrins in Memory Formation. *Int J Neuropsychopharmacol* **19**, doi:10.1093/ijnp/pyv106 (2016).
- 49 Dines, M. & Lamprecht, R. EphrinA4 mimetic peptide targeted to EphA binding site impairs the formation of long-term fear memory in lateral amygdala. *Transl Psychiatry* **4**, e450, doi:10.1038/tp.2014.76 (2014).



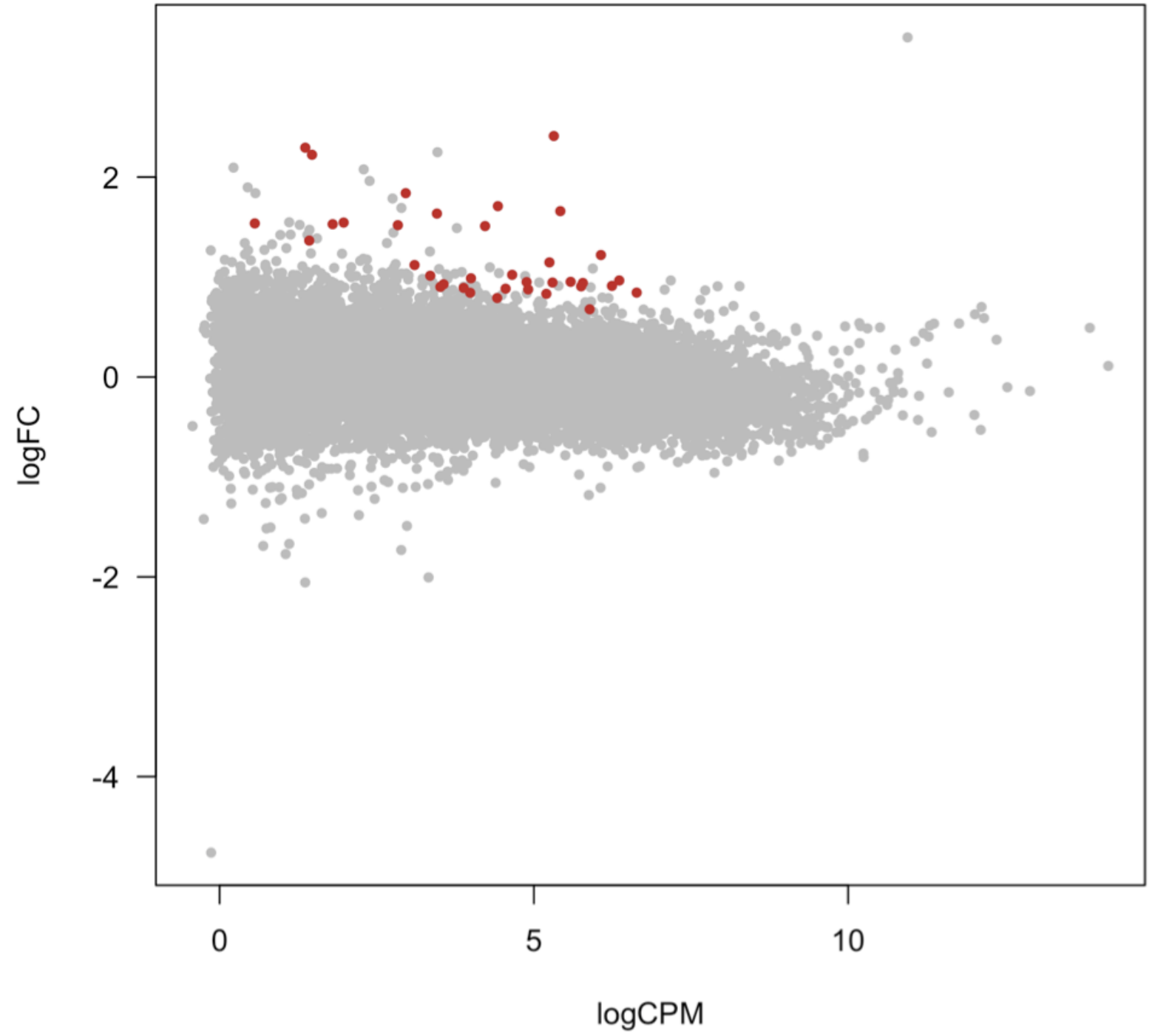
- 50 Attwood, B. K. *et al.* Neuropsin cleaves EphB2 in the amygdala to control anxiety. *Nature* **473**, 372-375, doi:10.1038/nature09938 (2011).
- 51 Boyd, A. W., Bartlett, P. F. & Lackmann, M. Therapeutic targeting of EPH receptors and their ligands. *Nat Rev Drug Discov* **13**, 39-62, doi:10.1038/nrd4175 (2014).
- 52 Lamminmaki, U., Nikolov, D. & Himanen, J. Eph Receptors as Drug Targets: Single-Chain Antibodies and Beyond. *Curr Drug Targets* **16**, 1021-1030 (2015).
- 53 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36, doi:10.1186/gb-2013-14-4-r36 (2013).
- 54 Zhao, Y., Yuan, J. & Chen, R. NONCODEv4: Annotation of Noncoding RNAs with Emphasis on Long Noncoding RNAs. *Methods Mol Biol* **1402**, 243-254, doi:10.1007/978-1-4939-3378-5\_19 (2016).
- 55 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
- 56 Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25, doi:10.1186/gb-2010-11-3-r25 (2010).
- 57 Zhang, Y. *et al.* An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci* **34**, 11929-11947, doi:10.1523/JNEUROSCI.1860-14.2014 (2014).
- 58 Fleige S, P. M. RNA integrity and the effect on the real-time qRT-PCR performance. *Mol Aspects Med.* (2006).
- 59 Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* **11**, R14, doi:10.1186/gb-2010-11-2-r14 (2010).
- 60 Chen, Y., Lun, A. T. & Smyth, G. K. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res* **5**, 1438, doi:10.12688/f1000research.8987.2 (2016).
- 61 Vitting-Seerup, K., Porse, B. T., Sandelin, A. & Waage, J. spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. *BMC Bioinformatics* **15**, 81, doi:10.1186/1471-2105-15-81 (2014).
- 62 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515, doi:10.1038/nbt.1621 (2010).
- 63 Falcon, S. & Gentleman, R. Using GStats to test gene lists for GO term association. *Bioinformatics* **23**, 257-258, doi:10.1093/bioinformatics/btl567 (2007).
- 64 Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800, doi:10.1371/journal.pone.0021800 (2011).
- 65 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).



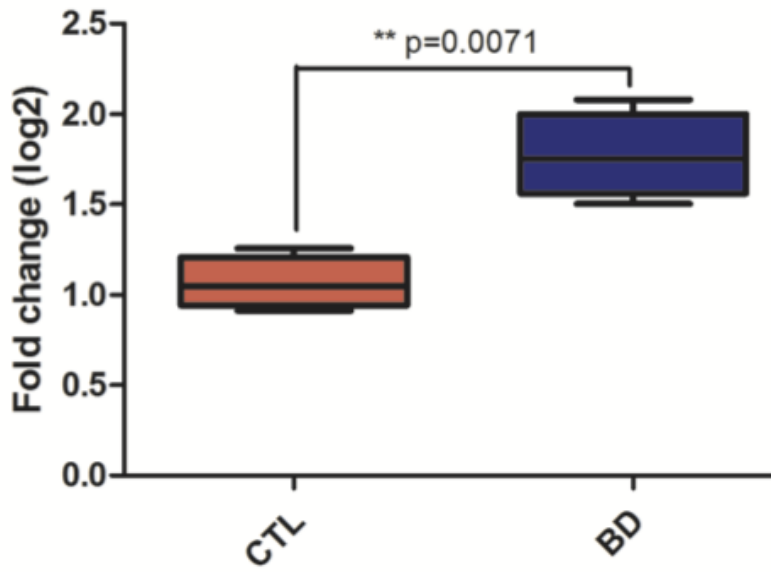


**Table 1.** Baseline characteristics of the study population. PMI = postmortem interval. M = male. F = female.

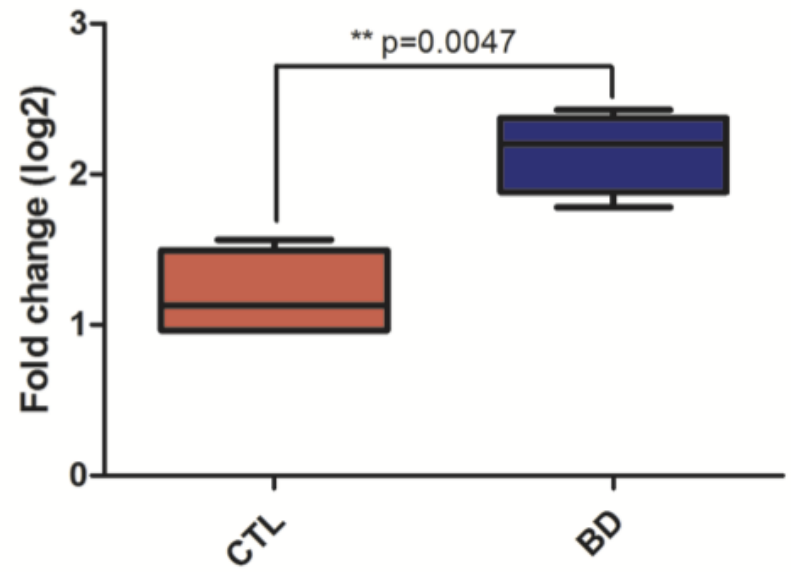
Cohort	Discovery (N=8)		Replication (N=8)	
	<i>Controls (N=4)</i>	<i>Bipolar disorder (N=4)</i>	<i>Controls (N=4)</i>	<i>Bipolar disorder (N=4)</i>
PMI	6h55	6h25	6h47	5h21
PMI range	6h15 - 7h15	4h50 - 8h00	5h05 - 8h10	4h35 - 6h40
pH	6.41	6.39	6.53	6.48
pH range	6.23 - 6.76	6.26 - 6.53	6.47 - 6.58	6.38 - 6.70
Sex	1F, 3M	1F, 3M	3F, 1M	4M
Mean age	65.7	72.5	70	73.5
Age range	49 - 88	70 - 79	57 - 93	68 - 81



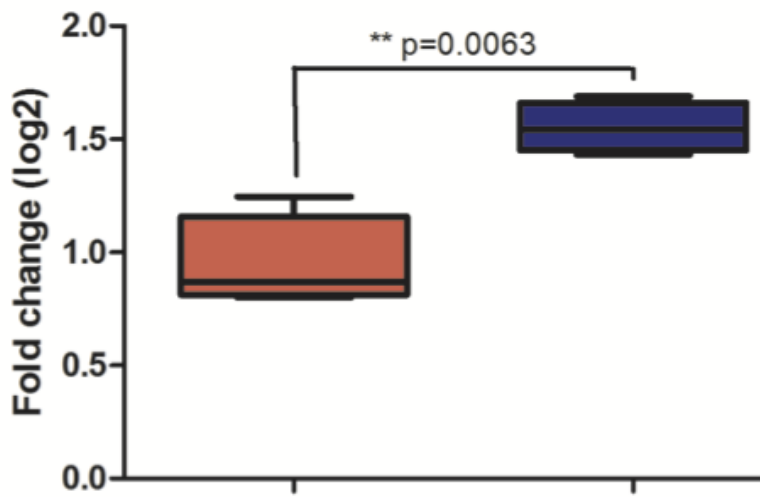
### ABCA1



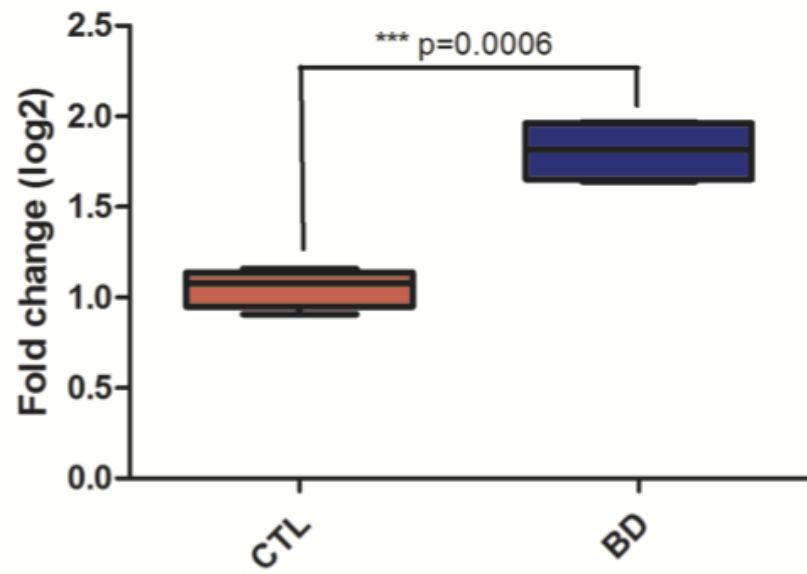
### BCL6B



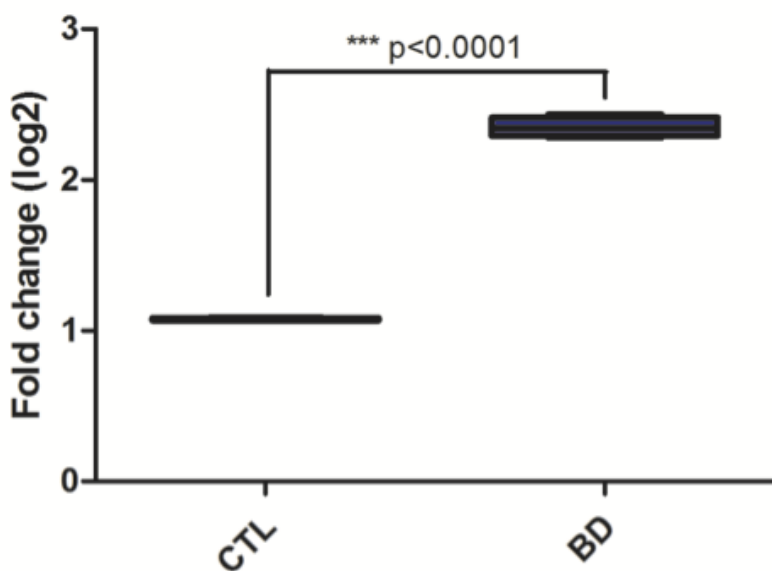
### CD93

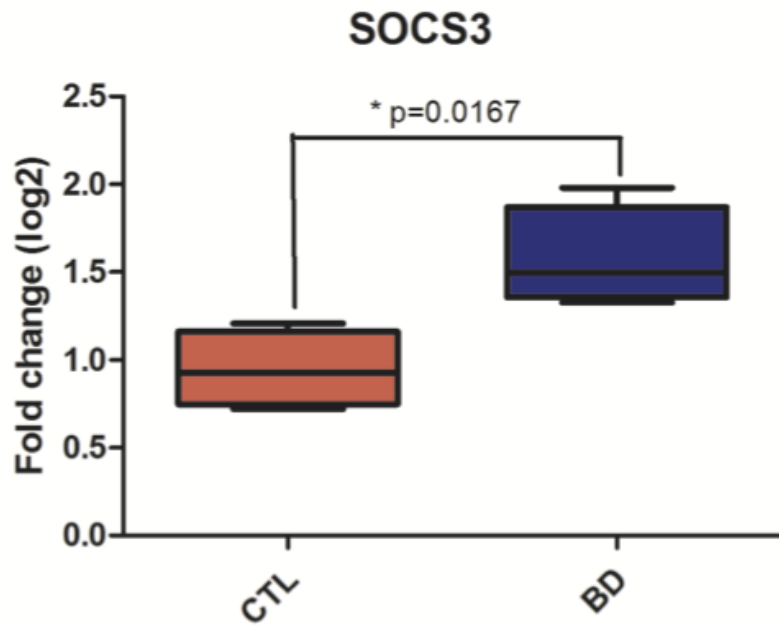
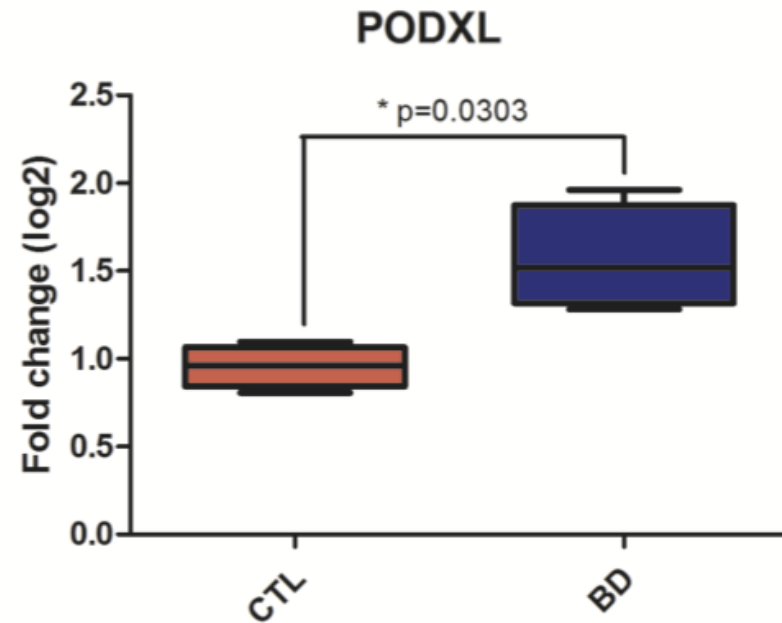
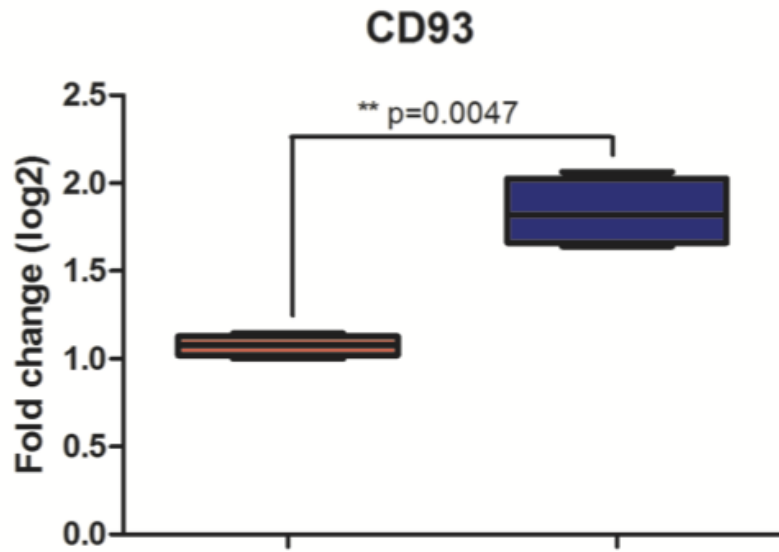
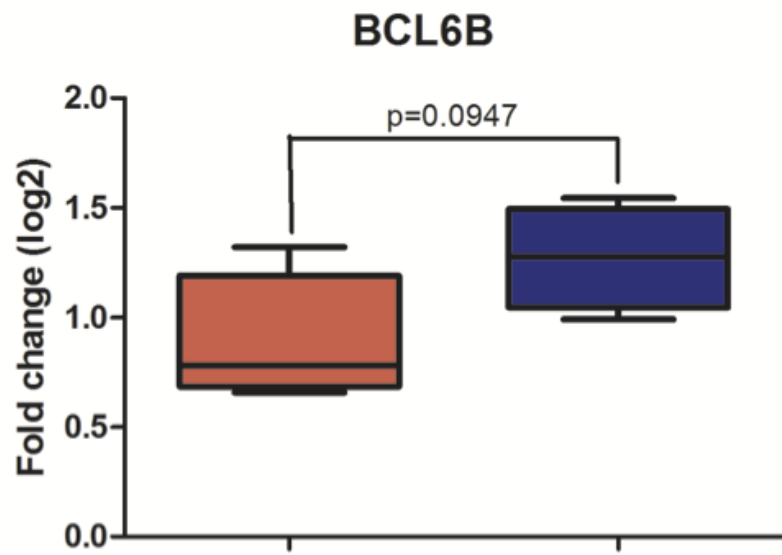
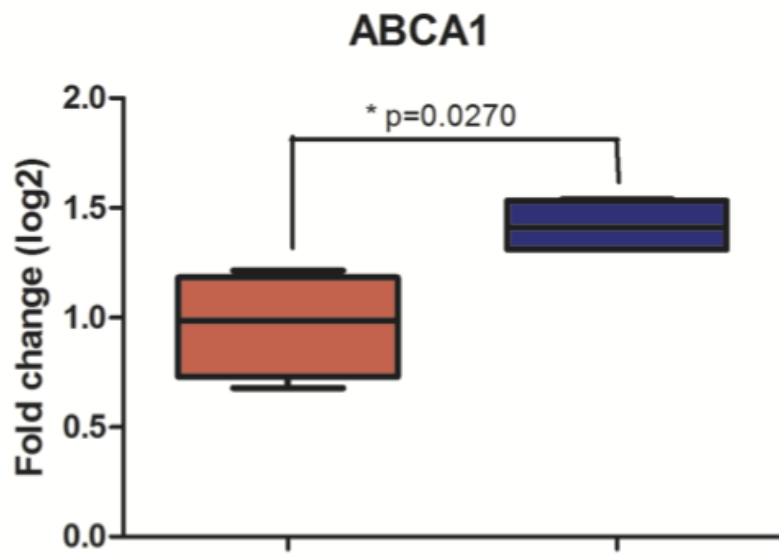


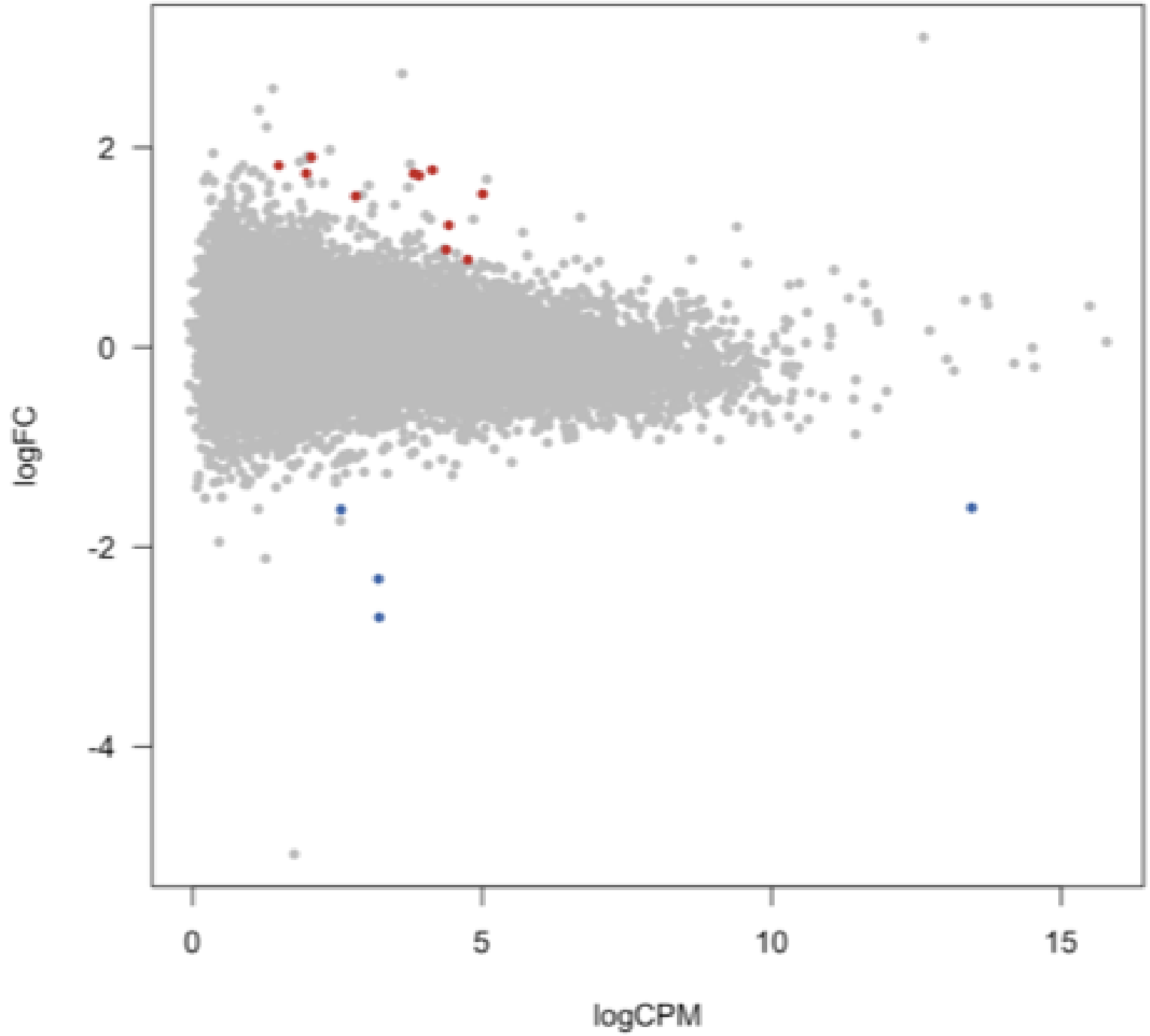
### PODXL

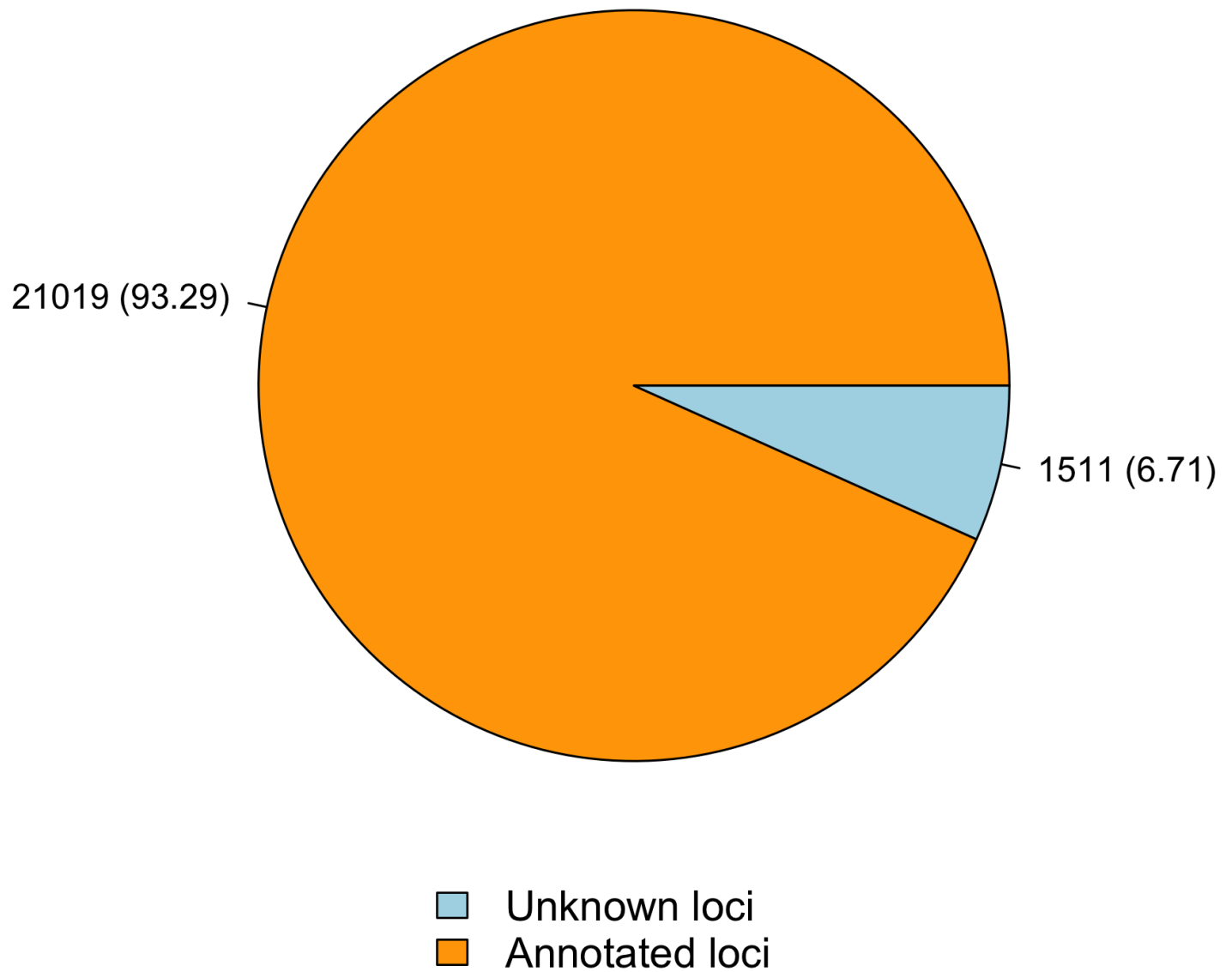


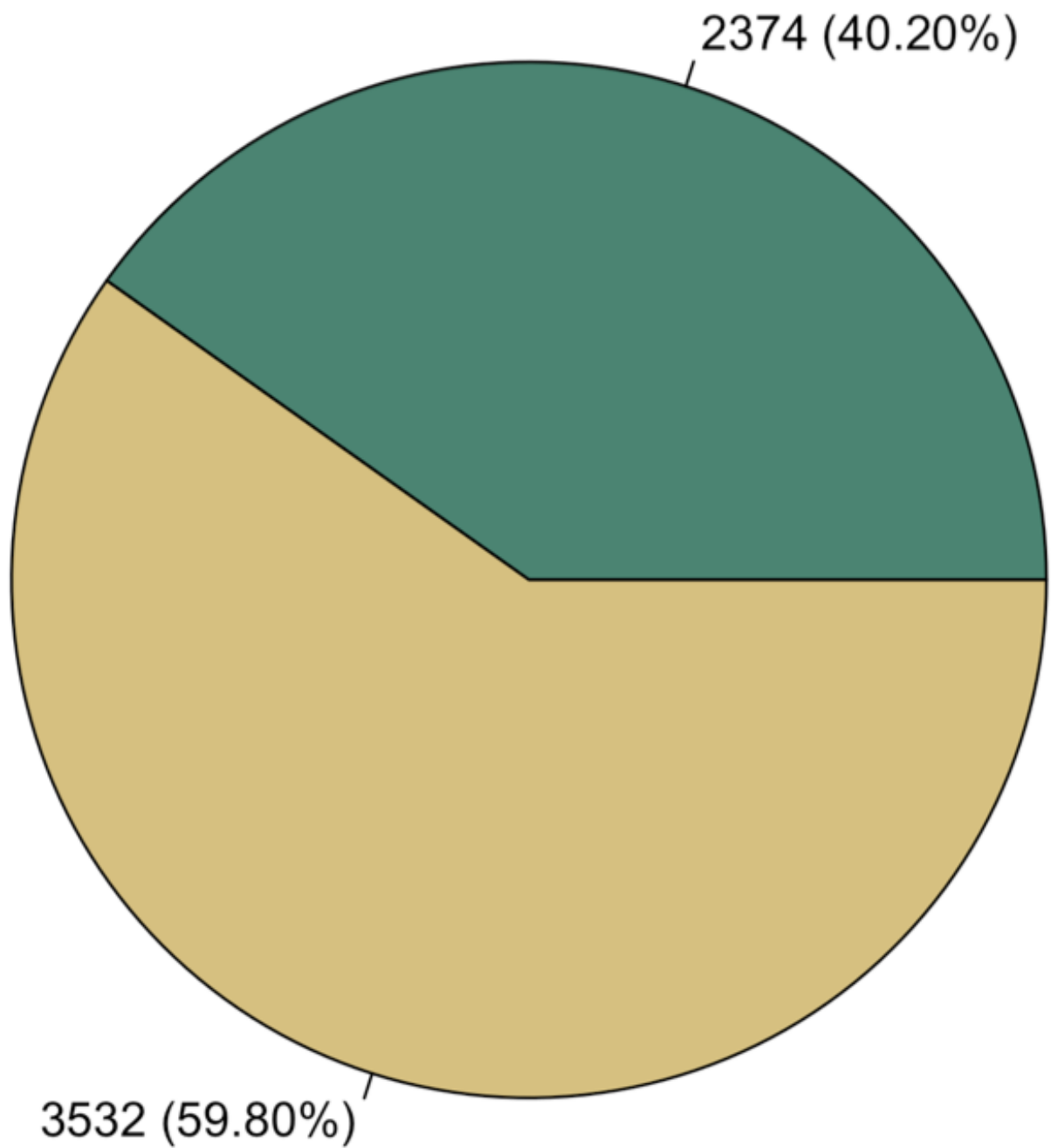
### SOCS3













-  Loci producing only 1 circRNA
-  Loci producing > 1 circRNA

