HHS Public Access

Author manuscript

Nat Neurosci. Author manuscript; available in PMC 2022 September 07.

Published in final edited form as:

Nat Neurosci. 2022 March ; 25(3): 381–389. doi:10.1038/s41593-022-01024-6.

Amygdala and anterior cingulate transcriptomes from individuals with bipolar disorder reveal down-regulated neuroimmune and synaptic pathways

Peter P. Zandi[1,2,+], **Andrew E. Jaffe**[1,2,3,4,5,6,7], **Fernando S. Goes**[1], **Emily E. Burke**[3], **Leonardo Collado-Torres**[3,4], **Louise Huuki-Myers**[3], **Arta Seyedian**[3], **Yian Lin**[1], **Fayaz Seifuddin**[8], **Mehdi Pirooznia**[8], **Christopher A. Ross**[1,7,9,10], **Joel E. Kleinman**[1,3], **Daniel R. Weinberger**[1,3,5,9], **Thomas M. Hyde**[1,3,9,+]

[1]Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD, USA

[2]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

[3]The Lieber Institute for Brain Development, Baltimore, MD, USA

[4]Center for Computational Biology, Johns Hopkins University, Baltimore, MD, USA

[5]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[6]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

[7]Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD, USA

[8]The National Heart, Lung and Blood Institute, The National Institute of Health, Bethesda, MD.

[9]Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD, USA

[+]Correspondence to: pzandi1@jhu.edu or thomas.hyde@libd.org.

Author Contributions

P.P.Z. contributed to the study design, data analysis, and data interpretation, and he led writing the manuscript. A.E.J. contributed to the study design, data processing, data analysis, and writing the manuscript. L.T.C. contributed to data processing, data analysis and writing the manuscript. F.S. and M.P. contributed to data analysis and data interpretation. E.E.B., L.H., A.S., and Y.L. contributed to data analysis. C.A.R. contributed to data interpretation and writing of the manuscript. T.M.H., J.E.K. contributed to the study design, data collection and generation, data interpretation, and writing the manuscript. D.R.W. and F.S.G. contributed to study design, data interpretation and writing the manuscript. All authors have approved the manuscript and agreed to take responsibility for their contribution to the work.

Conflict of Interest: None

Financial Disclosures: None

Competing Interests

A.E.J. is now a full-time employee and shareholder of Neumora Therapeutics, a for-profit biotechnology company, which is unrelated to the contents of this manuscript.

D.R.W. is on the Scientific Advisory Board of Sage Therapeutics.

All other authors report no competing interests.

Code Availability.

Code for all analyses described here is available on GitHub: https://github.com/LieberInstitute/zandiHyde_bipolar_rnaseq. The code has also been submitted to Zenodo at: https://doi.org/10.5281/zenodo.4463984.

[10]Department of Pharmacology and Molecular Sciences, Johns Hopkins School of Medicine, Baltimore, MD, USA

Abstract

Recent genetic studies have identified variants associated with bipolar disorder (BD), but it remains unclear how brain gene expression is altered in BD, and how genetic risk for BD may contribute to these alterations. Here, we obtained transcriptomes from subgenual anterior cingulate cortex and amygdala samples from post-mortem brains of individuals with BD and neurotypical controls, including 511 total samples from 295 unique donors. We examined differential gene expression between cases and controls, and the transcriptional effects of BD-associated genetic variants. We found two co-expressed modules that were associated with transcriptional changes in BD: one enriched for immune and inflammatory genes and the other with genes related to the post-synaptic membrane. Over 50% of BD genome-wide significant loci contained significant expression quantitative trait loci (eQTLs), and these data converged on several individual genes, including *SCN2A* and *GRIN2A*. Thus, these data implicate specific genes and pathways that may contribute to the pathology of bipolar disorder.

Introduction

Bipolar disorder (BD) is a highly heritable serious mental illness, with estimates reaching as high 80%[1]. Genome-wide association studies (GWAS) have attained the critical sample size needed to begin identifying specific genetic variants that contribute to BD heritability. A recent effort from the Psychiatric Genomics Consortium (PGC) reported 31 genome-wide significant loci and 850 additional suggestive loci[2]. As in GWAS of other major psychiatric disorders[3], the identified variants are common (minor allele frequencies [MAF] > 10%) and have small effect sizes (odds ratios [OR] < 1.1). The implicated variants combined explain only a small proportion of the genetic contribution to BD and the mechanisms by which they increase risk are unclear. Thus, it is critical to translate genetic associations into biological mechanisms to lay the foundation for developing new and more effective treatments.

Findings from GWAS of major psychiatric disorders are enriched in regulatory regions of the genome[4], suggesting that transcriptional mechanisms play an important role in their etiology. We therefore investigated the transcriptional changes and mechanisms that contribute to BD. As gene expression is differentially regulated across tissue types[5], we examined transcriptional mechanisms in the etiologically relevant tissue using post-mortem brain samples. Furthermore, because the brain is a complex organ composed of multiple distinct anatomical and functional regions with gene expression patterns that vary by brain region[6,7], we focused on samples from brain regions that may be most relevant to BD pathology. These included two key regions of the limbic system, the amygdala and the subgenual anterior cingulate cortex (sACC), which multiple converging lines of evidence[8,9] suggest underlie mood regulation and BD pathophysiology.

We report results of RNA sequencing (RNA-seq) data from one of the largest single post-mortem brain sample collections for BD and neurotypical subjects, involving 511 total samples across 295 unique donors, all of whom were of European ancestry. We tested

for transcriptional differences between BD cases versus unaffected controls and further examined the transcriptional effects of genetic variants identified by recent GWAS of BD (PGC-BD2)[2]. The findings converge on biological processes linked with synaptic signaling and suggest specific expression and splicing mechanisms by which certain genetic variants contribute to risk for BD. These findings provide an evidentiary base to motivate follow-up molecular work to elucidate the specific mechanisms that explain the genetic contribution to BD.

Results

A total of 511 samples from 295 individuals (138 cases and 157 controls) were available for study across the two brain regions, including 268 samples from the sACC (126 cases versus 142 controls) and 243 samples from the amygdala (121 cases versus 122 controls). The samples were 29% female with a mean age of 45.7 (SD=15.9, min=13.0, max=86.7) at death. The skew towards males in these samples is generally consistent with the sex ratio of the overall post-mortem brain collection in the Lieber Institute for Brain Development Human Brain Repository. See Supplementary Table 1 for further description of the study samples.

Differential Expression.

We first tested for differential expression between BD cases and neurotypical controls at the gene, transcript, exon, and junction expressed feature levels in each of the two brain regions. Results with all features tested are displayed as volcano plots in Supplementary Figure 1, and details of the differentially expressed features at FDR<5% are provided in Supplementary Table 2. There were 962 unique genes with at least one significant (FDR<5%) differentially expressed feature (Figure 1), with the overwhelming majority of implicated genes identified in the sACC compared to amygdala (858 vs 145). Most (86.8%) were protein-coding genes, while another 6% were long non-coding RNAs (lncRNA) and the remainder were other classes of non-coding sequence. Considerably more differentially expressed gene features (n=738) were identified compared to transcripts (n=101) in both brain regions, and the effect sizes were in general larger for genes (mean $|\log_2 FC|$ 0.18 vs 0.14 in sACC, P=0.01, K-S Test), in contrast to previous findings[10]. Furthermore, of the 962 unique genes identified as having DE features, 177 (18.4%) were implicated by a significant association with an exon or junction feature and would have been missed if only gene or transcript features were tested. Although only a small minority of genes (n=41; 4.3%) were implicated in both the sACC and amygdala, there was little evidence of significant differences in BD-related differential expression between the two brain regions (Supplementary Figure 2). In tests of interaction between diagnosis and brain region across all four feature types in region-combined analyses (see Methods), only one was FDR<5% significant among genes that showed any evidence of BD-related differential expression in either of the brain regions.

WGCNA Results.

To help resolve the biological implications of these findings, we carried out a weighted gene co-expression network analysis (WGCNA) with gene level data from cases and

controls and both brain regions together and tested if modules of co-expressed genes were associated with BD (Supplementary Table 3). A total of 20 modules were identified. The eigengene of the red module, consisting of 579 genes, was significantly associated with BD diagnosis after correction for multiple testing, with a trend towards down-regulation of module genes (β=−0.021, SE=0.006, P=2.76×10^{-4}) (Figure 2a). In addition, this module was enriched for genes that were differentially expressed with BD at FDR<5% (n=91, OR=5.08, P=8.55×10^{-51}), and moreover the differentially expressed genes tended to be concentrated among the more highly connected genes (i.e. putative hub genes) of this module (P=8.15×10^{-6}, Wilcoxon rank sum tests on module connectivity metric K_{within}). This module was also characterized by containing microglia-specific genes (P=7.60×10^{-39}, Fisher's Exact Test), which was consistent with an enrichment of microglia-specific genes observed among all differentially expressed genes that were down-regulated in the sACC (P=9.36×10^{-8}, Fisher's Exact Test) and amygdala (P=1.0×10^{-3}, Fisher's Exact Test). No other cell-type specific genes were enriched in the differentially expressed genes or in this module after correcting for multiple testing. This module was also enriched in numerous immune and inflammatory related GO pathways. Details of all enriched GO pathways at FDR<5% are shown in Supplementary Table 4. The 15 most significantly enriched of these GO pathways that contain genes with differentially expressed features are shown in Figure 2b. There were 31 unique genes with differentially expressed features among these selected pathways (Figure 2c). Of note, four of these genes (*A2M*, *TREM2*, *CTSB*, and *PLCG2*) have been associated with Alzheimer's disease. Another eight of these genes (*CD4*, *IKBKB*, *CX3CR1*, *PYCARD*, *NFKBID*, *SYK*, *BLNK*, and *TNFA1P8L2*), or over 25% of the differentially expressed genes in the top pathways, play a role directly or indirectly in NF-kB (nuclear factor kappa light chain enhancer of activated B cells) activation. In fact, five of these genes converge in the "regulation of I-kappaB kinase/NF-kappaB signalling" GO pathway, which was also significantly enriched in the red module (GO:0007249; Q=2.46×10^{-7}, Hypergeometric test). See Supplementary Table 2 for details of the differential expression results with these individual genes.

The second most significant module associated with BD was the pink module, which consisted of 224 genes that also tended to be down-regulated with BD (β=−0.014, SE=0.005, P=6.126×10^{-3}) (Figure 2d). The pink module did not survive correction for multiple testing, but it was also significantly enriched for individual genes that were differentially expressed with BD (n=17, OR=2.09, P=5.48×10^{-3}). This module was enriched in numerous overlapping synaptic related GO pathways. Details of all enriched GO pathways at FDR<5% are shown in Supplementary Table 4. The 15 most significantly enriched of these GO pathways that contain genes with differentially expressed features are shown in Figure 2e. The most significant was the "synaptic membrane" pathway (GO:0097060; Q=2.86×10^{-10}, Hypergeometric test), and it contained one of the most significant differentially expressed module genes, *GRASP*, which interestingly plays an important role in the organization of group 1 metabotropic glutamate receptors (mGluRs) at synapses. Other module genes in these pathways with significant differentially expressed features included *PTPRG*, *CBARP*, *ACTN1*, *TRAF3*, and *LINGO1* (highlighted in Figure 2e). See Supplementary Table 2 for details of differential expression results with these individual genes.

QTL Results.

Transcriptional differences between BD cases and controls could relate to some combination of the causes and/or consequences of illness. Therefore, to further interrogate the transcriptional mechanisms that may contribute to BD risk, we next tested for *cis* eQTLs or sQTLs in genetic loci implicated by prior BD GWAS. In particular, we tested for QTL associations between 10,777 SNPs and gene (n=4,647), transcript (n=14,434), exon (n=76,589) and junction (n=49,188) features related to 5,019 unique genes residing in the neighborhoods (as defined in the Methods section) of 881 loci (GWAS $p<1\times10^{-4}$ before replication) implicated by the PGC GWAS of BD[2]. The number of unique genes with features tested is more than the number of genes tested because some of the features are in genes with insufficient information to test the entire gene. For these analyses we combined the data from both cases and controls. Figure 3 shows a summary of the significant *cis* eQTLs/sQTLs (FDR<1%) with the lead SNP for these loci across the two brain regions. See Supplementary Table 5 for full results. Significant *cis* e/sQTLs were identified in at least one of the two brain regions in 172 (33.2%) of the 881 GWAS loci (518 with evaluable data), including 16 (51.6%) of the genome-wide significant and 156 (32.0%) of the suggestive loci. The greater proportion of significant e/sQTLs among the genome-wide significant loci is consistent with the notion that they are more enriched for true positive associations. Similar to the differential expression analysis, a greater number of significant QTLs was observed in the sACC (1,277 different features in 305 unique genes) compared to the amygdala (736 different features in 216 unique genes), but the findings were largely overlapping across the two brain regions. In tests of interaction between brain regions and the most significant lead SNP-feature at each locus, only five were observed to be significant after correction for multiple testing and each of these were quantitative interactions that showed the same trends across brain regions. Approximately 82% of the genes with at least one significant QTL across both brain regions were protein-coding, while another 5% were lncRNAs, in line with the differential expression analyses described above.

After carrying out a conditional analysis to isolate independent QTL effects[11], we identified 28 genes with significant conditionally independent QTLs in the 16 PGC-BD2 genome-wide significant loci and 218 genes in the 156 suggestive loci. In 9 of the 16 genome-wide significant loci (56.3%) and 118 of the 156 suggestive loci (75.6%), this QTL evidence pointed to at most a single gene when considering both brain regions, helping to narrow the identification of the genetic risk-associated genes in these loci and clarify the potential transcriptional mechanisms involved in BD risk. Among these 127 implicated genes, 15 (11.8%) involved an alternative splicing mechanism through an sQTL, 19 (15.0%) involved an eQTL associated with overall gene expression, 33 (26.0%) involved an eQTL with a specific transcript, and the remaining 60 (47.2%) involved a more ambiguous transcriptional mechanism that was identified solely through an altered exon or junction.

Table 1 shows the results for the individual genes implicated in 9 of the 16 genome-wide significant loci. Interestingly, two of these genes (*SCN2A* and *GRIN2A*) are ion channel components of the synaptic membrane pathway (GO:0097060) implicated in the WGCNA analysis described above and are predominantly expressed in neuronal cells as suggested by data from single nucleus RNA sequencing on 5 brain regions from 8 neurotypical

controls (Supplementary Figure 3). The most significant QTLs were observed in *SCN2A*. There are currently 18 annotated transcript isoforms of the gene, including 9 that are protein-coding (Figure 4a). Six of the protein-coding transcripts are translated into a canonical protein (UniProtKB Q99520) which is 2,005 amino acids long. The alternate allele of the lead SNP in this locus (rs17183814:G:A), which is associated with a decreased risk of BD, is a mis-sense coding variant (R19K0 that results in the substitution of lysine for arginine at position 19 of the canonical protein). In addition, it is a splicing variant that disrupts the canonical splice acceptor site in intron 1 of an alternative protein-coding transcript ENST00000636985, which encodes a truncated form of the protein that is only 1873 amino acid long ($\beta=-1.75$, $p=5.47\times10^{-22}$). This results in reduced expression of the junction spanning exons 1 (ENSE00003800530) and 2 (ENSE00003798965) ($\beta=-1.12$, $p=5.08\times10^{-29}$), which in turn leads to lower expression of the transcript ($\beta=-0.20$, $p=3.23\times10^{-14}$).

The other gene, *GRIN2A*, has 15 transcript variants in the genome databases, including 7 that are protein-coding (Figure 4b). The alternate allele of the lead SNP in this locus (rs11647445:T:G) has been associated with an increased risk of BD, and in our data it is associated with a decreased expression of exon 11 (ENSE00001327113) of the canonical transcript ENST00000396573 ($\beta=-0.07$, $p=1.32\times10^{-5}$). Interestingly, the lead SNP, which is approximately 10kb away from the affected exon, is not significantly associated with any other features in the region.

TWAS Results.

Lastly, we carried out a transcriptome-wide association study (TWAS) analysis to evaluate whether imputed gene expression levels were associated with BD allowing for the potential of multiple independent SNP-QTLs influencing overall gene transcription. A total of 13,822 genes passed our SNP-based heritability filter and were tested by TWAS in at least one of the two brain regions (see Supplementary Table 6). We identified 247 genes (n=155 in the sACC and n=125 in the amygdala) that were significantly associated with BD by TWAS at FDR<5% (Figure 5). Approximately 70% of these were protein coding and another 4.5% were lncRNAs, similar to the analyses reported above. Interestingly, there was little concordance between the TWAS and differential expression results (Supplementary Figure 4). By contrast, over 75% of these genes (n=192) were identified either in the PGC-BD2 genome-wide significant (n=53) or suggestive (n=139) loci, and the identified genes in these loci tended to implicate the same ones as those suggested by the QTL analysis above. Almost 40% of them also had a significant SNP-QTL, and in the 9 PGC-BD2 genome-wide significant loci where the QTL evidence pointed to a single gene 4 were also implicated by TWAS (*ASB16-AS1*, *ZCCHC2*, *LRRC57*, and *TRANK1*). Interestingly, 55 genes (n=31 in the sACC and n=27), almost 25% of those identified by TWAS, were outside of the PGC-BD2 significant or suggestive loci and represent novel loci of interest not previously implicated by the PGC-BD2 findings. See Supplementary Table 7 for a summary overview of the significant findings across the different analyses (differential expression, WGNCA, e/sQTL, and TWAS) carried out with the data.

Discussion

We used complementary analytic approaches to investigate transcriptional mechanisms underlying BD risk in two brain regions of the limbic system, the subgenual anterior cingulate cortex (sACC) and amygdala, thought to subserve mood dysregulation in BD[8,9]. Using WCGNA, we found two co-expressed modules downregulated in BD, one enriched with genes involved in immune and inflammatory responses and the other with genes related to the synaptic membrane. We further found that approximately 34% of loci reported by PGC-BD2, including over 50% of genome-wide significant loci, contained significant QTLs, implicating specific transcriptional mechanisms that potentially explain their contribution to risk. In 9 of the genome-wide significant loci, the QTL evidence converged on a single gene, including two of particular note (*SCN2A* and *GRIN2A*) that are membrane proteins involved in neuronal signaling. Associated alleles for these loci predicted reduced expression of these genes. The current findings thus implicate specific gene pathways in BD pathology and provide clues to transcriptional mechanisms that underlie risk and motivate further functional follow-up. They also provide circumstantial evidence of the involvement of specific cell-types in BD that will require further examination with approaches such as single nucleus RNA sequencing.

There have been at least 10 reported RNA sequencing studies of transcriptional changes in BD using post-mortem brain samples[10,12–20], and all but one had samples considerably smaller than the current study. The one exception was from the PsychEncode project which focused on the dorso-lateral prefrontal cortex (dlPFC). The findings from these studies have been largely inconsistent, possibly due to different brain regions studied, different experimental methods (e.g, polyA enrichment versus Ribo-Zero depletion), different analytic approaches (in particular qSVA used herein), and significant lack of power for most studies. As a result of our relatively larger sample size, we observed considerably more significant differentially expressed genes than previous studies. Interestingly, although there was little evidence of differences in expression changes with BD across brain regions, we observed almost eight times more differentially expressed gene features in the sACC versus amygdala. This difference cannot be explained solely by the slightly larger sample available for the sACC, nor by greater variability in feature expression levels in the sACC because the variability was comparable in both brain regions (Supplementary Figure 5). The magnitude of the findings in the sACC may reflect a more prominent role for this brain region both in mood regulation in general and BD specifically.

We identified two co-expressed gene modules associated with BD and enriched for differentially expressed genes. As mentioned above, one was enriched for a number of overlapping pathways traditionally involved in immune activation and inflammatory responses. This module, which was also enriched in microglia specific genes, was down-regulated in BD. Down-regulation of microglia related modules was also highlighted in the recent report from the PsychEncode[10], which included data on post-mortem brain samples from 28 overlapping BD cases with the current study. However, the current study included data on nearly four times as many new cases not previously studied from two different brain regions selected because of their putative involvement in mood disorders. Thus, the current findings provide additional support for the involvement of microglia in BD, although

the mechanism of this involvement is unclear. How to interpret the potential pathogenicity of down-regulation of the microglia related gene module, especially when contrasted with results from other studies that have reported elevation of both peripheral[21] and central markers of inflammation as well as putative markers of activated microglial cells in the brain in BD[22–25], is uncertain. We would not expect activated microglia to be represented by relative downregulation of their cognate genes.

It is noteworthy that several differentially expressed genes in the most enriched immune/inflammatory pathways of the microglia specific module have also been implicated in Alzheimer's Disease[26–29]. There is clinical and epidemiological evidence of a link between BD and Alzheimer's disease[30], and a recent re-analysis of GWAS data on BD and Alzheimer's disease found a significant overlap in polygenic risk between the two[31]. It has been hypothesized that common inflammatory processes underlying both disorders potentially explain the relationship between them[30]. Consistent with this, two differentially expressed genes, *TREM2* and *PLCG2*, may be of particular interest. Rare coding variants associated with Alzheimer's disease have recently been identified in both genes, and they were noted to participate in the same interaction network[29]. These genes are noteworthy because they are primarily expressed in microglia in the brain and contribute to microglia-mediated innate immunity and modulate inflammatory responses[29]. Moreover, both genes exert downstream effects on the NF-kB pathway, which was implicated by our analysis. *PLCG2* may be of additional interest because it encodes a protein that drives phosphoinositide recycling which has been proposed as a target of lithium's therapeutic mechanism of action in BD[32].

Including *TREM2* and *PLCG2*, over a quarter of the differentially expressed genes in the most enriched immune/inflammatory pathways of the microglia specific module converge on the NF-kB pathway. NF-kB is a transcription factor complex that regulates multiple aspects of innate and adaptive immune functions and mediates inflammatory responses[33]. The previous report from the PsychEncode project also highlighted NF-kB related modules in association with BD, as well as schizophrenia and autism[10]. Curiously, that study found an up-regulation of NF-kB related modules, whereas we observed primarily down-regulation of specific genes along the pathway. As a result, while the evidence suggests a possible role of the NF-kB pathway in BD pathology[34], more work is needed to confirm this and untangle the precise molecular mechanisms that may be involved.

In interpreting these results, it may be relevant that microglia play a role in neuronal circuit activity and synaptic function independent of their traditional classification as "immune cells" of the central nervous system. In the adult brain, microglia can influence neuronal activity acutely and over the long term, and appear to monitor the integrity of synaptic function[35]. It may be that downregulation of microglia in BD produces changes in both neuronal activity and synaptic function, rather than altering classic inflammatory processes associated with infection and autoimmune diseases. In line with this, the NF-kB pathway is also known to have pleiotropic effects on structural and synaptic plasticity in the brain that may play a role in cognitive behaviors, including learning and memory[36]. Because these effects appear to be highly cell dependent, it will be crucial to follow-up the current findings

with single cell RNA sequencing approaches to unravel the pathogenic role of disruption of this pathway in BD.

The other BD associated module identified from our study was enriched for a number of overlapping gene sets related to the synaptic membrane. Several of these pathways were reported in previous RNA sequencing studies of BD with post-mortem brain samples, lending credence to their role in BD. Of note, two postsynaptic pathways - GO:0014069 (postsynaptic density) and GO:0045211 (postsynaptic membrane) - were found to be enriched in a gene co-expression module associated with BD in a previous study of dlPFC samples from 11 cases versus 11 controls[16]. A similar postsynaptic density specific module was also associated with BD by the much larger study from PsychEncode[10]. In addition, the regulation of neuron projection development pathway (GO: 0010975) was reported to be enriched in one of two gene co-expression modules associated with BD in another study of samples from the dorsal striatum of 18 BD cases versus 17 controls[19]. The postsynaptic density pathway, in particular, has been the focus of increased interest in psychotic disorders. Several recent sequencing studies have implicated genes in these pathways in both schizophrenia[37] and now also in BD[38,39]. Our current findings add to this evidence. It is noteworthy, and perhaps not surprising, that across the various sequencing studies there has been less convergence of significant findings for specific genes than at the level of pathways. However, we anticipate that as sample sizes for both DNA and RNA sequencing studies increase there will be greater consistency of findings that point to specific genes driving BD pathology among the interacting network of neurotransmitter receptors and cytoskeletal and signaling proteins that spatially and functionally organize them at the postsynaptic membrane. In the meantime, it is tempting, to proffer that our results, both the evidence of downregulation of microglia related molecular activity and of synaptic signaling genes, converge on the hypothesis that synaptic signaling and plasticity are core genetic and molecular underpinnings of BD.

The QTL analyses may help in this regard as they provided further evidence for the contribution to BD risk of specific genes implicated by the latest GWAS. Using a conditional analysis approach, we identified 246 potentially relevant genes from the reported GWAS loci, including 28 genes in 16 PGC-BD2 genome-wide significant loci. In almost a quarter of the genome-wide significant loci, the evidence pointed to a gene that was involved in synaptic membrane pathways highlighted in the co-expression module described above.

Perhaps the most compelling individual finding was the eQTL association with *SCN2A*. This gene encodes the alpha subunit of a voltage-gated sodium ion channel ($Na_v1.2$), a transmembrane glycoprotein complex that is highly expressed in the brain, especially in the axon initial segment of excitatory glutamatergic neurons[7,40]. *De novo* and rare loss-of-function and missense coding variants in this gene have been associated with infantile epileptic encephalopathy (IEE), benign familial infantile seizures (BIS), and autism spectrum disorders and intellectual disabilities (ASD/ID)[41–43]. The association of *SCN2A* with three disorders of early brain development suggests that perturbations in this gene may produce dysmaturation of brain circuits resulting in life-long effects with the potential to profoundly alter adult behavior. Moreover, there is evidence suggesting an increased risk of BD in high-functioning individuals with ASD[44].

The canonical protein isoform of Nav1.2 is 2,005 amino acids long and coding mutations associated with IEE, BIS and ASD/ID have been identified along the entire length of the protein. The genetic variant associated with reduced risk of BD causes a coding change in six known transcripts that encode the canonical protein isoform leading to a lysine for arginine substitution at position 19, residing in the N-terminal cytoplasmic domain where other mis-sense mutations have been observed to predominantly cause ASD/ID[43]. While the R19K alteration substitutes one basic amino acid for another, it could have functional consequences, for instance via alterations of post-translational modifications[45]. Interestingly, the implicated variant in *SCN2A* also disrupts the splice acceptor site of an alternative transcript (ENST00000636985) which encodes an isoform of the protein that is only 1,873 amino acids long. This leads to a significant reduction in expression of the truncated transcript. It is unclear how the mis-sense coding change to the canonical isoform and/or reduction in expression of the truncated isoform reduce BD risk, but further investigation is warranted.

Another gene implicated by the eQTL analysis, *GRIN2A*, encodes a postsynaptic density protein (GluN2A) that makes up a subunit of the NMDA receptor, a glutamate-gated ion channel involved in synaptic plasticity, learning, and memory. Several lines of evidence suggest that dysregulation of glutamatergic transmission plays a role in a range of neuropsychiatric conditions[46], and may contribute to BD as well[47]. In the current analysis, we found that the identified risk allele for BD was associated with reduced expression of an exon that is constitutive to all of the protein's known full-length transcripts and is central to encoding the glutamate binding domain of the protein. More work is needed to determine the biological significance of this finding, but it is generally consistent with prior reports of reduced expression of *GRIN2A*[48] and reduced NMDA-mediated glutamatergic activity in general in BD[49]. It is noteworthy that we would not have observed any finding with *GRIN2A* if we had not analyzed exon features.

Human post-mortem brain samples are an increasingly important resource for studying neurogenetic mechanisms underlying risk for psychiatric disorders[10,50]. In the current analysis, we used one of the largest individual collections of post-mortem brain samples in BD that is also part of the PsychENCODE effort. By examining differential expression and QTL effects across multiple expression features in this large sample, we comprehensively surveyed the potential transcriptional changes in BD. The findings reported herein should help guide follow-up functional work to further elucidate the mechanisms by which implicated transcriptional changes in these genes contribute to risk for BD. Such work will advance our understanding of the etiology of BD and, hence, provide a stronger foundation for improved treatments.

Methods

Sample Collection

The post-mortem brain samples were supplied by the Lieber Institute for Brain Development Human Brain Repository, all collected with informed consent of next of kin. A number of brains in the LIBD Human Brain Repository were transferred from the National Institute of Mental Health Clinical Brain Disorders Branch under a material transfer agreement,

after having been collected under NIMH Protocol 90-M-0142 and processed approved by the NIMH/National Institutes of Health Institutional Review Board. Additional cases in the LIBD repository were collected from the Office of the Chief Medical Examiner for the State of Maryland under State of Maryland Department of Health and Mental Hygiene Protocol 12–24, and from the Kalamazoo County Michigan Medical Examiners' Office under Western Institutional Review Board Protocol 20111080. All samples were collected and processed using a protocol specifically developed to minimize sample heterogeneity and technical artifacts. In the current study, we used only samples from individuals of European ancestry to further minimize sample heterogeneity and maximize power of downstream analyses.

The procedures for sample collection, curation and diagnosis have been described in detail elsewhere[51]. Briefly, clinical and demographic information was gathered by the review of medical and psychiatric records[52] and via a structured interview[53–55] with next of kin conducted within 1 week of donation. Psychiatric narratives were prepared on each case, summarizing information obtained from all available sources, and each case was independently reviewed by two board-certified psychiatrists, who arrived at consensus DSM-IV Axis I lifetime diagnoses[56]. History of cigarette smoking at the time of death was collected during the initial telephone screening and verified through toxicological analysis of nicotine and cotinine levels in blood or brain tissue. Toxicological analyses of blood, vitreous humor fluid, occipital pole, and/or urine were conducted by a forensic toxicologist. Neuropathological examination was performed on each case by a board-certified neuropathologist. Subjects with evidence of clinically or neuropathologically significant cerebrovascular disease (infarcts or hemorrhages), subdural hematoma, neuritic pathology, or other significant pathological features were excluded from further study. Subjects with acute subarachnoid hemorrhages that were directly related to the immediate cause of death were not excluded. The cause and manner of death and any contributory causes were obtained from medical examiner documents. Neurotypical controls had no history of significant psychiatric symptoms or substance abuse, as determined by both telephone interviews with next of kin and medical examiner documentation. By definition, cases were excluded from the neurotypical control group if the manner of death was suicide, if death was due to drug overdose or poisoning, or toxicology results were positive. Agonal state was assessed by gathering data regarding specific medical conditions and treatment proximate to the date of death (e.g., coma, hypoxia, seizures) and the duration of the terminal phase. Age at death was verified by obtaining both date of birth and date of death through medical records, medical examiner documents, and family interviews. Postmortem interval (PMI) was defined as the time elapsed, in hours, between the pronounced time of death and time of tissue freezing. In the final sample, the manner of death among cases was 38.0% suicide, 21.2% natural 21.2% accidental, and 19.7% undetermined; while among controls it was 77.7% natural, 21.0% accidental, and 1.3% homicide.

Sample Preparation

After removal from the calverium, brains were wrapped in plastic and cooled on wet ice. A detailed macroscopic inspection was performed of the brain, meninges, attached blood vessels, and when possible, the pituitary and pineal glands. After removal of the brainstem

and cerebellum, the forebrains then were hemisected, cut into 1.5 cm coronal slabs, flash frozen in a pre-chilled dry-ice/isopentane slurry bath (−40°C), and stored at −80°C. The time from when the tissue was stored at −80°C until the RNA was extracted was considered the freezer time (mean ± SD: 43.8 ± 2.8 months). A block of lateral superior cerebellar cortex was cut transversely to the folia. pH was measured by inserting a probe into the right parietal neocortex and again into the right cerebellar hemisphere. The amygdala was identified on the frozen coronal slabs from the medial temporal lobe and dissected from one hemisphere with a hand held dental drill. The temporal lobe was sectioned in the coronal plane at a level designed to expose the amygdala at its largest circumference. The whole amygdala was then dissected out at the level of its largest circumference - including all subnuclei at this level - just anterior to the hippocampus, ventral to the anterior commissure and claustrum, and medial to the entorhinal cortex. Given the size of the dissected region it is unlikely there would be significant variation in the subnuclei included in the dissection between cases and controls that might lead to false positive associations in downstream analyses, although it is possible the combining of different subnuclei in bulk tissue may obscure true associations. The subgenual anterior cingulate cortex (sACC) also was dissected under visual guidance from the medial aspect of the forebrain at the level of the rostrum of the corpus callosum, using a hand held dental drill. It was dissected ventral to the corpus callosum, and dorsal to the orbital frontal cortex (BA11). Medially it is bounded by the interhemispheric fissure while laterally it is bounded by the corona radiata/centrum semiovale. Approximately 60% of both the amygdala and sACC samples were dissected from the left hemisphere and these were equally balanced between cases and controls. Moreover, all dissections were performed by one of us (TH), and they were performed blind to the case/control status of each brain, helping to minimize the potential for systematic differences that could confound downstream analyses.

RNA Sequencing

RNA was extracted from the above tissue dissections, and sequencing libraries constructed using Illumina TruSeq Stranded Total RNA Ribo-Zero sample Prep Kit following the manufacturer's protocol. Briefly, after ribosomal RNA depletion, we generated 200 base pair (bp) fragments, added paired end adapters, and added a unique nucleotide index/barcode to permit multiplexing. The sequencing libraries underwent quality control steps, and then were sequenced using an Illumina HiSeq 2000.

Reads were quality checked with FastQC[57], and where needed, leading bases were trimmed from the reads using Trimmomatic[58] as appropriate. Quality checked reads were mapped to the hg38/GRCh38 human reference genome with splice-aware aligner HISAT2 version 2.0.4[59]. We then generated estimates of expression at the gene, exon, junction, and transcript levels based on GENCODE release 25 (GRCh38.p7) annotation. Gene and exon expression levels were calculated with *featureCounts* v1.5.0-p3[60] which implements the popular HTseq[61] python-based approach in a fast and parallelizable framework. We also generated consensus junction locations, and the number of reads that support each junction in each sample, using regtools[62] v. 0.1.0 and the bed_to_juncs program from TopHat2[63] to retain the number of supporting reads (in addition to returning the coordinates of the spliced sequence, rather than the maximum fragment range). These junction reads have the

ability to approximate transcript-level expression without incurring the known inaccuracies in transcript assembly and abundance estimations[64]. Lastly, we used the program Salmon v0.7.2 to estimate transcript abundances[65]. Synthetic ERCC transcripts were quantified with Kallisto version 0.43.0[66].

Estimated feature expressions levels were normalized to reads per kilobase of feature per million mapped reads (RPKM) for genes and exons, reads per 10 million mapped reads (RP10M) for junctions and transcripts per million (TPM) for transcripts. Features were then filtered if they did not exceed the following mean expression thresholds: 0.25 RPKM for genes, 0.30 RPKM for exons, 0.35 RP10M for junctions and 0.40 TPM for transcripts as determined with the expression_cutoff() function from the jaffelab R package[67]. Cutoffs were applied based on samples from the amygdala and subgenual anterior cingulate cortex.

SNP Genotyping

Genotype data was processed and imputed as previously described for this brain collection[11]. Briefly, genotype imputation was performed on high-quality observed genotypes (removing low quality and rare variants) using the prephasing/imputation stepwise approach implemented in IMPUTE2[68] and Shape-IT2[69], with the imputation reference set from the full 1000 Human Genomes Project Phase 3 dataset[70] separately by Illumina platform using genome build hg19. We retained common variants (MAF > 5%) that were present in the majority of samples (missingness < 10%) and were in Hardy Weinberg equilibrium (at $p>1\times10^{-6}$) using the Plink tool kit version 1.90b3a[71]. Multidimensional scaling (MDS) was performed on autosomal LD-independent SNPs to construct genomic ancestry components on each sample, which can be interpreted as quantitative levels of ethnicity, but can also measure other technical factors. This processing and quality control steps resulted in 5,980,012 common variants in this dataset of 295 unique subjects. We remapped variants to hg38 first using dbSNP databases[72] from v142 on hg19 then to v149 on hg38, and then used the liftOver tool[73] for unmapped variants (that were dropped in dbSNP v149).

Analysis Sample

Data on a total of 540 amygdala and sACC samples was generated. From these, five samples were removed as outliers based on alignment metrics, including overall map rate<0.5, gene assignment rate<0.3, or mitochondrial map rate>0.1. Another eight samples were removed due to mis-labelled sample numbers based on comparisons of DNA genotype data above with a panel of 740 exonic/coding SNPs called with the RNA-seq data; two samples were removed due to low microarray quality data; and 14 samples were removed due to apparent region discrepancy based on comparisons of expression levels of the 1,000 most region associated genes. This left a total of 511 samples (268 sACC and 243 AMYG) for downstream analyses.

Data Analysis

Quality surrogate variable analysis (qSVA).—We used quality surrogate variables to account for potential latent RNA quality confounding and adapted a previous multi-region strategy for defining quality surrogate variables[74]. First, we performed tissue

degradation experiments in tissue dissected from sACC and amygdala from a common set of 5 neurotypical donors from both sexes (2 males and 3 females) between ages 27–69 containing multiple ancestries. Each dissection created four tissue aliquots that were placed on ice, and either never taken off ice (0 minutes) or left at room temperature for 15, 30 and 60 minutes. RNA was extracted, sequenced, and processed as above, resulting in 20 RNA-seq samples in each brain region (5 donors and 4 degradation time points). RNA integrity numbers (RINs) dropped from a mean of 7.74 to 6.06 in AMYG and from 8.36 to 7.3 in sACC across this tissue degradation experiment. We then combined all 40 RNA-seq samples to identify common degradation regions across these two brain regions for calculating quality surrogate variables in our larger target dataset, since we needed to use the same set of qSVs for direct comparisons of differentially expressed genes across the two brain regions. We therefore calculated mean coverage separately by strand across all 40 combined samples to define expressed regions with greater than 5 normalized reads and greater than 50bp[75]. We then fit a linear model to each expression region as a function of degradation time adjusting for brain region and donor as fixed effects. We then ranked the expressed regions by the degradation effect and created an input bed file with the top 1000 degradation-susceptible regions for coverage-based quantification in the 511 post-QC RNA-seq samples described above. Subsequently, quality surrogate variables (qSVs) for each sample were calculated once for the entire project from the top principal components (PCs) of the expression in these 1000 degradation regions across all 511 samples, where k was calculated as selected to be 18 using the BE algorithm[76] with the sva Bioconductor package[77].

Differential Expression.—We tested for expression differences in both brain regions between BD cases and unaffected controls at the gene, transcript, exon and junction feature levels with a series of linear regression models using the limma package and “voom”[78,79]. Given their distribution, feature level data were modeled as $\log_2$ of the reads counts normalized as described above. To control for potential confounding, we included in the models terms for age at death, sex, mitochondrial rate, ribosomal RNA rate, total gene assignment, RIN, ERCC spike-in rates, the top 3 principal components of the genotype data (calculated using ~100,000 LD-independent SNPs), and 18 quality surrogate variables (qSVs) described above[74,80]. We restricted the analyses to White European samples, so concerns about confounding by population stratification should be minimized, but we included ancestry related principal components of the genotype data to further protect against possible residual confounding and potential differences in microarray platforms in the imputation procedure. We could not control for certain factors such as substance abuse because by our selection criteria they could only occur among cases and there was no variability among the controls. However, we note that in previous work on similar outcomes sensitivity analyses suggested that accounting for these factors did not materially change the results (10.1101/2021.01.12.426438). We accounted for multiple testing by controlling the false discovery rate (FDR) via the Benjamani-Hochberg algorithm[81], and used FDR<5% to declare findings genome-wide significant.

We carried out an additional cell-type deconvolution analysis to explore if the differential expression results were potentially confounded by cell proportion differences between the

cases and controls. We estimated the proportion of major brain cell types of the bulk samples using a single-nucleus RNA sequence reference sample generated with the 10X protocol from 5 brain regions (the, sACC, AMYG, DLPFC, hippocampus and nucleus encumbens) on 8 neurotypical controls (10.1101/2020.10.07.329839). We used the combined data from all five brain regions for our reference sample to maximize the sample size of the reference sample for determining cell-types profiles[82]. Combining data from all five brain regions and using only controls may add "noise" into the selection of cell-type specific marker genes for subsequent deconvolution of the bulk sACC and AMYG samples as described below. However, because we have found that cell-type differences typically contribute a greater proportion of variance in gene expression than either brain region or disease status, we concluded it was reasonable to proceed with this reference sample, especially given larger, more appropriate reference samples for these brain regions were not available. The deconvolution was performed with the ReferenceBasedDecomposition function from the R package *BisqueRNA* version 1.0.4 (10.1038/s41467-020-15816-6), using the use.overlap = FALSE option. The cell types considered in the deconvolution of the bulk samples were astrocytes, microglia, oligodendrocytes, oligodendrocyte progenitor cells, excitatory neurons, inhibitory neurons, T-cells, vascular smooth muscle cells and endothelial cells. Marker genes were selected by first filtering for genes common between the bulk data and the reference data, then calculating the ratio of the mean expression of each gene in the target cell type over the highest mean expression of that gene in a non-target cell type. The 25 genes with the highest ratios for each cell type were selected as markers. We found that the estimated proportion of major cell types did not systematically differ between cases and controls (Supplementary Figure 6). Moreover, when we explored differential expression models that additionally controlled for cell proportion estimates per sample over and above the covariates described above, the results were nearly unchanged with correlations greater than 0.95 (Supplementary Figure 7). As a result, we reported results without the estimated cell-type proportions.

We further carried out a weighted gene co-expression network analysis (WGCNA) using the gene level count data from both cases and controls and both brain regions analyzed together to identify modules of genes that were co-expressed across conditions[83]. Before running the WGCNA, the unwanted fixed effect covariates controlled for in the differential expression analysis were first regressed out, while preserving the wanted biological/clinical effects of diagnosis, brain region, their statistical interaction, age and sex. We then tested whether the identified modules of genes were enriched in genes that were differentially expressed between BD cases and unaffected controls using simple logistic regression, or whether the eigengene of each module was associated with BD status using linear mixed effects models adjusting for the wanted biological effects above and treating donor as a random intercept to account for the correlations in WGCNA sample eigengenes from the same donor. We further tested whether the modules implicated by the above analyses were enriched for genes of specific GO pathways using simple hypergeometric tests, and whether the more highly connected genes (i.e. hub genes) in the implicated modules were further enriched for differentially expressed genes over the background of all genes in the module using Wilcoxon rank sum tests with the K_{within} metric to define intramodular connectivity. We also tested for cell type enrichments of implicated modules and FDR<5% significant

differentially expressed genes in either brain region using Fisher's exact tests with cell type-specific genes for each reported cell type (astrocytes, endothelial cells, excitatory neurons, inhibitory neurons, microglia, oligodendrocytes, oligodendrocyte progenitor cells, T-cells, and vascular smooth muscle cells) defined based on the single nucleus RNA-seq data described above.

Expression/Splicing QTLs.—In subsequent analyses, we modeled the gene, transcript, exon and junction feature level data as a function of SNP genotypes using linear regression to identify candidate expression quantitative trait loci (eQTL) in each brain region. SNP genotypes were coded as the number of minor alleles (0, 1, or 2). All models controlled for diagnosis, sex, the top principal components of the expression data at each feature level to adjust for technical and potential "batch" effects, and the top 5 principal components of the genotype data to adjust for population sub-structure as above. Given the limited power to detect *trans* eQTL effects, and the more opaque mechanisms underlying their associations, we focused on *cis* effects and tested for associations with features within 1 Mb of each SNP (500kb up- and down-stream). For our primary analysis, we examined SNPs in 881 loci implicated by the PGC study of bipolar disorder with suggestive associations at $p<1\times10^{-4}$ before replication, including 31 loci that were genome-wide significant at $p<5\times10^{-8}$ after replication[2]. The loci were defined by an index SNP with the strongest p-value of association in the PGC analysis plus all nearby proxy SNPs with $r^2> 0.8$ with the index SNP. Statistical significance for genotype effects on each feature type was controlled using false discovery rate (FDR) considering all tests with that feature type across the 881 loci.

We retained for further analyses those loci with significant SNP-feature associations at FDR<1%, focusing specifically on associations with the lead SNP, defined as the index SNP or the closest proxy SNP with the most significantly associated feature if the index SNP was not genotyped or not significantly associated with a feature. Given the considerable correlation between features in a given locus, we carried out a conditional analysis in which we sequentially conditioned on the most significantly associated feature and tested the remaining features to determine which were independently associated with the lead SNP[11]. We also tested for differences in the association of the lead SNP with the most significant feature across brain regions by combining data from the brain regions and including interaction terms between SNP genotype and indicator variables for each region.

We also analyzed the data for splicing QTLs (sQTL) using leafcutter.v.0.2.8[84] to process the filtered junction read counts generated above and FastQTL[85] to test for sQTLs with SNPs in the 881 loci implicated by PGC-BD2. Per leafcutter recommendations, we filtered any junction reads that were found in less than 40% of the sample. In the sQTL tests we controlled for the same covariates as included in the models testing for eQTLs, except we included the top principal components of an analysis of the intron ratios generated by leafcutter. As with the eQTLs, we focused on lead sQTLs at FDR<1%.

Transcriptome-wide Association Study (TWAS).—We adapted a previously-developed TWAS workflow[74] that was, in turn, modified from the published TWAS approach by Gusev and colleagues[86]. We first filtered target SNPs from the sACC and

amygdala samples to those present in our eQTL analyses and harmonized the names and coordinates of the SNPs across sample datasets and GWAS summary statistics. We also converted the GWAS summary statistics from hg19 to hg38 coordinates[2]. We then computed feature weights at the gene level separately for the amygdala and sACC and applied the weights to the GWAS summary statistic SNPs and calculated the functional-GWAS association statistics using modified versions of TWAS-FUSION R scripts provided by Gusev and colleagues (http://gusevlab.org/projects/fusion/).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by a grant from the National Institute of Mental Health (R01MH105898). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The authors gratefully acknowledge the contributions of the Offices of the Chief Medical Examiner of Maryland, Washington DC, and Northern Virginia for collaborating in the accession of post-mortem human brain donations that were used in this study. Dr. Lewellyn B. Bigelow and members of the Neuropathology Section of the Lieber Institute for Brain Development made important contributions in the clinical characterization and diagnosis of the donors. Brianna Barry also contributed to the analyses of the RNA sequencing data. Finally, the authors would like to express their gratitude to families of the brain donors whose generosity made this study possible.

Data Availability.

All data generated in this study are made available through the PsychENCODE Consortium. Access to the data is managed by the NIMH Repository & Genomics Resource (NRGR) and the data are distributed via Synapse under the BipSeq study (syn5844980). Instructions for access to the data through NRGR are provided here: https://www.nimhgenetics.org/resources/psychencode

More information for browsing the available data and instructions for accessing the data is also provided on the PsychENCODE Knowledge Portal: https://psychencode.synapse.org/

References

1. Craddock N & Sklar P Genetics of bipolar disorder. Lancet 381, 1654–1662 (2013). [PubMed: 23663951]
2. Stahl EA et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat. Genet 51, 793–803 (2019). [PubMed: 31043756]
3. Sullivan PF et al. Psychiatric genomics: an update and an agenda. Am. J. Psychiatry 175, 15–27 (2018). [PubMed: 28969442]
4. Gallagher MD & Chen-Plotkin AS The Post-GWAS Era: From Association to Function. Am. J. Hum. Genet 102, 717–730 (2018). [PubMed: 29727686]
5. GTEx Consortium et al. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017). [PubMed: 29022597]
6. Roth RB et al. Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. Neurogenetics 7, 67–80 (2006). [PubMed: 16572319]
7. Kang HJ et al. Spatio-temporal transcriptome of the human brain. Nature 478, 483–489 (2011). [PubMed: 22031440]
8. Drevets WC, Savitz J & Trimble M The subgenual anterior cingulate cortex in mood disorders. CNS Spectr. 13, 663–681 (2008). [PubMed: 18704022]

9. Strakowski SM et al. The functional neuroanatomy of bipolar disorder: a consensus model. Bipolar Disord. 14, 313–325 (2012). [PubMed: 22631617]
10. Gandal MJ et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science 362, (2018).
11. Jaffe AE et al. Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. Nat. Neurosci 21, 1117–1125 (2018). [PubMed: 30050107]
12. 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 (2015). [PubMed: 26238605]
13. Darby MM, Yolken RH & Sabunciyan S Consistently altered expression of gene sets in postmortem brains of individuals with major psychiatric disorders. Transl. Psychiatry 6, e890 (2016). [PubMed: 27622934]
14. Kim S, Hwang Y, Webster MJ & 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 (2016). [PubMed: 26077692]
15. 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 (2014). [PubMed: 24393808]
16. Akula N, Wendland JR, Choi KH & McMahon FJ An integrative genomic study implicates the postsynaptic density in the pathogenesis of bipolar disorder. Neuropsychopharmacology 41, 886–895 (2016). [PubMed: 26211730]
17. Hu J et al. Systematically characterizing dysfunctional long intergenic non-coding RNAs in multiple brain regions of major psychosis. Oncotarget 7, 71087–71098 (2016). [PubMed: 27661005]
18. Ramaker RC et al. Post-mortem molecular profiling of three psychiatric disorders. Genome Med. 9, 72 (2017). [PubMed: 28754123]
19. Pacifico R & Davis RL Transcriptome sequencing implicates dorsal striatum-specific gene network, immune response and energy metabolism pathways in bipolar disorder. Mol. Psychiatry 22, 441–449 (2017). [PubMed: 27350034]
20. Luykx JJ, Giuliani F, Giuliani G & Veldink J Coding and Non-Coding RNA Abnormalities in Bipolar Disorder. Genes (Basel) 10, (2019).
21. Modabbernia A, Taslimi S, Brietzke E & Ashrafi M Cytokine alterations in bipolar disorder: a meta-analysis of 30 studies. Biol. Psychiatry 74, 15–25 (2013). [PubMed: 23419545]
22. Pinto JV et al. Neuron-glia Interaction as a Possible Pathophysiological Mechanism of Bipolar Disorder. Curr. Neuropharmacol 16, 519–532 (2018). [PubMed: 28847296]
23. Rao JS, Harry GJ, Rapoport SI & Kim HW Increased excitotoxicity and neuroinflammatory markers in postmortem frontal cortex from bipolar disorder patients. Mol. Psychiatry 15, 384–392 (2010). [PubMed: 19488045]
24. Söderlund J et al. Elevation of cerebrospinal fluid interleukin-1ß in bipolar disorder. J. Psychiatry Neurosci 36, 114–118 (2011). [PubMed: 21138659]
25. Stertz L, Magalhães PVS & Kapczinski F Is bipolar disorder an inflammatory condition? The relevance of microglial activation. Curr. Opin. Psychiatry 26, 19–26 (2013). [PubMed: 23196997]
26. Varma VR et al. Alpha-2 macroglobulin in Alzheimer's disease: a marker of neuronal injury through the RCAN1 pathway. Mol. Psychiatry 22, 13–23 (2017). [PubMed: 27872486]
27. Guerreiro R et al. TREM2 variants in Alzheimer's disease. N. Engl. J. Med 368, 117–127 (2013). [PubMed: 23150934]
28. Hook V et al. Cathepsin B in neurodegeneration of Alzheimer's disease, traumatic brain injury, and related brain disorders. Biochim. Biophys. Acta Proteins Proteom 1868, 140428 (2020). [PubMed: 32305689]
29. Sims R et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. Nat. Genet 49, 1373–1384 (2017). [PubMed: 28714976]
30. Diniz BS et al. History of Bipolar Disorder and the Risk of Dementia: A Systematic Review and Meta-Analysis. Am. J. Geriatr. Psychiatry 25, 357–362 (2017). [PubMed: 28161155]

31. Drange OK et al. Genetic overlap between alzheimer's disease and bipolar disorder implicates the MARK2 and VAC14 genes. Front. Neurosci 13, 220 (2019). [PubMed: 30930738]
32. Alda M Pharmacogenetics of lithium response in bipolar disorder. J. Psychiatry Neurosci 24, 154–158 (1999). [PubMed: 10212559]
33. Liu T, Zhang L, Joo D & Sun S-C NF-κB signaling in inflammation. Signal Transduct. Target. Ther 2, (2017).
34. Elhaik E & Zandi P Dysregulation of the NF-κB pathway as a potential inducer of bipolar disorder. J. Psychiatr. Res 70, 18–27 (2015). [PubMed: 26424419]
35. Wake H, Moorhouse AJ & Nabekura J Functions of microglia in the central nervous system--beyond the immune response. Neuron Glia Biol. 7, 47–53 (2011). [PubMed: 22613055]
36. Dresselhaus EC & Meffert MK Cellular Specificity of NF-κB Function in the Nervous System. Front. Immunol 10, 1043 (2019). [PubMed: 31143184]
37. Purcell SM et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature 506, 185–190 (2014). [PubMed: 24463508]
38. Goes FS et al. De novo variation in bipolar disorder. Mol. Psychiatry (2019) doi:10.1038/s41380-019-0611-1.
39. Toma C et al. An examination of multiple classes of rare variants in extended families with bipolar disorder. Transl. Psychiatry 8, 65 (2018). [PubMed: 29531218]
40. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660 (2015). [PubMed: 25954001]
41. Ben-Shalom R et al. Opposing effects on nav1.2 function underlie differences between SCN2A variants observed in individuals with autism spectrum disorder or infantile seizures. Biol. Psychiatry 82, 224–232 (2017). [PubMed: 28256214]
42. Wolff M et al. Genetic and phenotypic heterogeneity suggest therapeutic implications in SCN2A-related disorders. Brain 140, 1316–1336 (2017). [PubMed: 28379373]
43. Sanders SJ et al. Progress in Understanding and Treating SCN2A-Mediated Disorders. Trends Neurosci. 41, 442–456 (2018). [PubMed: 29691040]
44. Selten J-P, Lundberg M, Rai D & Magnusson C Risks for nonaffective psychotic disorder and bipolar disorder in young people with autism spectrum disorder: a population-based study. JAMA Psychiatry 72, 483–489 (2015). [PubMed: 25806797]
45. Baek J-H, Rubinstein M, Scheuer T & Trimmer JS Reciprocal changes in phosphorylation and methylation of mammalian brain sodium channels in response to seizures. J. Biol. Chem 289, 15363–15373 (2014). [PubMed: 24737319]
46. Myers SJ et al. Distinct roles of GRIN2A and GRIN2B variants in neurological conditions. [version 1; peer review: 2 approved]. F1000Res. 8, (2019).
47. de Sousa RT et al. Genetic studies on the tripartite glutamate synapse in the pathophysiology and therapeutics of mood disorders. Neuropsychopharmacology 42, 787–800 (2017). [PubMed: 27510426]
48. Itokawa M et al. Genetic analysis of a functional GRIN2A promoter (GT)n repeat in bipolar disorder pedigrees in humans. Neurosci. Lett 345, 53–56 (2003). [PubMed: 12809987]
49. Fountoulakis KN The possible involvement of NMDA glutamate receptor in the etiopathogenesis of bipolar disorder. Curr. Pharm. Des 18, 1605–1608 (2012). [PubMed: 22280433]
50. Li M et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science 362, (2018).
51. Lipska BK et al. Critical factors in gene expression in postmortem human brain: Focus on studies in schizophrenia. Biol. Psychiatry 60, 650–658 (2006). [PubMed: 16997002]
52. Zalcman S & Endicott J Diagnostic Evaluation after Death, Developed for NIMH Neurosciences Research Branch. Department of Research Assessment and Training (1983).
53. Kelly TM & Mann JJ Validity of DSM-III-R diagnosis by psychological autopsy: a comparison with clinician ante-mortem diagnosis. Acta Psychiatr. Scand 94, 337–343 (1996). [PubMed: 9124080]
54. Deep-Soboslay A et al. Reliability of psychiatric diagnosis in postmortem research. Biol. Psychiatry 57, 96–101 (2005). [PubMed: 15607306]

55. First MB, Spitzer RL, Gibbon M & Williams JBW Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV). American Psychiatric Press, Inc (1996).
56. Association AP & Others. Diagnostic and statistical manual of mental disorders-IV-TR. Washington, DC: American Psychiatric Association (2000).
57. Babraham Bioinformatics. FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2016).
58. Bolger AM, Lohse M & Usadel B Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014). [PubMed: 24695404]
59. Kim D, Langmead B & Salzberg SL HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360 (2015). [PubMed: 25751142]
60. Liao Y, Smyth GK & Shi W featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930 (2014). [PubMed: 24227677]
61. Anders S, Pyl PT & Huber W HTSeq — a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015). [PubMed: 25260700]
62. McDonnell Genome Institute, T. G. L. regtools https://regtools.readthedocs.io/en/latest/.
63. Kim D et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013). [PubMed: 23618408]
64. Steijger T et al. Assessment of transcript reconstruction methods for RNA-seq. Nat. Methods 10, 1177–1184 (2013). [PubMed: 24185837]
65. Patro R, Duggal G, Love MI, Irizarry RA & Kingsford C Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417–419 (2017). [PubMed: 28263959]
66. Bray NL, Pimentel H, Melsted P & Pachter L Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol 34, 525–527 (2016). [PubMed: 27043002]
67. Collado-Torres L & Jaffe AE jaffelab: Commonly used functions by the Jaffe lab. (2017).
68. Howie BN, Donnelly P & Marchini J A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009). [PubMed: 19543373]
69. Delaneau O, Coulonges C & Zagury J-F Shape-IT: new rapid and accurate algorithm for haplotype inference. BMC Bioinformatics 9, 540 (2008). [PubMed: 19087329]
70. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. Nature 526, 68–74 (2015). [PubMed: 26432245]
71. Purcell S et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet 81, 559–575 (2007). [PubMed: 17701901]
72. Sherry ST et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29, 308–311 (2001). [PubMed: 11125122]
73. Hinrichs AS et al. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 34, D590–8 (2006). [PubMed: 16381938]
74. Collado-Torres L et al. Regional Heterogeneity in Gene Expression, Regulation, and Coherence in the Frontal Cortex and Hippocampus across Development and Schizophrenia. Neuron 103, 203–216.e8 (2019). [PubMed: 31174959]
75. Collado-Torres L et al. Flexible expressed region analysis for RNA-seq with derfinder. Nucleic Acids Res. 45, e9 (2017). [PubMed: 27694310]
76. Buja A & Eyuboglu N Remarks on parallel analysis. Multivariate Behav. Res 27, 509–540 (1992). [PubMed: 26811132]
77. Leek JT & Storey JD Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 3, 1724–1735 (2007). [PubMed: 17907809]
78. Law CW, Chen Y, Shi W & Smyth GK voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15, R29 (2014). [PubMed: 24485249]
79. Ritchie ME et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47 (2015). [PubMed: 25605792]
80. Jaffe AE et al. qSVA framework for RNA quality correction in differential expression analysis. Proc Natl Acad Sci USA 114, 7130–7135 (2017). [PubMed: 28634288]

81. Benjamini Y & Hochberg Y Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological) 57, 289–300 (1995).

82. Jew B et al. Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. Nat. Commun 11, 1971 (2020). [PubMed: 32332754]

83. Langfelder P & Horvath S WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559 (2008). [PubMed: 19114008]

84. Li YI et al. Annotation-free quantification of RNA splicing using LeafCutter. Nat. Genet 50, 151–158 (2018). [PubMed: 29229983]

85. Ongen H, Buil A, Brown AA, Dermitzakis ET & Delaneau O Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics 32, 1479–1485 (2016). [PubMed: 26708335]

86. Gusev A et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet 48, 245–252 (2016). [PubMed: 26854917]

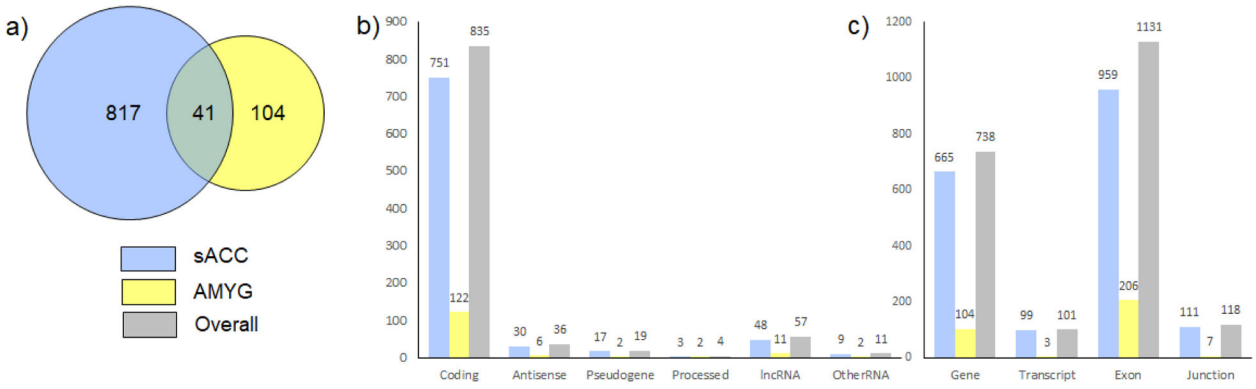

Figure 1.

Summary of FDR<5% significant differentially expressed features overall and by brain region. A total of 25,136 genes, 73,214 transcripts, 396,818 exons, and 266,197 junctions were tested in the sACC and amygdala. Results shown include: a) Venn diagram of the overlap in the number of unique genes with a significant feature by brain region; b) breakdown of gene types for the unique set of genes represented in a); and c) breakdown of the significant features (gene, transcript, exon or junction) in the implicated genes.

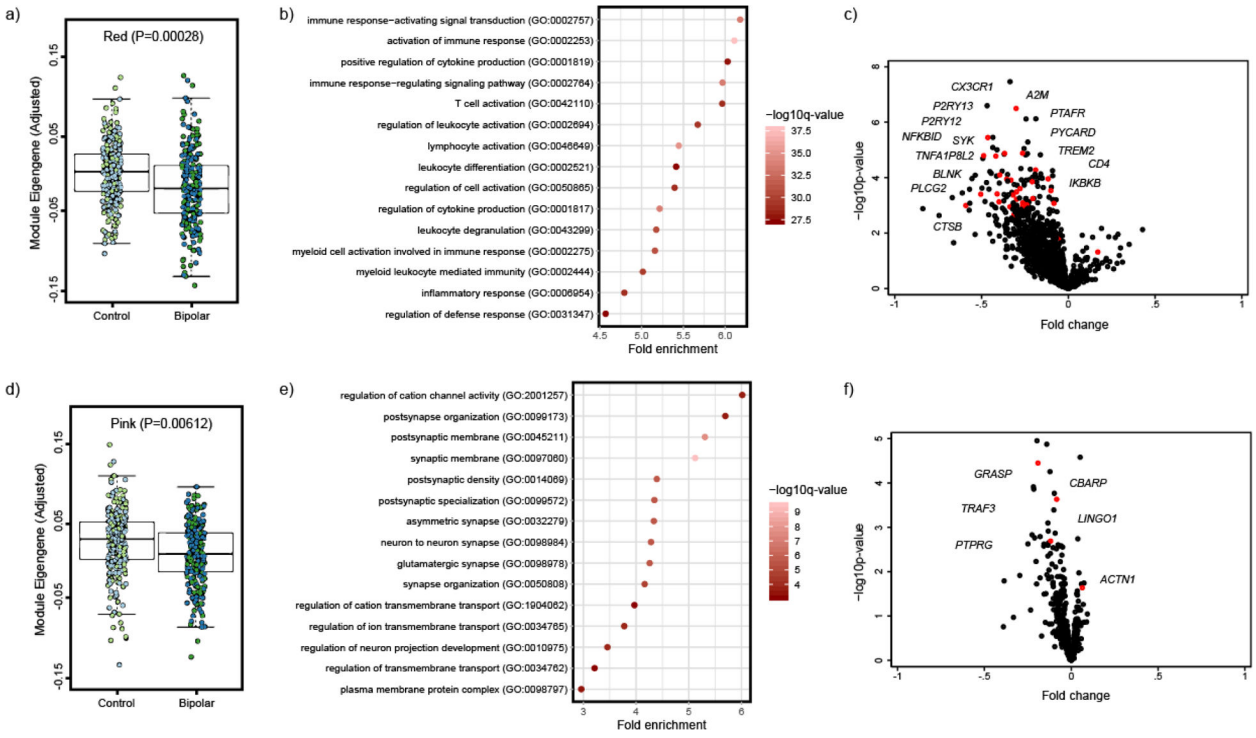

Figure 2.

Results of WGCNA analysis: a) boxplot of red module eigengene values for each sample by case-control status controlling for age and brain region; b) top 15 enriched pathways in the red module containing genes with FDR<5% significant differentially expressed features; c) volcano plot of fold change in gene level expression by −log10pvalue for genes in the red module, with those genes in the top 15 pathways that have a significant differentially expressed features highlighted in red (note some may have differentially expressed features other than at the gene level which is shown here); d) boxplot of pink module eigengene values for each sample by case-control status, controlling for age and brain region; e) top 15 enriched pathways in the pink module containing genes with FDR<5% significant differentially expressed features; and f) volcano plot of fold change in gene level expression by −log10pvalue for genes in the pink module, with those genes in the top 15 pathways that have a significant differentially expressed features highlighted in red (note some may have differentially expressed features other than at the gene level which is shown here). Boxplots in a) and d) show data on 126 BD cases and 142 controls in the sACC (green circles) together with 121 BD cases and 122 controls in the amygdala (blue circles). The boxplots display the median as the center line, the interquartile range (IQR; 25th – 75th percentile) as the box range, and 1.5 times the IQR as the whiskers unless a minimum/maximum is reached.

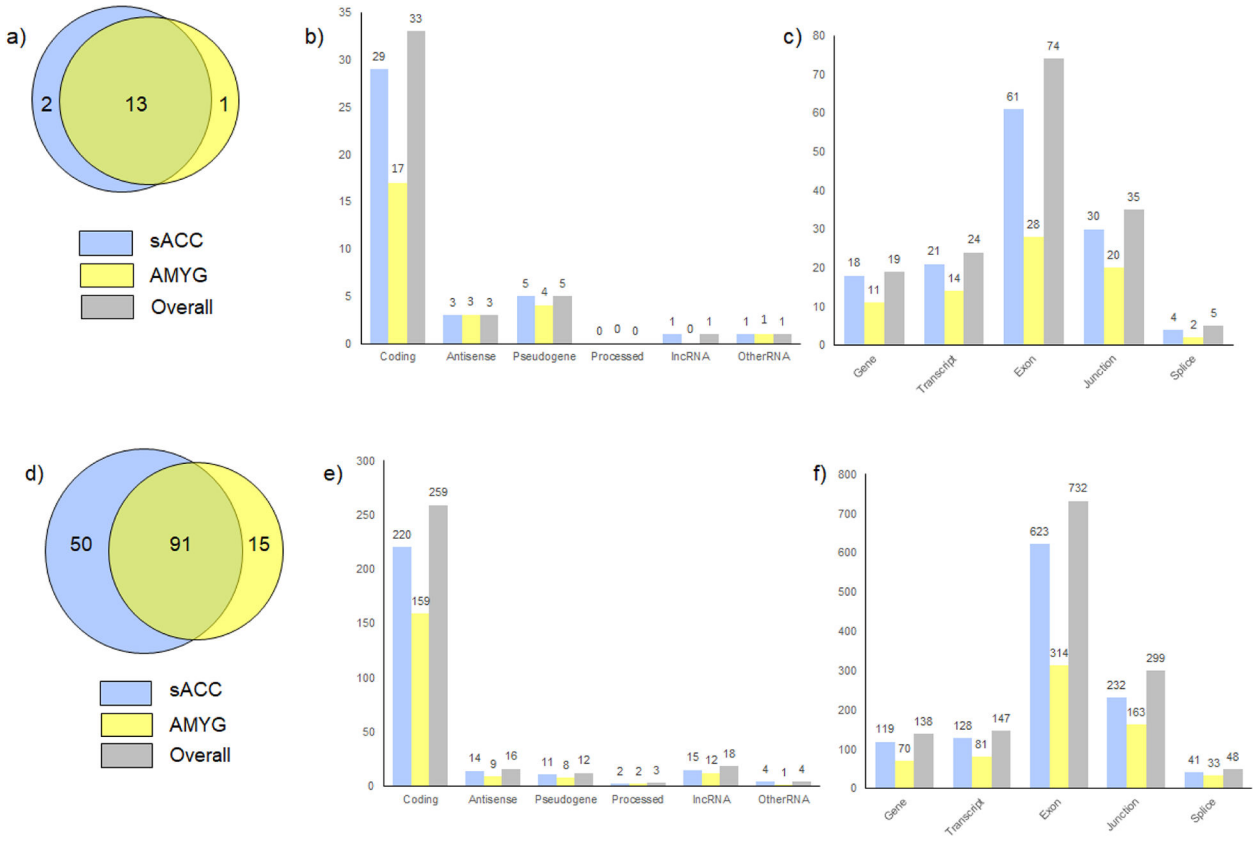

Figure 3.

Summary of FDR<1% expression (eQTL) and splicing (sQTL) quantitative trait loci in loci suggested by the latest Psychiatric Genomics Consortium (PGC) genome-wide association study of bipolar disorder. Tests were carried out for QTL associations between 10,777 SNPs and gene (n=4,647), transcript (n=14,434), exon (n=76,589) and junction (n=49,188) features in these loci, and results summarized for the lead SNP in each locus. a) Overlap of loci with significant eQTL/sQTL by brain region among the 30 genome-wide significant loci. b) and c) Breakdown of gene types for the implicated genes in a) and the features that are associated in the genome-wide significant loci. d) Overlap of loci with significant eQTL/sQTL by brain region among the 850 suggestive genome-wide loci. e) and f) Breakdown of gene types for the implicated genes in d) and the features that are associated in these loci.

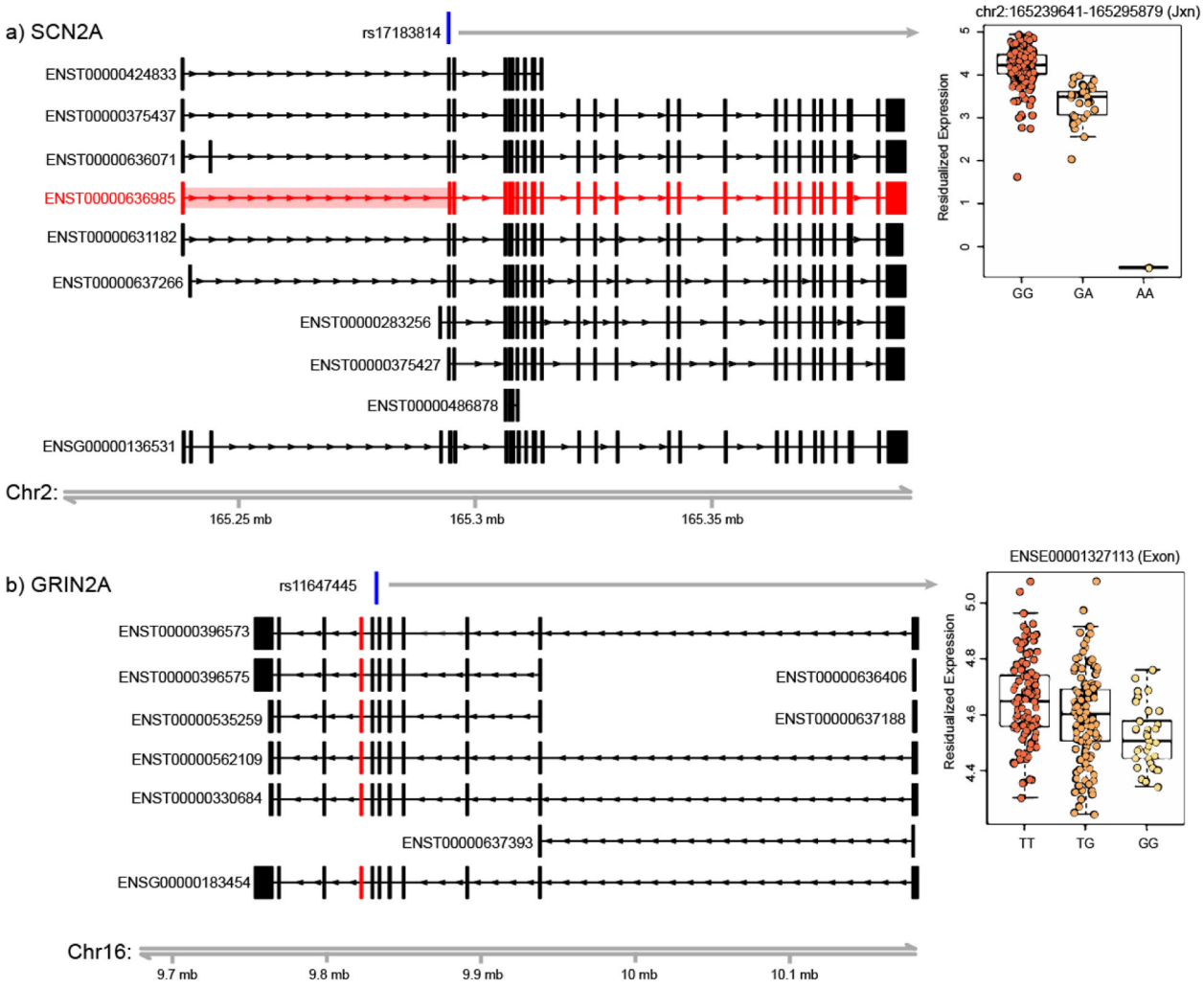

Figure 4.

Gene visualization plots and accompanying SNP-feature scatterplots for the lead SNP with significant e/sQTLs (FDR<1%). The gene visualization plots show the location of the lead SNP in blue and exon/intron models in black for all protein coding transcripts and the union gene model for each gene. Specific gene features are shown in red if they are down-regulated in the e/sQTL and in green if they are up-regulated. Red/green highlighted boxes over introns represent down/up regulated junctions between corresponding exons. The SNP-feature boxplots show residualized expression levels (from the eQTL linear regression models described in the Methods) of the most significant feature with the lead SNP, with SNP genotypes shown from most to least common: a) *SCN2A* ($\beta=-1.12$, $p=5.08\times10^{-29}$) and b) *GRIN2A* ($\beta=-0.07$, $p=1.32\times10^{-5}$). Boxplots for a) and b) show results based on 262 samples (126 BD cases and 142 controls) from the sACC and display the median as the center line, the IQR as the box range, and 1.5 times the IQR as the whiskers unless a minimum/maximum is reached.

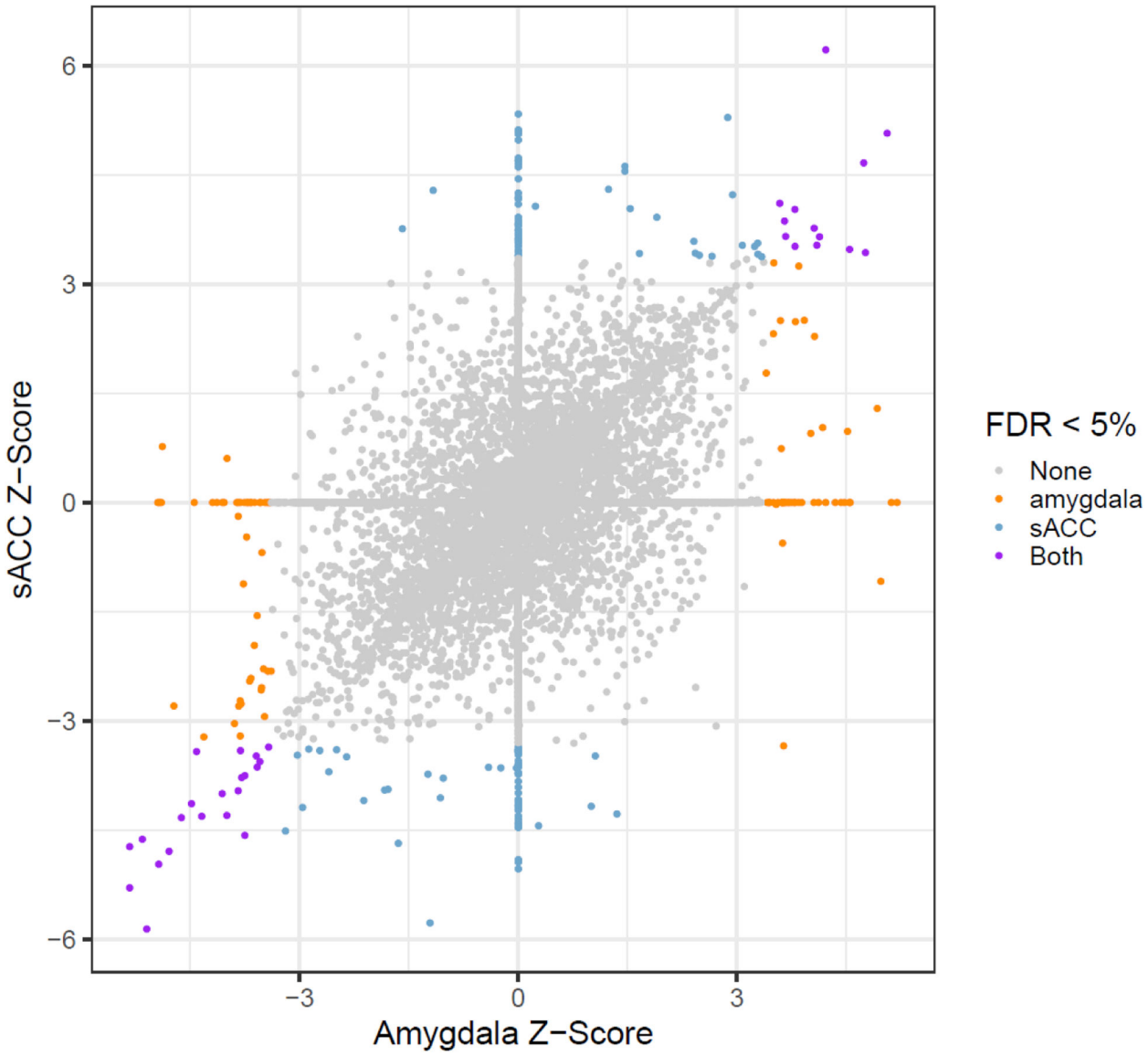

Figure 5.

Scatterplot of Z-score test statistics from TWAS of genes in the sACC versus amygdala. A total of 13,822 genes were tested across both brain regions. Points highlighted in colors are FDR<5% significant in the amygdala (n=125), sACC(n=156), or both regions (n=34). Points along the X=0 or Y=0 axes were not estimated in the amygdala or sACC, respectively, typically because heritability estimates failed in one of the regions and subsequent models could not be estimated.

Table 1.

Details of the FDR<1% cis eQTLs with Lead SNPs in GWAS Significant Loci by Brain Region *

Index SNP	Chr:Pos:Ref:Alt	AMYG	sACC	Best P	GWAS Effect (+/−)	e/sQTL Effect (+/−)
rs17183814	Chr2:166152389:G:A	SCN2A (Sp)	SCN2A (**Sp**)	5.47E-22	−	−
rs11557713	Chr18:60243876:G:A	ZCCHC2 (Ex)	ZCCHC2 (**Tx**)	3.25E-13	+	−
rs112114764	Chr17:42201041:G:T	ASB16-AS1 (Gn)	ASB16-AS1 (**Gn**)	9.85E-08	+	−
rs9834970	Chr3:36856030:T:C	N/A	TRANK1 (**Gn**)	2.32E-07	−	−
rs4447398	Chr15:42904904:A:C	LRRC57 (Gn)	LRRC57 (**Gn**)	5.54E-07	−	−
rs11647445	Chr16:9926966:T:G	N/A	GRIN2A (**Ex**)	1.32E-05	+	−
rs57195239	Chr2:97376407:A:AT	LMAN2L (**Gn**)	LMAN2L (Jx)	2.19E-05	−	+
rs3804640	Chr3:107793709:A:G	BBX (**Ex**)	N/A	2.45E-05	−	−
rs10035291	Chr5: 80796368:T:C	SSBP2 (Jx)	SSBP2 (**Ex**)	1.01E-04	−	−

Gn=gene; Tx=transcript; Ex=exon; Jx=Junction; Sp=Splicing

*Shown are conditionally independent associations with the lead SNPs at FDR<1% across the 2 brain regions where the evidence points to a single gene in one of the 31 genome-wide significant loci from PGC-BD2[2]. Gene symbols and the features in these genes that are associated are shown. The feature in parenthesis is the most significant feature in the specific brain region, and the bolded feature is the most significant feature across the 2 brain regions. The Best P shows the p-value for the most significant SNP-feature across both brain regions. GWAS and e/sQTL effects show the observed direction of effect of the alternate allele with + for increased risk/up-regulation and − for decreased risk/down-regulation. Results are based on QTL models described in the Methods.