OPEN

# Does single gene expression omnibus data mining analysis apply for only tumors and not mental illness? A preliminary study on bipolar disorder based on bioinformatics methodology

Xu You, PhD (iD), Yunqiao Zhang, PhD, Qing Long, PhD, Zijun Liu, MM, Ziqiao Feng, MM, Wengyu Zhang, MM, Zhaowei Teng, PhD\*, Yong Zeng, PhD\*

## Abstract

**Background:** Bipolar disorder (BD), a common kind of mood disorder with frequent recurrence, high rates of additional comorbid conditions and poor compliance, has an unclear pathogenesis. The Gene Expression Omnibus (GEO) database is a gene expression database created and maintained by the National Center for Biotechnology Information. Researchers can download expression data online for bioinformatics analysis, especially for cancer research. However, there is little research on the use of such bioinformatics analysis methodologies for mental illness by downloading differential expression data from the GEO database.

**Methods:** Publicly available data were downloaded from the GEO database (GSE12649, GSE5388 and GSE5389), and differentially expressed genes (DEGs) were extracted by using the online tool GEO2R. A Venn diagram was used to screen out common DEGs between postmortem brain tissues and normal tissues. Functional annotation and pathway enrichment analysis of DEGs were performed by using Gene ontology and Kyoto Encyclopedia of Genes and Genomes analyses, respectively. Furthermore, a protein-protein interaction network was constructed to identify hub genes.

**Results:** A total of 289 DEGs were found, among which 5 of 10 hub genes [HSP90AA1, HSP90AB 1, UBE2N, UBE3A, and CUL1] were identified as susceptibility genes whose expression was downregulated. Gene ontology and Kyoto Encyclopedia of Genes and Genomes analyses showed that variations in these 5 hub genes were obviously enriched in protein folding, protein polyubiquitination, apoptotic process, protein binding, the ubiquitin-mediated proteolysis pathway, and protein processing in the endoplasmic reticulum pathway. These findings strongly suggested that HSP90AA1, UBE3A, and CUL 1, which had large areas under the curve in receiver operator curves ($P < .05$), were potential diagnostic markers for BD.

**Conclusion:** Although there are 3 hub genes [HSP90AA1, UBE3A, and CUL 1] that are tightly correlated with the occurrence of BD, mainly based on routine bioinformatics methods for cancer-related disease, the feasibility of applying this single GEO bioinformatics approach for mental illness is questionable, given the significant differences between mental illness and cancer-related diseases.

**Abbreviations:** ASDs = autism spectrum disorders, BD = bipolar disorder, BP = biological process, CC = cellular component, DEGs = differentially expressed genes, GEO = gene expression omnibus, GO = Gene ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes, MF = molecular function, PPI = protein-protein interaction.

**Keywords:** bioinformatics, Bipolar disorder, CUL1, HSP90AA1, UBE3A

## 1. Introduction

Bipolar disorder (BD) is a widely distributed mental disorder that affects approximately 1% of the population worldwide. Epidemiological studies have indicated that the prevalence of BD during the human lifetime ranges from 2% to 4%, placing heavy burdens on families and society.[1] As reported, BD affects 4.4% of the population in the USA,[2] and its prevalence in European countries ranges from 0.1% to 6.0%.[3] Currently, although the diagnosis of BD is mainly dependent on insights into the characteristic clinical symptoms, including repeated and alternating presentation of manic/mild manic and depressive episodes with irregular forms of seizures, the cause of BD remains mysterious. Unfortunately, BD patients who have not been diagnosed and treated appropriately have serious adverse consequences, such as a higher rate of suicide and an increased number of additional psychiatric and physical comorbidities,[4] making it the tenth most common disabling condition.[5] Many hypotheses have been proposed to clarify the pathogenesis of BD. It is well known and seems to be undeniable that an increasing number of studies focused on affective disorders are becoming involved in the domain of the environment and genes.[6–10] It was revealed in a recent magnetic resonance imaging study that the brain tissue structure of BD patients is widely affected, particularly with reductions in gray matter volume,[11,12] accompanied by enlargement of the brain ventricle[13,14] and corpus callosum impairment.[15,16] Accumulating evidence suggests that the occurrence of BD is implicated in various disorders, such as inflammatory disorders,[17] dysfunction of neuroplasticity,[18] oxidative stress, and mitochondrial dysfunction.[19] Therefore, for more accurate diagnosis, novel candidate biomarkers of BD should be searched for and identified in these interrelated biological processes (BP).[20] To date, it has been widely accepted that brain-derived neurotrophic factor is the most popular biomarker for BD due to its participation in the maintenance of adult neuroplasticity and the regulation of synaptic activity and neurotransmitter synthesis.[21–23] Explorations of novel and sensitive diagnostic targets for BD in clinical practice are both challenging and urgently needed, giving new insight to improve therapeutic effects in the future.

In recent years, microarray technology has been widely utilized in the investigation of general genetic abnormalities.[24,25] gene expression omnibus (GEO)2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/) provides online datasets freely, and researchers can obtain information on differentially expressed genes (DEGs) by comparing disease samples with normal samples based on GEO series.[26] Therefore, GEO2R can offer an ideal direction for obtaining a comprehensive understanding of the underlying pathogenesis of various diseases with the help of profiling gene expression datasets. However, to date, few studies have focused on the pathogenesis of mental disorders by using the bioinformatics tool GEO2R to elucidate the genetic factors involved. In the present study, we attempted to search for novel potential biomarkers, taking them as potential BD diagnostic targets dependent on the analysis from 3 main data banks. Importantly, the integrated bioinformatics approaches we used in this study greatly contributed to the search for and identification of 3 promising genes (UBE3A, CUL 1, and HSP90AA1), which are at least related to the incidence of BD. Our results will undoubtedly shed new light on superior diagnostic approaches and timely therapies for BD, and they also provide some insight into whether routine bioinformatics analysis using single GEO databases can be applied for mental illness.

## 2. Materials and methods

### 2.1. Data source

The gene expression datasets analyzed in this current study were extracted from the GEO database. This study aimed to mine data for bioinformatics analysis and did not involve original animal and human clinical experiments, so ethical review was not applicable. A total of 82 BD cases were retrieved from the database. When the type of organism was limited to *Homo sapiens*, 77 cases were obtained by further screening.

Following a careful review, the 3 gene expression profiles included in this study (GSE12649, GSE5388 and GSE5389) were selected due to the common platform, and their experimental types were all expression profiling by microarray all based on the GPL96 [HG-U133A] Affymetrix Human Genome U133A Array. The array data from GSE12649 contained 102 samples, among which 67 samples from the postmortem human prefrontal cortex were used for the present study, including 33 BD samples and 34 healthy samples; the array data from GSE5388 contained 61 postmortem dorsolateral prefrontal cortex samples, including 30 BD samples and 31 healthy samples, and the array data from GSE5389 contained 21 orbitofrontal cortex samples, including 10 BD samples and 11 healthy samples.

### 2.2. Data extraction for DEGs

The GEO2R analysis tool was employed online to exploit the DEGs among all samples. There were significant differences in gene expression levels between patients with mental illness and those with other diseases (such as tumors). For example, if we chose the screening criteria for DEGs of tumor diseases as |log fold change (log FC)| > 1 or 2, even 0.2 to 0.9, there were only approximately 10 DEGs that could be screened out for mental diseases. To avoid missing more valuable DEGs, we artificially defined $P < 0.05$ and |log FC| > 0.1 as inclusion criteria for DEGs. DEGs with log FC < 0 were considered downregulated genes, while DEGs with log FC > 0 were considered upregulated genes. The data from the 3 gene expression profiles (GSE12649, GSE5388 and GSE5389) were statistically analyzed, volcano maps of 3 datasets were drawn using the volcano plotting tool (https://shengxin.ren), and the intersection of upregulated and downregulated DEGs was identified using the Venn diagram web tool (http://bioinformatics.psb.ugent.be/webtools/Venn/).

### 2.3. Functional Gene ontology (GO) and pathway enrichment analysis of DEGs

GO annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of the DEGs were performed by using the online tool of the DAVID database (6.8) (https://david.ncifcrf.gov/). The GO project covers 3 classical domains: BP, cellular component (CC) and molecular function (MF). KEGG is a useful database for investigating genomes, biological pathways, diseases, chemicals and drugs. In this study, GO annotations and KEGG enrichment pathways were both considered to have statistical value if they met the cut-off criterion of $P < 0.05$ and ≥5 genes.

### 2.4. Analysis of the protein-protein interaction (PPI) network and hub proteins

The open-source software of the Search Tool for the Retrieval of Interacting Genes/Proteins named STRING (https://string-db.

**Statistical data for GSE12649, GSE5388, and GSE5389 derived from the GEO database.**

| Dataset ID | BD | Control | Platform | Sample type | Gene counts (upregulation) | Gene counts (downregulation) |
|---|---|---|---|---|---|---|
| GSE12649 | 33 | 34 | GPL96 | human prefrontal cortex (BA46) | 618 | 1064 |
| GSE5388 | 30 | 31 | GPL96 | dorsolateral prefrontal cortex | 1263 | 1396 |
| GSE5389 | 10 | 11 | GPL96 | orbitofrontal cortex | 1947 | 1392 |

GEO = gene expression omnibus.

org/) was utilized to construct the network of PPI information. The previously determined DEGs were mapped to the STRING database to assess potential PPI relationships. PPI pairs were extracted with a combined score > 0.4. The visualized network of PPIs was generated via Cytoscape software (https://cyto scape.org/), and the CytoHubba plugin downloaded from Cytoscape was used to compute the connectivity degree of individual protein nodes. Finally, in the current study, the top 10 genes were considered hub genes and crucial in the PPI network.

### 2.5. Screening out more potential hub genes and mining their expression patterns in the normal human brain

To study the potential signaling pathways of the 10 hub genes and explore more potential hub genes, KEGG pathway enrichment analysis was employed by using DAVID (6.8) (P < .05). In addition, the online software BrainCloud (http://braincloud.jhmi.edu/) constructed by Colantuoni et al[27] was used to roughly elucidate the expression levels of these additional potential genes in normal tissues, assessing the temporal dynamics and genetic control of transcription of the human prefrontal cortex throughout the lifetime. With this method, we preliminarily compared the expression levels of potential genes in BD patients with healthy individuals and identified the differences. In addition to providing a dynamic understanding of the expression changes of these hub genes, the National Center for Biotechnology Information (NCBI) database can provide information on the static expression levels of these genes in normal brain tissue for further investigation.

### 2.6. Diagnostic value of 5 hub genes

The application of Prism(8.0) (https://www.graphpad.com/scientific-software/prism/) was used to construct the receiver operating characteristic curve, which could help determine the diagnostic value of the hub genes for predicting BD. P < 0.05 was considered to indicate a statistically significant difference.

## 3. Results

### 3.1. Identification of DEGs

A total of 3 microarray datasets (GSE12649, GSE5388 and GSE5389) were used in this study. Based on the criteria of $P < .05$ and $|logFC| > 0.1$, a total of 2659 DEGs were identified from GSE5388, among which 1263 genes were upregulated and 1396 genes were downregulated. There were 3339 DEGs identified in GSE5389, including 1947 upregulated genes and 1392 down-regulated genes. A total of 1682 DEGs were identified in GSE12649, among which 618 were upregulated and 1064 were downregulated (see Table 1).

Volcano plot analysis was performed to visualize the DEGs of the 3 datasets (GSE12649, GSE5388 and GSE5389) between the BD samples and the healthy samples. In the volcano plot, the DEGs between the BD and healthy samples were presented by all nodes. The nodes were significant when conforming to the cut-off criteria ($P < .05$ and fold-change $> 0.1$ or $< -0.1$) and were marked as green or red. The downregulated DEGs are presented by the green nodes, and the upregulated DEGs are presented by the red nodes. Volcano plots of GSE12649, GSE5388 and GSE5389 are shown in Figure 1A, 1B and 1C. Subsequently, the
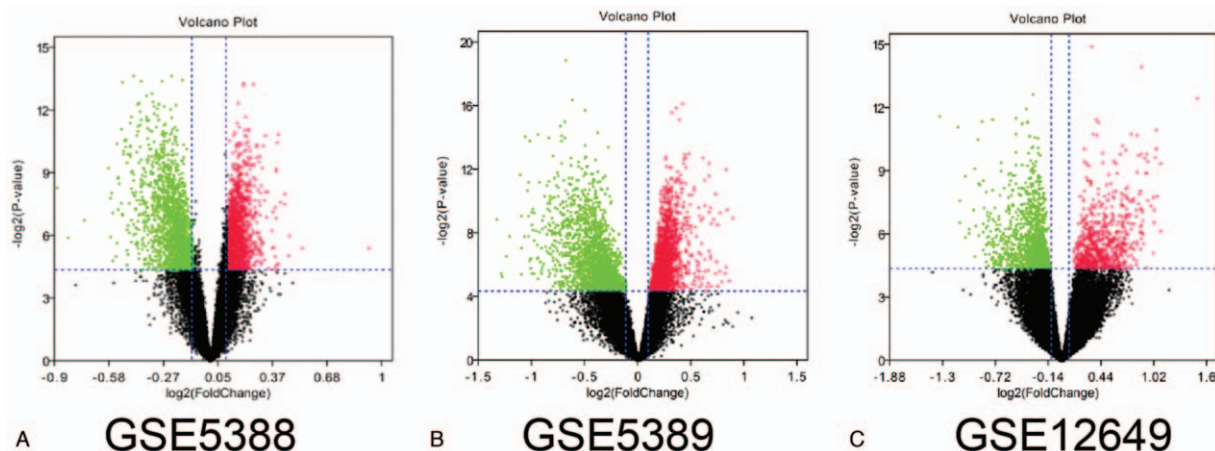


**Figure 1.** Identification of DEGs in brain samples between BD patients and healthy control individuals. The X-axis represents the fold change (log-scaled), and the Y-axis represents the *P* value (log-scaled). Each symbol represents a different gene. The red symbols show the upregulated genes; green symbols show the downregulated genes.
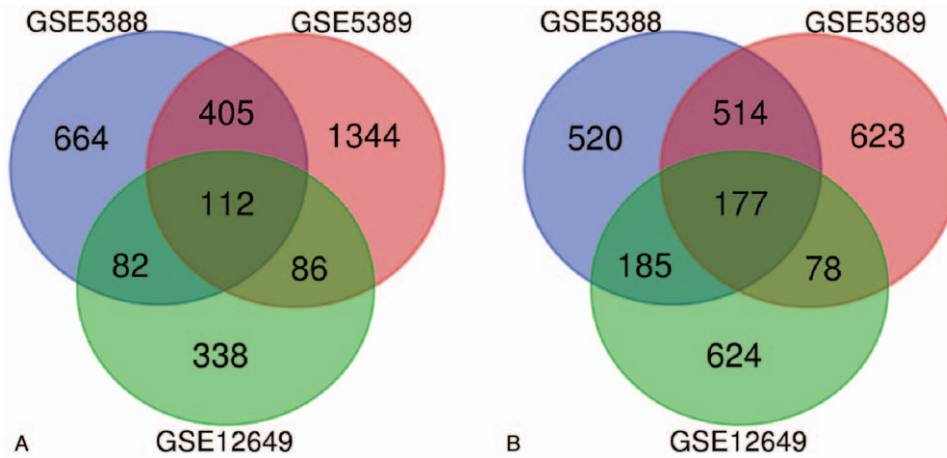
**Figure 2.** Venn diagram of DEGs from 3 GEO datasets. A. upregulated genes. B. downregulated genes. DEGs = differentially expressed genes, GEO = gene expression omnibus.

results of the intersection of these 3 DEG profiles were obtained by Venn diagram analysis. There were 289 DEGs in these 3 sets, as shown in Figure 2(A and B), among which 112 genes were upregulated and 177 genes were downregulated.

### 3.2. GO analysis and KEGG pathway analysis of DEGs

According to the inclusion criteria of $P < .05$ and gene counts $\geq 5$, 23 distinct GO terms and 3 significant KEGG pathways were

obtained to functionally classify the 289 DEGs. GO analysis of the upregulated DEGs and downregulated DEGs consisted of 3 items (BP, CC and MF), which are presented in Figure 3. These genes may play an important role in the development of BD.

Results of the GO analysis indicated that upregulated DEGs associated with BP were mainly enriched in negative regulation of growth; cellular response to zinc ions; cellular response to cadmium ions; peptidyl-tyrosine phosphorylation; negative
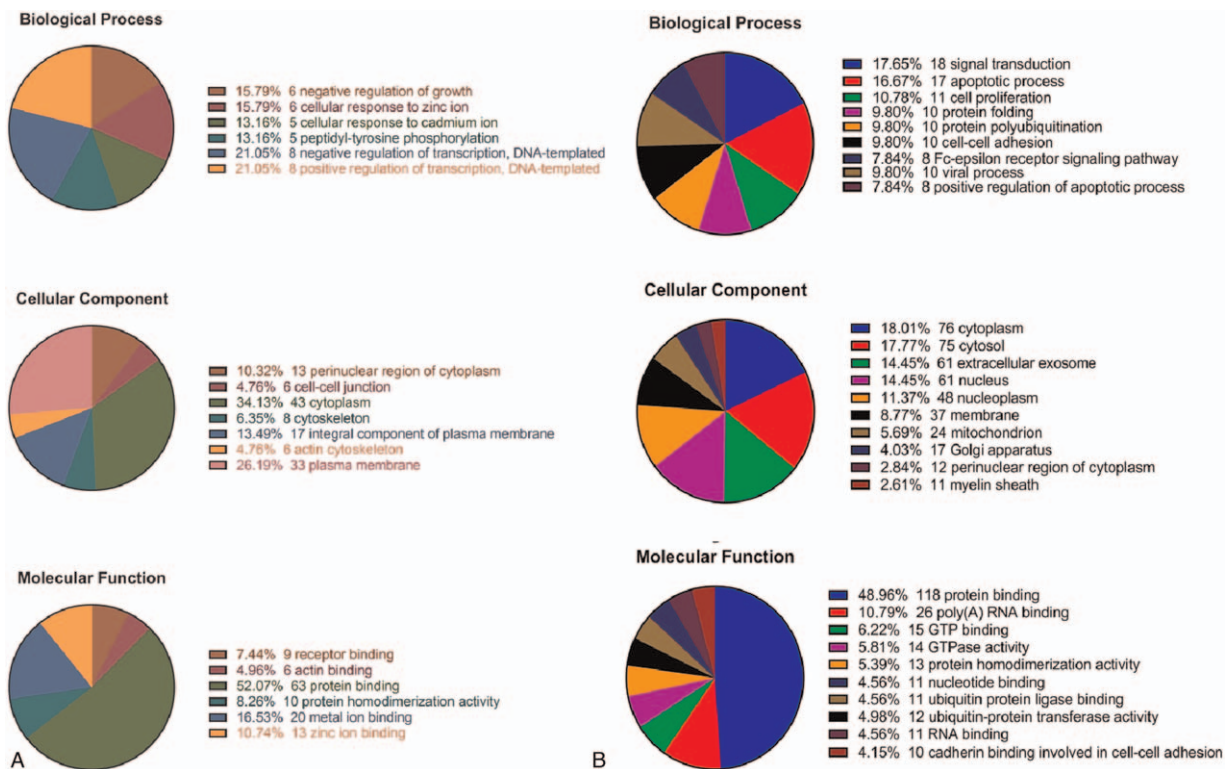


**Figure 3.** Significant GO terms (Biological process; Cellular component; Molecular function) associated with DEGs. The chart shows the annotations enriched by each GO term and their proportions. A. GO analysis of the upregulated DEGs. B. GO analysis of the downregulated DEGs. DEGs = differentially expressed genes, GO = Gene ontology.

**Table 2**

**KEGG pathway enrichment analysis of upregulated and downregulated DEGs.**

| Type | Category | Term | Description | Count | P Value | Genes |
|---|---|---|---|---|---|---|
| Upregulated DEGs | KEGG pathway | hsa04978 | Mineral absorption | 6 | 1.94E-05 | MT2A, MT1E, MT1H, MT1X, MT1G, MT1F |
| Downregulated DEGs | KEGG pathway | hsa04120 | Ubiquitin-mediated proteolysis | 9 | .00053 | UBE2N, UBE2D3, CUL5, UBE2D2, UBE3A, RCHY1, CDC27, CUL1, TRIP12 |
| Downregulated DEGs | KEGG pathway | hsa04141 | Protein processing in endoplasmic reticulum | 9 | .00208 | HSP90AB1, UBE2D3, HSP90AA1, UBE2D2, DNAJA1, PDIA6, EDEM3, CUL1, SEC61G |

DEGs = differentially expressed genes.

regulation of transcription, DNA-templated; and positive regulation of transcription, DNA-templated. Variations in upregulated DEGs associated with CC were significantly enriched in perinuclear region of cytoplasm, cell-cell junction, cytoplasm, cytoskeleton, integral component of plasma membrane, actin cytoskeleton, and plasma membrane. Upregulated DEGs for MF were significantly enriched in receptor binding, actin binding, protein binding, protein homodimerization activity, metal ion binding, and zinc ion binding.

Likewise, the results of the GO analysis indicated that downregulated DEGs linked with BP were mainly enriched in signal transduction, apoptotic process, cell proliferation, protein folding, protein polyubiquitination, cell-cell adhesion, viral process, Fc-epsilon receptor signaling pathway, and positive regulation of apoptotic process. Downregulated DEGs associated with CC were significantly enriched in cytoplasm, cytosol, extracellular exosome, nucleus, nucleoplasm, membrane, mitochondrion, Golgi apparatus, perinuclear region of cytoplasm, and myelin sheath. For MF, downregulated DEGs were significantly enriched in protein binding, poly(A) RNA binding,

GTP binding, GTPase activity, protein homodimerization activity, ubiquitin protein transferase activity, ubiquitin protein ligase binding, nucleotide binding, RNA binding, and cadherin binding involved in cell-cell adhesion.

In addition, the results of KEGG pathway analysis of the upregulated and downregulated DEGs are shown in Table 2. The upregulated DEGs were mainly enriched in pathways in mineral absorption, and the downregulated DEGs were mainly enriched in pathways in ubiquitin-mediated proteolysis and protein processing in the endoplasmic reticulum.

### 3.3. PPI network construction and identification of hub genes

Based on the online STRING and Cytoscape software (3.7.1), the PPI network was constructed, and hub genes were selected. The PPI network information involves 275 nodes and 728 edges, as presented in Figure 4. The top 10 genes evaluated by the degree of connectivity in the PPI network were identified, as presented in Figure 5. Detailed gene descriptions and connectivity degrees of
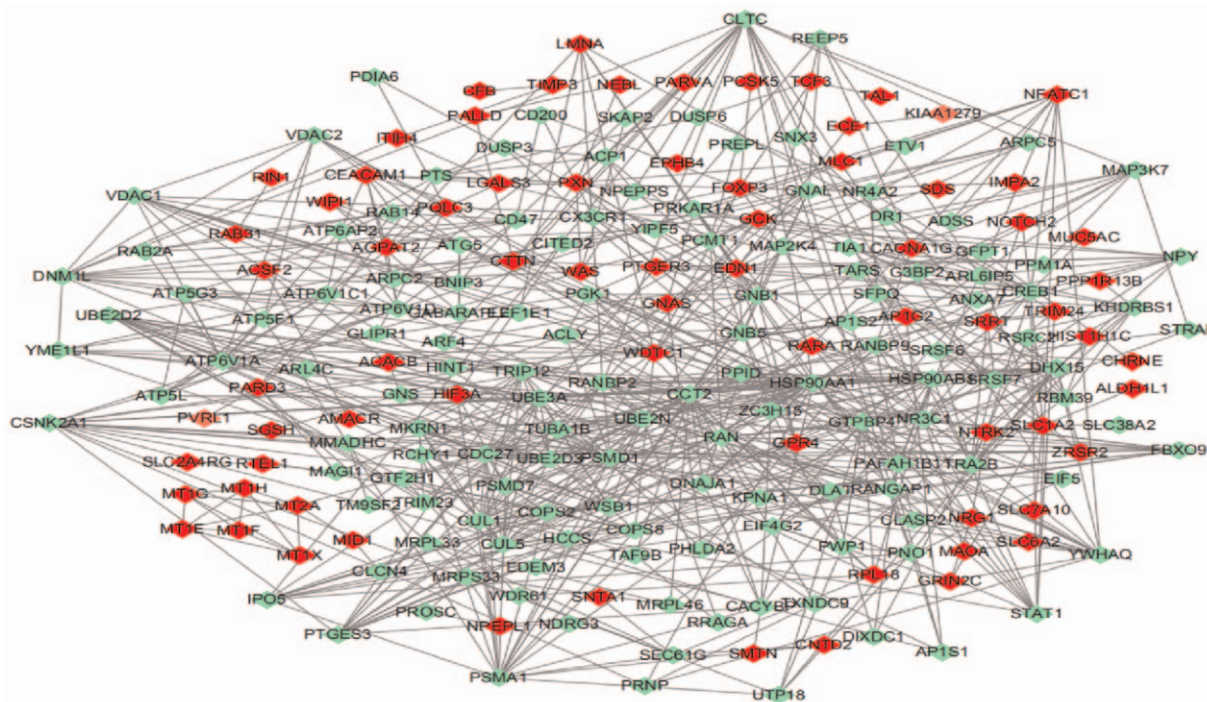


**Figure 4.** Protein-protein interaction network constructed with the DEGs. Red nodes represent upregulated genes, and green nodes represent downregulated genes.
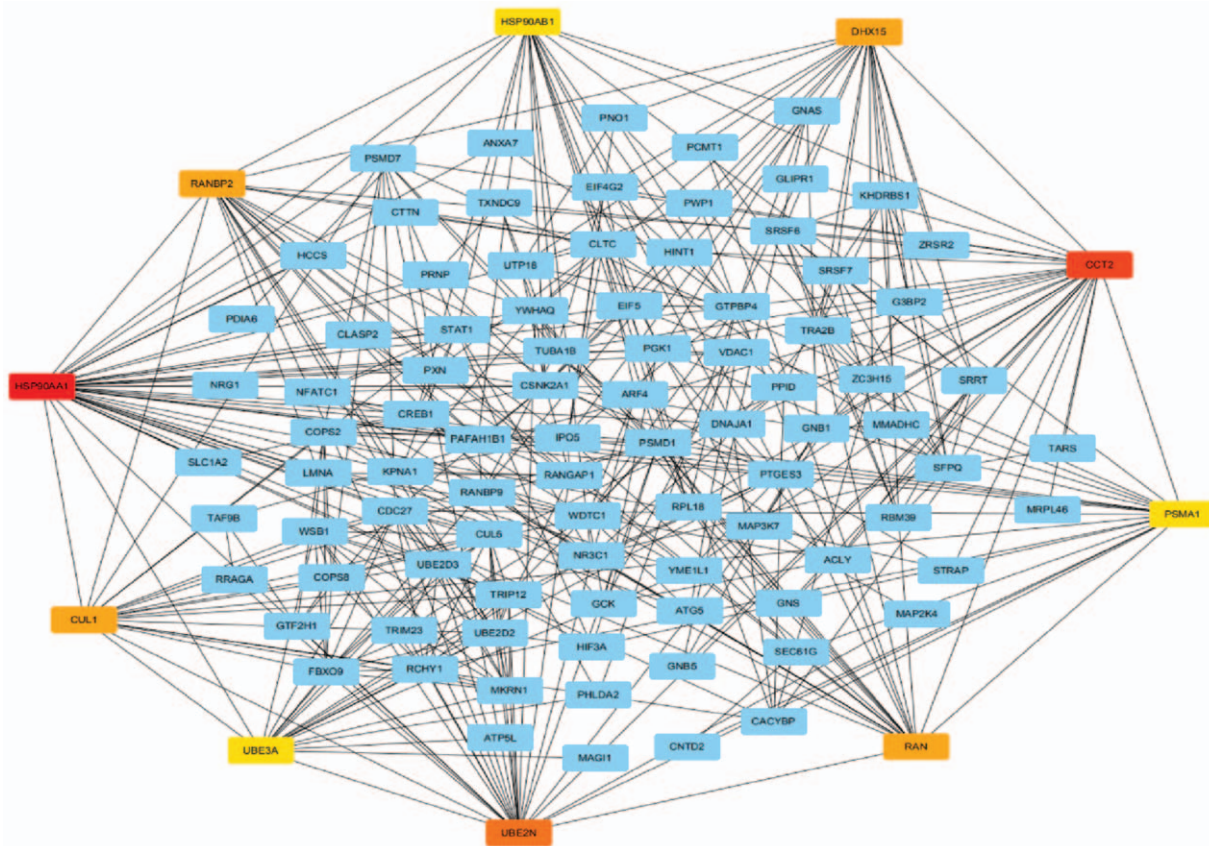
**Figure 5.** PPI network constructed with 10 hub genes and other DEGs. The darker a hub gene's color is, the heavier its weight across the network. DEGs = differentially expressed genes, PPI = protein-protein interaction.

the 10 hub genes are presented in Table 3. All of these hub genes were downregulated in BD samples.

### 3.4. Selection of more potential hub genes

To further elucidate more valuable hub genes from the initial 10 hub genes, KEGG pathway enrichment analysis was performed again ($P < .05$). The results showed that 3 genes (HSP90AA1, HSP90AB1, and CUL1) were enriched in the protein processing pathway in the endoplasmic reticulum, and 3 genes (UBE2N, UBE3A, and CUL1) were enriched in the ubiquitin-mediated proteolysis pathway (Table 4).

### 3.5. Exploration of the temporal expression levels of 5 hub genes in brain tissue from normal humans

The online software BrainCloud was utilized to obtain the expression levels of these 5 hub genes in the brain tissues of normal humans whose age ranged from 1 to 80 years old (Fig. 6) compared with those of BD patients. The expression level of HSP90AA1 showed a rapid decline in the period from 1 to 20 years old, regardless of whether the individual was male or female, and began to maintain a steady state after 20 years of age in males and females. The expression level of UBE3A gradually increased from 1 to 20 years of age in male

### Table 3

**Gene description and connectivity degree of 10 hub genes.**

| Gene symbol | Gene description | Degree | Type |
|---|---|---|---|
| HSP90AA1 | Heat shock protein 90 alpha family class A member 1 | 40 | Downregulation |
| CCT2 | Chaperonin containing TCP1 subunit 2 | 27 | Downregulation |
| UBE2N | Ubiquitin-conjugating enzyme E2 N | 26 | Downregulation |
| RANBP2 | RAN-binding protein 2 | 20 | Downregulation |
| RAN | RAN, member of the RAS oncogene family | 20 | Downregulation |
| CUL1 | Cullin 1 | 20 | Downregulation |
| DHX15 | DEAH-box helicase 15 | 20 | Downregulation |
| PSMA1 | Proteasome subunit alpha 1 | 19 | Downregulation |
| HSP90AB1 | Heat shock protein 90 alpha family class B member 1 | 19 | Downregulation |
| UBE3A | Ubiquitin protein ligase E3A | 19 | Downregulation |

**Reanalysis of 10 hub genes via KEGG pathway enrichment.**

| Category | Term | Description | Count | P Value | Genes |
|----------|------|-------------|-------|---------|-------|
| KEGG pathway | hsa04120 | Ubiquitin-mediated proteolysis | 3 | .010191483 | UBE2N, UBE3A, CUL1 |
| KEGG pathway | hsa04141 | Protein processing in endoplasmic reticulum | 3 | .015242141 | HSP90AB1, HSP90AA1, CUL1 |

individuals, while its expression in females showed the opposite pattern. After the age of 20 years, the expression level of UBE3A in males increased steadily, while women maintained a steady trend. From 1 to 80 years old, the expression level of HSP90AB1 showed a steady decline in both males and females. From 1 to 80 years of age, the expression level of UBE2N maintained a stable trend in both males and females. From 1 to 80 years of age, the expression level of CUL1 in males showed a steady upward trend, while the opposite trend was observed in females.
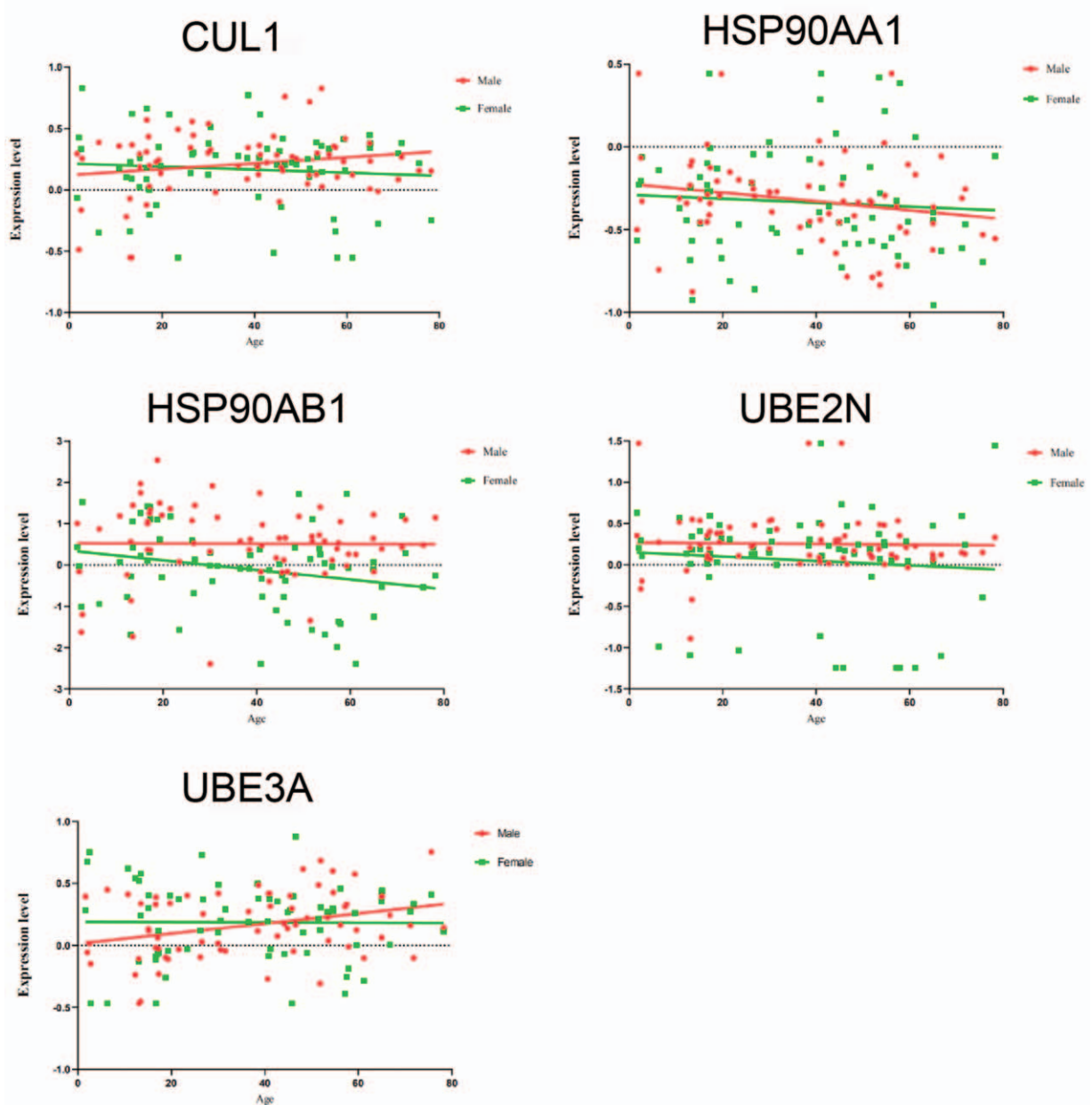


**Figure 6.** Expression of 5 genes (HSP90AA1, HSP90AB1, UBE2N, UBE3A, and CUL1) in the normal prefrontal cortex of males and females. Expression data were used for plotting from BrainCloud. Red represents male, and green represents female. The red trend line shows the gene expression of normal males from 1 to 80 years old. The green trend line shows the gene expression of females from 1 to 80 years old.
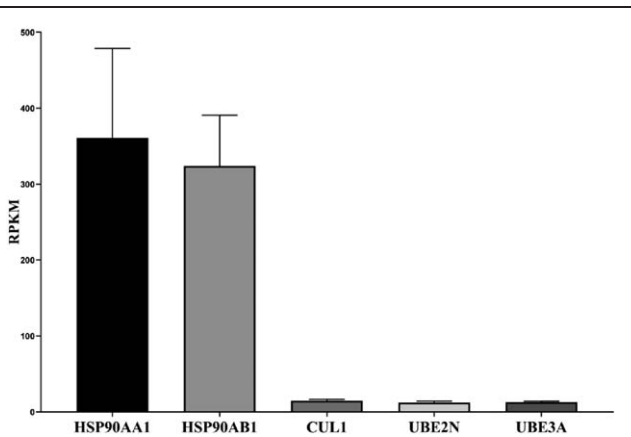
**Figure 7.** The expression levels of 5 hub genes in the same 3 normal brain samples (PMID: 24309898). CUL1: Cullin 1; HSP90AA1: Heat shock protein 90 alpha family class A member 1; HSP90AB1: Heat shock protein 90 alpha family class B member 1; UBE2N: Ubiquitin-conjugating enzyme E2 N; UBE3A: Ubiquitin protein ligase E3A; RPKM: Reads Per Kilobase Million.

### 3.6. Mining the expression levels of 5 hub genes in normal brain tissue based on the NCBI database

The NCBI database was used to mine the expression levels of 5 hub genes in normal brain tissue (https://www.ncbi.nlm.nih.gov/gene/). RNA-seq (quantitative transcriptomics analysis) was performed to detect the expression levels of HSP90AA1, HSP90AB1, CUL1, UBE2N, and UBE3A in 27 different tissues from 95 human individuals.[28] The expression level of HSP90AA1 ranked first among 27 tissues in 3 normal brain samples, with a mean Reads Per Kilobase Million (RPKM) of $360.601 \pm 96.334$; HSP90AB1 ranked second, with a mean RPKM of $324.029 \pm 54.479$; CUL1 ranked seventh with a mean RPKM of $14.51 \pm 1.734$; UBE2N ranked second with a mean RPKM of $12.165 \pm 1.724$; and UBE3A ranked third with a mean RPKM of $12.959 \pm 1.031$. Figure 8 displays the expression levels of 5 hub genes in the same 3 normal brain samples. As shown in Figure 7, HSP90AA1 was the most significantly expressed in healthy brain tissues among the 5 genes. However, according to the expression level of a single gene in 27 different types of tissues, these 5 genes were highly expressed in brain tissues compared to other tissues.

### 3.7. Diagnostic evaluation of HSP90AA1, HSP90AB1, CUL1, UBE2N, and UBE3A

Expression levels in BD samples were evaluated using ROC curves in order to illustrate the diagnostic value of the 5 hub genes (HSP90AA1, HSP90AB1, CUL1, UBE2N, and UBE3A). As displayed in Figure 8, the area under the curve (AUC) values for CUL1, HSP90AA1, HSP90AB1, UBE2N and UBE3A in BD patients and healthy controls determined for the GSE5388 dataset were 0.69 [95% confidence interval (CI), 0.5489–0.8210; $P = .0131$], 0.68 [95% CI, 0.542–0.818; $P = .0160$], .6129 [95% CI, 0.4698–0.7560; $P = .1298$], 0.6667 [95% CI, 0.5296–0.8038; $P = .0253$], and 0.7022 [95% CI, 0.5722–0.8321; $P = .0067$], respectively.

As presented in Figure 9, the AUC values for CUL1, HSP90AA1, HSP90AB1, UBE2N and UBE3A in BD patients and healthy controls determined for the GSE5389 dataset were 0.7818 [95% CI, 0.5641–0.9995; $P = .0290$], 0.9091 [95% CI,

0.7384–1.0000; $P = .0015$], 0.6545 [95% CI, 0.4137–0.8954; $P = .2313$], 0.6909 [95% CI, 0.4467–0.9351; $P = .1392$], and 0.8818 [95% CI, 0.7067–1.000; $P = .0031$], respectively.

As indicated in Figure 9, the AUC values for CUL1, HSP90AA1, HSP90AB1, UBE2N, and UBE3A in BD patients and healthy controls determined for the GSE12649 dataset were 0.6885 [95% CI, 0.5622–0.8148; $P = .0080$], 0.6729 [95% CI, 0.5426–0.8032; $P = .0150$], 0.5615 [95% CI, 0.4218–0.7012; $P = .3868$], 0.6052 [95% CI, 0.4689–0.7415; $P = .1389$], and 0.7103 [95% CI, 0.5873–0.8333; $P = .0031$], respectively. Combined with the results above, UBE3A, HSP90AA1, and CUL1 may be diagnostic genes for BD based on the inclusion criterion of $P < .05$ and AUC: 0.60–1.00.

As indicated in Figure 10, the AUC values for CUL1, HSP90AA1, HSP90AB1, UBE2N and UBE3A in BD patients and healthy controls determined for the GSE12649 dataset were 0.6885 [95% CI, 0.5622–0.8148; $P = .0080$], 0.6729 [95% CI, 0.5426–0.8032; $P = .0150$], 0.5615 [95% CI, 0.4218–0.7012; $P = .3868$], 0.6052 [95% CI, 0.4689–0.7415; $P = .1389$], and 0.7103 [95% CI, 0.5873–0.8333; $P = .0031$], respectively. Combined with the results above, UBE3A, HSP90AA1, and CUL1 may be diagnostic molecular markers for BD based on the inclusion criteria of $P < .05$ and AUC: 0.60–1.00.

## 4. Discussion

BD is the leading cause of disability worldwide, with unsatisfactory treatment and an unclear biological basis.[29] To date, drug therapy is still widely accepted as the most rapidly efficacious method to treat affective episodes and their most disturbing symptoms (agitation in mania), contributing to substantial reductions in the length of hospital stay and the rehospitalization rate.[30] However, the initial misdiagnosis of BD, mainly attributed to the presence of symptoms and the overlap of clinical features with other mental disorders, has been a great obstacle to effective and timely therapy.[31] Therefore, there is an urgent need to identify new diagnostic biomarkers and seek sensitive drugs to target signaling pathways for BD to develop effective and robust therapeutic strategies.

Currently, bioinformatics prediction and computer technology have some beneficial applications for all aspects of biomedical analysis.[32] However, there are few systematic bioinformatics analyses for mental diseases because the gene expression profiles associated with psychosis, which may be significantly different from those associated with other diseases, have not yet been distinguished. In the present study, ChIP analysis was first used to screen out DEGs in BD brain tissues, providing a novel strategy to exploit potential genes to provide insight into the pathogenesis of BD. The results showed that 289 common DEGs were identified from 3 array sets, among which 112 were upregulated and 177 were downregulated. By integrated bioanalysis, 3 promising genes, UBE3A, HSP90AA1, and CUL1, were identified, and all exhibited downregulated expression. This strongly suggests that the ubiquitin-mediated proteolysis pathway and protein processing in the endoplasmic reticulum pathway are of great importance and may be related to the incidence of BD.

UBE3A, also known as ubiquitin protein ligase E3A, functions to conjugate ubiquitin groups to a unique set of proteins[33,34] and is responsible for the degradation of multiple proteins.[35] Previous studies strongly indicated that mutation of the UBE3A catalytic domain is sufficient to induce the development of Angelman syndrome, a severe neurological disorder character-
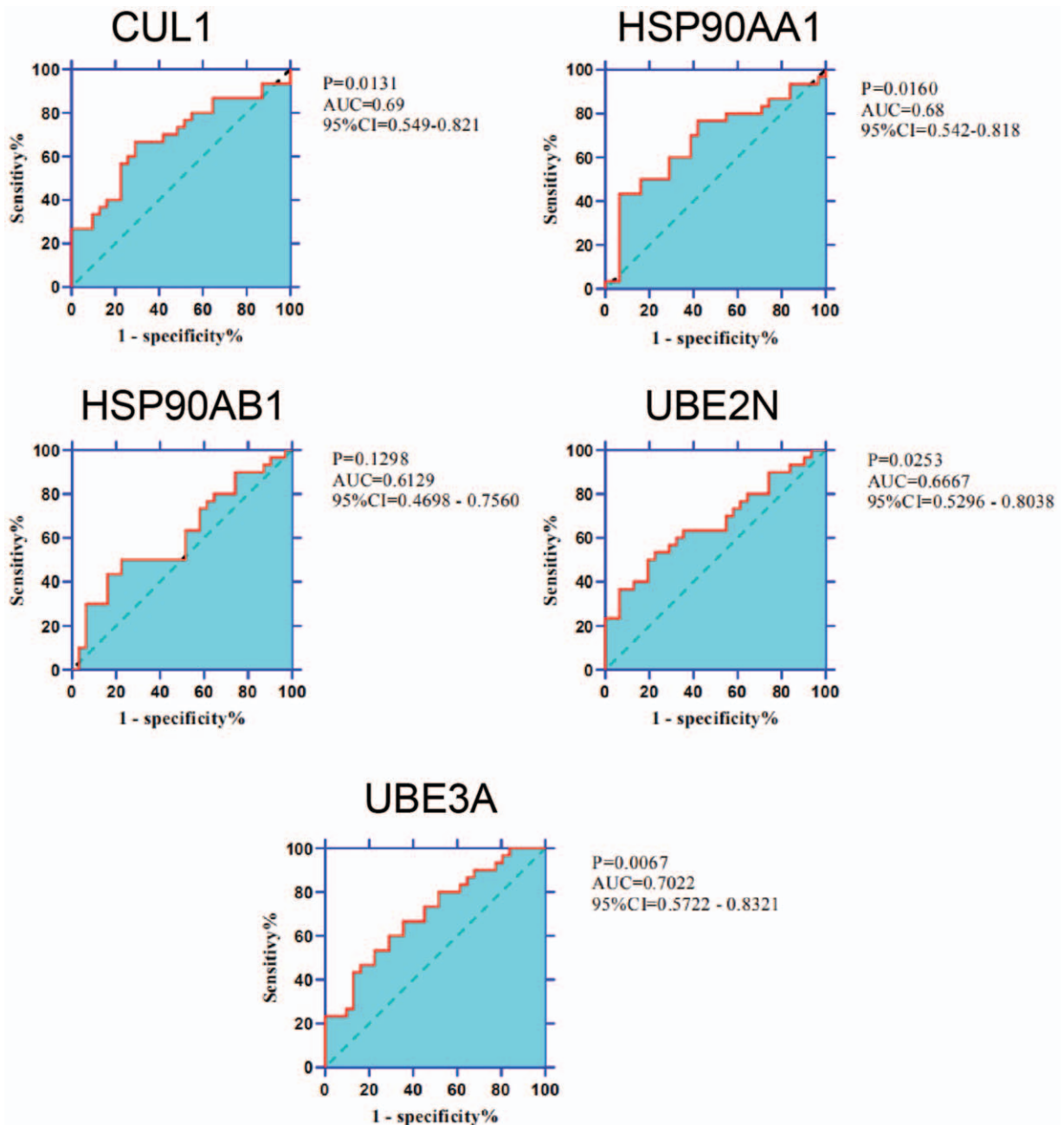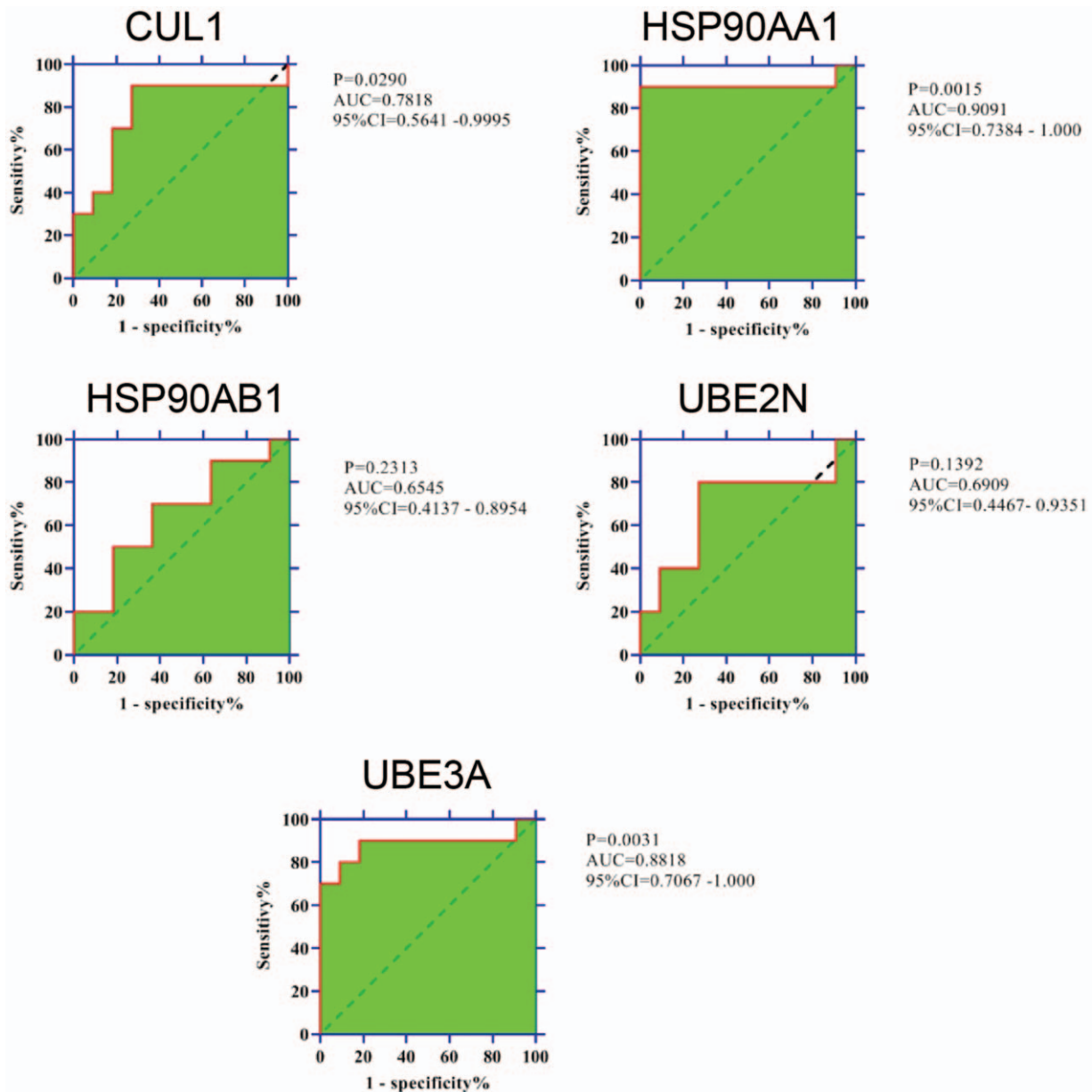
**Figure 8.** Diagnostic evaluation of 5 hub genes from the GSE5389 dataset. ROC curves and AUC values for the 5 hub genes (HSP90AA1, HSP90AB1, CUL1, UBE2N, and UBE3A). ROC: receiver operating characteristic; AUC: area under the ROC curve; CI, confidence interval; CUL1: Cullin 1; HSP90AA1: Heat shock protein 90 alpha family class A member 1; HSP90AB1: Heat shock protein 90 alpha family class B member 1; UBE2N: Ubiquitin-conjugating enzyme E2 N; UBE3A: Ubiquitin protein ligase E3A.

ized by mental retardation, absent speech, ataxia, seizures, and hyperactivity.[36–38] A study[39] revealed that maternal UBE3A deficiency in mice resulted in neuronal dysplasia and fewer branched apical dendrites. A report also indicated that mouse models with maternal deletions of UBE3A showed many Angelman-like phenotypes, including learning and memory deficits, motor phenotypes and seizures.[40] Some of these related phenotypes of Angelman syndrome could be seen in other neurodevelopmental disorders,[41] such as autism spectrum disorders (ASDs). Moreover, duplication of chromosomal regions containing UBE3A is linked with ASDs.[42] Coincidentally, the most prominent comorbidities of ASDs are BD, seizures, and migraine.[43] Potent evidence supports the idea that these superficially distinct diseases share common genetic changes and pathways with 1 another.[44]

In this study, UBE3A maintained a high expression level from birth, but UBE3A in patients with BD patients was substantially downregulated relative to the healthy controls, contrary to the expression in the normal human prefrontal cortex from the BrainCloud data (Fig. 7). Similarly, according to data from the
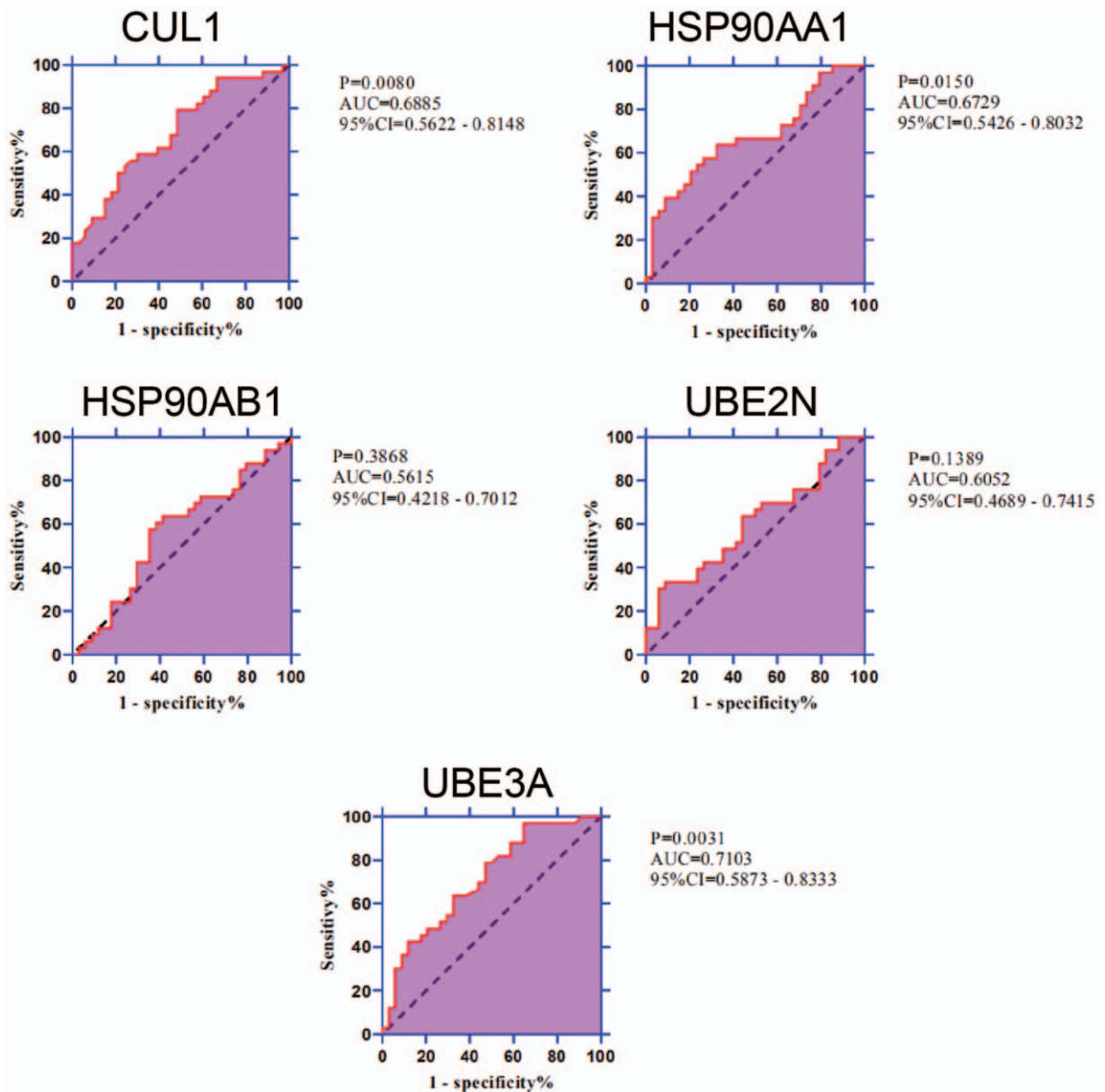
**Figure 9.** Diagnostic evaluation of 5 hub genes from the GSE5389 dataset. ROC curves and AUC values for the 5 hub genes (HSP90AA1, HSP90AB1, CUL1, UBE2N, and UBE3A). ROC: receiver operating characteristic; AUC: area under the ROC curve; CI, confidence interval; CUL1: Cullin 1; HSP90AA1: Heat shock protein 90 alpha family class A member 1; HSP90AB1: Heat shock protein 90 alpha family class B member 1; UBE2N: Ubiquitin-conjugating enzyme E2 N; UBE3A: Ubiquitin protein ligase E3A.

NCBI database, UBE3A presents a trend of high expression in normal brain tissue, but not the testis or thyroid, compared with other tissues. In our study, the expression of UBE3A in the brain tissues of BD patients was downregulated. Therefore, it can be inferred that UBE3A probably plays a key synergistic role in the pathogenesis of BD. In addition, UBE3A exhibited a large AUC in the 3 array datasets (GSE5388, AUC: 0.7022; GSE5389, AUC: 0.8818; and GSE12649, AUC: 0.7103). Thus, UBE3A may serve as a potential biomarker for the diagnosis of BD with the help of integrated bioinformatics technology.

Cullin protein is a molecular scaffold that plays a pivotal role in ubiquitin-mediated posttranslational modification of cellular proteins. CUL1, as 1 of eight members of the mammalian cullin protein family, can assemble the multisubunit Cullin-Ring (a truly interesting new gene) E3 ubiquitin ligase complex.[45] Experiments in mice have confirmed that CUL1 plays an indispensable role in the cell cycle and embryogenesis.[46,47] Previous research[48] on *C. elegans* indicated that CUL1 is involved in germline apoptosis. Studies with model organisms such as Drosophila have shown that CUL1 participates in the cell cycle[49] and eye development.[50] In addition to these processes, CUL1 plays important roles in signal transduction, cell cycle progression, and ubiquitin dependence,[51] serving as a skeleton of the Skp1-CUL1/Rbx1-F-box protein ubiquitin E3 ligase complex.[45]

In summary, since UBE3A is anticipated to be involved in the pathogenesis of BD, we also propose that CUL1 may be

**Figure 10.** Diagnostic evaluation of 5 hub genes from the GSE12649 dataset. ROC curves and AUC values for the 5 hub genes (HSP90AA1, HSP90AB1, CUL1, UBE2N, and UBE3A). ROC: receiver operating characteristic; AUC: area under the ROC curve; CI, confidence interval; CUL1: Cullin 1; HSP90AA1: Heat shock protein 90 alpha family class A member 1; HSP90AB1: Heat shock protein 90 alpha family class B member 1; UBE2N: Ubiquitin-conjugating enzyme E2 N; UBE3A: Ubiquitin protein ligase E3A.

considered to play an indirect but key role in the mechanism of BD. The expression of CUL1 maintained a relatively high expression level after birth in normal brain tissue, as shown by data from the BrainCloud and NCBI databases. In addition, this gene is downregulated in BD patients. In combination with the information from 3 array sets (GSE5388, AUC: 0.69; GSE5389, AUC: 0.7818; and GSE12649, AUC: 0.6885), it is clear that CUL1 provides a novel direction for the diagnosis of BD.

HSP90AA1, known as heat shock protein 90 alpha family class A member 1, is encoded on the complementary strand of chromosome 14q32.33 and spans over 59 kbp. Several pseudogenes of HSP90AA1 exist throughout the human genome, located on chromosomes 3, 4, 11, and 14.[52,53] In the last 20 years, the overexpression of HSP90AA1 has emerged as an intriguing hallmark of cancers and is thought to have important regulatory roles in invasion and migration through extensive interactions with other family members.[54–56] A previous study[57] has shown that HSP90AA1 is highly expressed in hepatocellular carcinoma in patients with depression. One study identified HSP90AA1 as 1 of several hub genes (CAMK2A, HSP90AA1 and PLCG1) among 184 risk genes via genome-wide association studies and exome sequencing studies, but it was not implicated as a drug target.[58] A study using RNA-Seq and qPCR in 2 postmortem cohorts of 34 BD patients and 55 controls illustrated that HSP90AA1 was upregulated without diagnostic differences.[59]

Few studies have been performed to investigate HSP90AA1 among mental illnesses, and the expression level of HSP90AA1 was lower in the present 3 array sets, contrary to the results of previous studies. It is thought that posttranslational modifications (phosphorylation, acetylation, S-nitrosylation, oxidation, methylation, sumoylation and ubiquitination) have a great impact on Hsp90 function and regulation.[60] Therefore, we speculated that the variation in HSP90AA1 is mainly dependent on different types of diseases and impacts posttranslational modifications. In our study, not only datasets (GSE5388, AUC: 0.68; GSE5389, AUC: 0.9091; and GSE12649, AUC: 0.6729) but also BrainCloud (see Fig. 7) and NCBI databases (shown in Fig. 8) revealed that HSP90AA1 has a large AUC in the 3 arrays and a high expression level in normal brain tissues. Based on these findings, it has been strongly suggested that the downregulated expression of HSP90AA1 may be involved in the pathogenesis of BD.

However, this study has some obvious deficiencies. First, objectively speaking, our bioinformatics research on BD based on the GEO database was conducted in strict accordance with the routine bioinformatics strategy for tumors or cancer. In terms of methodology, the research idea is relatively clear. However, cancer and mental illness are 2 different diseases, and statistically speaking, if DEGs were selected in strict accordance with the inclusion criteria for tumors, the extraction of DEGs from this study could not be further analyzed. The log fold change value is almost always limited to 1 in the extensive literature on bioinformatics analysis of tumors, indicating a more statistically significant and convincing limit, which is generally accepted by the research community. In our study, we tried to download the expression data for a variety of mental diseases (such as schizophrenia, depression, and BD) in the same or different chip platforms of the GEO database. When the log fold change value was artificially assigned a range of 0.2 to 1, there were only a small number of DEGs. However, it is interesting to note that we can obtain the titer of similar tumor inclusion criteria for the next GO and KEGG analysis when the log fold change value is artificially set as 0.1. This is perhaps the greatest statistical flaw in our study; of course, this may be the reason why few studies have performed bioinformatics analysis on mental illness data from a single GEO database. Second, although we identified 3 differentially expressed susceptibility genes, further validation was not performed.

## 5. Conclusion

Although 3 hub genes [HSP90AA1, UBE3A, and CUL 1] that are tightly correlated with BD occurrence were first found here, mainly based on routine bioinformatics methods for cancer-related disease, the feasibility of applying this single GEO bioinformatics approach for mental illness is questionable, given the significant differences between mental illness and cancer-related disease.

## Acknowledgments

All authors are grateful to the participated person and their family members.

## Author contributions

Conceptualization: Yong Zeng, Zhaowei Teng.

## References

[1] Grande I, Berk M, Birmaher B, et al. Bipolar disorder. Lancet 2016;387:1561–72.

[2] Merikangas KR, Akiskal HS, Angst J, et al. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. Arch Gen Psychiatry 2007;64:543–52.

[3] Pini S, de Queiroz V, Pagnin D, et al. Prevalence and burden of bipolar disorders in European countries. Eur Neuropsychopharmacol 2005;15:425–34.

[4] Crump C, Sundquist K, Winkleby MA, et al. Comorbidities and mortality in bipolar disorder: a Swedish national cohort study. JAMA Psychiatry 2013;70:931–9.

[5] Plana-Ripoll O, Pedersen CB, Agerbo E, et al. A comprehensive analysis of mortality-related health metrics associated with mental disorders: a nationwide, register-based cohort study. Lancet 2019;394:1827–35.

[6] Kraemer HC. Determining gene moderation of environmental risk factors for a mental disorder: a "perfect storm" of methodological problems. Int J Methods Psychiatr Res 2012;21:185–94.

[7] Fergusson DM, Horwood LJ, Miller AL, et al. Life stress, 5-HTTLPR and mental disorder: findings from a 30-year longitudinal study. Br J Psychiatry 2011;198:129–35.

[8] Wang Q, Shelton RC, Dwivedi Y. Interaction between early-life stress and FKBP5 gene variants in major depressive disorder and post-traumatic stress disorder: a systematic review and meta-analysis. J Affect Disord 2018;225:422–8.

[9] Hunsberger JG, Chibane FL, Elkahloun AG, et al. Novel integrative genomic tool for interrogating lithium response in bipolar disorder. Transl Psychiatry 2015;5:e504.

[10] Segura AG, Mitjans M, Jiménez E, et al. Association of childhood trauma and genetic variability of CRH-BP and FKBP5 genes with suicidal behavior in bipolar patients. J Affect Disord 2019;255:15–22.

[11] Yu K, Cheung C, Leung M, et al. Are bipolar disorder and schizophrenia neuroanatomically distinct? An anatomical likelihood meta-analysis. Front Hum Neurosci 2010;4:189.

[12] Anderson D, Ardekani BA, Burdick KE, et al. Overlapping and distinct gray and white matter abnormalities in schizophrenia and bipolar I disorder. Bipolar Disord 2013;15:680–93.

[13] Quigley SJ, Scanlon C, Kilmartin L, et al. Volume and shape analysis of subcortical brain structures and ventricles in euthymic bipolar I disorder. Psychiatry Res 2015;233:324–30.

[14] Cousins DA, Moore PB, Watson S, et al. Pituitary volume and third ventricle width in euthymic patients with bipolar disorder. Psychoneuroendocrinology 2010;35:1074–81.

[15] Wise T, Radua J, Nortje G, et al. Voxel-based meta-analytical evidence of structural disconnectivity in major depression and bipolar disorder. Biol Psychiatry 2016;79:293–302.

[16] Coffman JA, Bornstein RA, Olson SC, et al. Cognitive impairment and cerebral structure by MRI in bipolar disorder. Biol Psychiatry 1990;27:1188–96.

[17] Colpo GD, Leboyer M, Dantzer R, et al. Immune-based strategies for mood disorders: facts and challenges. Expert Rev Neurother 2018;18:139–52.

[18] Pfaffenseller B, Fries GR, Wollenhaupt-Aguiar B, et al. Neurotrophins, inflammation and oxidative stress as illness activity biomarkers in bipolar disorder. Expert Rev Neurother 2013;13:827–42.

[19] Morris G, Walder K, McGee SL, et al. A model of the mitochondrial basis of bipolar disorder. Neurosci Biobehav Rev 2017;74.(Pt A):1–20.

[20] Rowland T, Perry BI, Upthegrove R, et al. Neurotrophins, cytokines, oxidative stress mediators and mood state in bipolar disorder: systematic review and meta-analyses. Br J Psychiatry 2018;213:514–25.

[21] Mora E, Portella MJ, Piñol-Ripoll G, et al. High BDNF serum levels are associated to good cognitive functioning in bipolar disorder. Eur Psychiatry 2019;60:97–107.

[22] Aas M, Djurovic S, Ueland T, et al. The relationship between physical activity, clinical and cognitive characteristics and BDNF mRNA levels in patients with severe mental disorders. World J Biol Psychiatry 2019;20:567–76.

[23] Gideons ES, Lin P-Y, Mahgoub M, et al. Chronic lithium treatment elicits its antimanic effects via BDNF-TrkB dependent synaptic downscaling. Elife 2017;6:e25480.

[24] Butz H, Patócs A. Brief summary of the most important molecular genetic methods (PCR, qPCR, Microarray, Next-Generation Sequencing, etc). Exp Suppl 2019;111:33–52.

[25] Yao Z, Liu C, Yu X, et al. Microarray profiling and coexpression network analysis of long noncoding rnas in adipose tissue of obesity-T2DM mouse. Obesity (Silver Spring) 2019;27:1644–51.

[26] Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res 2013;41(Database issue): D991–5.

[27] Colantuoni C, Lipska BK, Ye T, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature 2011;478:519–23.

[28] Fagerberg L, Hallström BM, Oksvold P, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics 2014;13:397–406.

[29] Harrison PJ, Geddes JR, Tunbridge EM. The emerging neurobiology of bipolar disorder. Focus (Am Psychiatr Publ) 2019;17:284–93.

[30] Azorin J-M, Simon N. Dopamine receptor partial agonists for the treatment of bipolar disorder. Drugs 2019;79:1657–77.

[31] Mitchell PB, Loo CK, Gould BM. Diagnosis and monitoring of bipolar disorder in general practice. Med J Aust 2010;193(S4):S10–3.

[32] Zou Q, Mrozek D, Ma Q, et al. Scalable data mining algorithms in computational biology and biomedicine. Biomed Res Int 2017;2017:5652041.

[33] Huang L, Kinnucan E, Wang G, et al. Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. Science 1999;286:1321–6.

[34] Scheffner M, Huibregtse JM, Vierstra RD, et al. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell 1993;75:495–505.

[35] Eletr ZM, Kuhlman B. Sequence determinants of E2-E6AP binding affinity and specificity. J Mol Biol 2007;369:419–28.

[36] Kishino T, Lalande M, Wagstaff J. UBE3A/E6-AP mutations cause Angelman syndrome. Nat Genet 1997;15:70–3.

[37] Matsuura T, Sutcliffe JS, Fang P, et al. De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. Nat Genet 1997;15:74–7.

[38] Cooper EM, Hudson AW, Amos J, et al. Biochemical analysis of Angelman syndrome-associated mutations in the E3 ubiquitin ligase E6-associated protein. J Biol Chem 2004;279:41208–17.

[39] Miao S, Chen R, Ye J, et al. The Angelman syndrome protein Ube3a is required for polarized dendrite morphogenesis in pyramidal neurons. J Neurosci 2013;33:327–33.

[40] Jiang YH, Armstrong D, Albrecht U, et al. Mutation of the Angelman ubiquitin ligase in mice causes increased cytoplasmic p53 and deficits of contextual learning and long-term potentiation. Neuron 1998;21:799–811.

[41] Tan W-H, Bird LM, Thibert RL, et al. If not Angelman, what is it? A review of Angelman-like syndromes. Am J Med Genet A 2014;164A:975–92.

[42] Sell GL, Margolis SS. From UBE3A to Angelman syndrome: a substrate perspective. Front Neurosci 2015;9:322.

[43] Gargus JJ. Genetic calcium signaling abnormalities in the central nervous system: seizures, migraine, and autism. Ann N Y Acad Sci 2009;1151:133–56.

[44] Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 2013;381:1371–9.

[45] Klok MD, Alt SR, Irurzun Lafitte AJM, et al. Decreased expression of mineralocorticoid receptor mRNA and its splice variants in postmortem brain regions of patients with major depressive disorder. J Psychiatr Res 2011;45:871–8.

[46] Dealy MJ, Nguyen KV, Lo J, et al. Loss of Cul1 results in early embryonic lethality and dysregulation of cyclin E. Nat Genet 1999;23:245–8.

[47] Wang Y, Penfold S, Tang X, et al. Deletion of the Cul1 gene in mice causes arrest in early embryogenesis and accumulation of cyclin E. Curr Biol 1999;9:1191–4.

[48] Gao MX, Liao EH, Yu B, et al. The SCF FSN-1 ubiquitin ligase controls germline apoptosis through CEP-1/p53 in C. elegans. Cell Death Differ 2008;15:1054–62.

[49] Moberg KH, Bell DW, Wahrer DC, et al. Archipelago regulates Cyclin E levels in Drosophila and is mutated in human cancer cell lines. Nature 2001;413:311–6.

[50] Ou C-Y, Lin Y-F, Chen Y-J, et al. Distinct protein degradation mechanisms mediated by Cul1 and Cul3 controlling Ci stability in Drosophila eye development. Genes Dev 2002;16:2403–14.

[51] Nakayama KI, Nakayama K. Ubiquitin ligases: cell-cycle control and cancer. Nat Rev Cancer 2006;6:369–81.

[52] Ozawa K, Murakami Y, Eki T, et al. Mapping of the gene family for human heat-shock protein 90 alpha to chromosomes 1, 4, 11, and 14. Genomics 1992;12:214–20.

[53] Chen B, Piel WH, Gui L, et al. The HSP90 family of genes in the human genome: insights into their divergence and evolution. Genomics 2005;86:627–37.

[54] Workman P, Burrows F, Neckers L, et al. Drugging the cancer chaperone HSP90: combinatorial therapeutic exploitation of oncogene addiction and tumor stress. Ann N Y Acad Sci 2007;1113:202–16.

[55] Yang Y, Rao R, Shen J, et al. Role of acetylation and extracellular location of heat shock protein 90alpha in tumor cell invasion. Cancer Res 2008;68:4833–42.

[56] Wong DS, Jay DG. Emerging Roles of Extracellular Hsp90 in Cancer. Adv Cancer Res 2016;129:141–63.

[57] Xiang X, You X-M, Li L-Q. Expression of HSP90AA1/HSPA8 in hepatocellular carcinoma patients with depression. Onco Targets Ther 2018;11:3013–23.

[58] Li H, Zhou D-S, Chang H, et al. Interactome Analyses implicated CAMK2A in the genetic predisposition and pharmacological mechanism of Bipolar Disorder. J Psychiatr Res 2019;115:165–75.

[59] Sinclair D, Fillman SG, Webster MJ, et al. Dysregulation of glucocorticoid receptor co-factors FKBP5, BAG1 and PTGES3 in prefrontal cortex in psychotic illness. Sci Rep 2013;3:3539.

[60] Zuehlke AD, Beebe K, Neckers L, et al. Regulation and function of the human HSP90AA1 gene. Gene 2015;570: