



## A novel multi-class classification model for schizophrenia, bipolar disorder and healthy controls using comprehensive transcriptomic data

Qingxia Yang<sup>a,\*</sup>, Yi Li<sup>b</sup>, Bo Li<sup>c</sup>, Yaguo Gong<sup>d,\*\*</sup>

<sup>a</sup> Department of Bioinformatics, Smart Health Big Data Analysis and Location Services Engineering Lab of Jiangsu Province, School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China

<sup>b</sup> Institute of Genomic Medicine, Wenzhou Medical University, Wenzhou, China

<sup>c</sup> College of Life Sciences, Chongqing Normal University, Chongqing, Chongqing, 401331, China

<sup>d</sup> School of Pharmacy, Macau University of Science and Technology, Macau, China

### ARTICLE INFO

#### Keywords:

Multi-class classification  
Schizophrenia  
Bipolar disorder  
Gene signature  
Partial least squares discriminant analysis

### ABSTRACT

Two common psychiatric disorders, schizophrenia (SCZ) and bipolar disorder (BP), confer lifelong disability and collectively affect 2% of the world population. Because the diagnosis of psychiatry is based only on symptoms, developing more effective methods for the diagnosis of psychiatric disorders is a major international public health priority. Furthermore, SCZ and BP overlap considerably in terms of symptoms and risk genes. Therefore, the clarity of the underlying etiology and pathology remains lacking for these two disorders. Although many studies have been conducted, a classification model with higher accuracy and consistency was found to still be necessary for accurate diagnoses of SCZ and BP. In this study, a comprehensive dataset was combined from five independent transcriptomic studies. This dataset comprised 120 patients with SCZ, 101 patients with BP, and 149 healthy subjects. The partial least squares discriminant analysis (PLS-DA) method was applied to identify the gene signature among multiple groups, and 341 differentially expressed genes (DEGs) were identified. Then, the disease relevance of these DEGs was systematically performed, including ( $\alpha$ ) the great disease relevance of the identified signature, ( $\beta$ ) the hub genes of the protein-protein interaction network playing a key role in psychiatric disorders, and ( $\gamma$ ) gene ontology terms and enriched pathways playing a key role in psychiatric disorders. Finally, a popular multi-class classifier, support vector machine (SVM), was applied to construct a novel multi-class classification model using the identified signature for SCZ and BP. Using the independent test sets, the classification capacity of this multi-class model was assessed, which showed this model had a strong classification ability.

### 1. Introduction

As severe psychiatric disorders, schizophrenia (SCZ) and bipolar disorder (BP) are associated with shortened lifespan and are the leading causes of disability worldwide [1]. Patients with these two psychiatric disorders have higher rates of chronic medical conditions and die at a younger age than the general population [2]. Moreover, SCZ and BP collectively affect 2% of the population of the world [3], and SCZ and BP are disorders of thoughts and emotions as well as showing alterations in cognitive and affective processing [4]. Currently, psychiatry is the last area of medicine where the disease is diagnosed based only on symptoms, because biomarkers to assist the diagnosis remain to be developed [5]. Developing more effective methods for diagnosis of SCZ and BP has

been reported to be a major international public health priority [6].

However, there is substantial overlap between these two psychiatric disorders based on the genetic and epidemiological studies [7,8], and heritability estimates range from 60 to 80% for both SCZ and BP [9]. Despite these two disorders sharing genetics and symptomatology, the diagnostic systems adhere to BP or SCZ as independent categorical entities based on the differentiated clinical presentation [10]. In particular, SCZ is considered a prototypical psychotic disorder, but BP is characterized by an episodic disease course and predominant mood symptoms [1]. Moreover, the clarity of the underlying etiology and pathology is still lacking for moving from clinical syndromes to diseases for these two disorders [11]. Because of the overlap and lack of clarity between SCZ and BP, how to discriminate these two psychiatric disorders has been a

\* Corresponding author.

\*\* Corresponding author.

E-mail addresses: [yangqx@njupt.edu.cn](mailto:yangqx@njupt.edu.cn) (Q. Yang), [gongyglab@gmail.com](mailto:gongyglab@gmail.com) (Y. Gong).

complicated question [12].

Therefore, the objective measures of the genetic differences between these two disorders would assist psychiatrists with obvious benefits to the efficiency of diagnosis [13,14]. These objective measures consist of gene signatures and biological processes, which are different in SCZ and BP patients. To seek objective measures, high-throughput transcriptomic analysis using microarrays has emerged as a powerful tool for detecting gene expression [15]. In many transcriptomic studies, the prefrontal cortex (PFC) regions have been applied to explore the objective molecular measures for the diagnosis of SCZ and BP [16].

Recently, objective measures for these two disorders have been revealed based on many transcriptomic studies. For example, a reduction in key oligodendrocyte- and myelin-related genes was discovered by Tkachev et al. [17] between SCZ and BP based on microarray data. In the study of de Baumont et al. [18], the innate immune response was dysregulated between SCZ and BP. And 28 differentially expressed genes (DEGs), including CASP4, TYROBP, CCR1, SERPINA1, CCR5, and C1QA, were identified to play a central role in the manifestation of diseases. It was reported that SCZ and BP were characterized by abnormal Wnt gene expression in the study of Hoseh et al. [15]. As reported by Fillman et al., some of the heterogeneity in SCZ and BP could be attributed to inflammation and stress interactions [19]. Although many studies on SCZ and BP have been conducted to search for DEGs and biological processes, a classification model with higher consistency remains to be constructed based on transcriptomic data.

In this work, a comprehensive dataset of SCZ and BP was first combined from five independent transcriptomic studies. This dataset was composed of 120 SCZ patients, 101 BP patients, and 149 healthy controls. The gene signatures among the three groups were identified applying the partial least squares-discriminant analysis (PLS-DA). As a result, 341 DEGs were discovered for differentiating SCZ, BP, and healthy subjects. Then, the disease relevance of the gene signature was systematically performed, including (α) the great disease relevance of the identified signature was proven, (β) the hub genes of the protein-protein interaction (PPI) network played a key role in psychiatric disorders, and (γ) GO (gene ontology) terms and KEGG (Kyoto

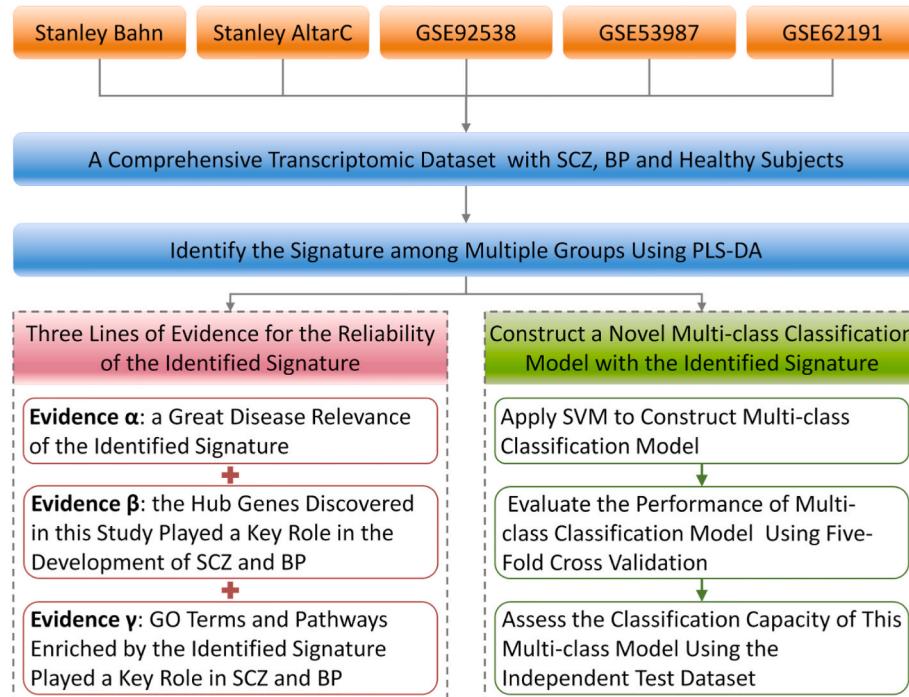
Encyclopedia of Genes and Genomes) pathways played a key role in psychiatric disorders. Finally, a novel multi-class classification model was constructed using support vector machine (SVM) based on the identified gene signature. The classification capacity of this model was assessed by the independent test set, which demonstrated a strong ability to classify different samples of SCZ, BP, and healthy controls.

## 2. Materials and methods

The detailed flowchart of this study is shown in Fig. 1, including (1) five microarray datasets were collected from multiple independent studies, (2) a comprehensive transcriptomic dataset was obtained by combining five independent datasets, (3) the gene signature among SCZ, BP and healthy subjects was identified using the PLS-DA method, (4) the disease relevance of the gene signature for psychiatric disorders was systematically performed, (5) a novel multi-class classification model was constructed using SVM based on the gene signature among multiple groups, and (6) the classification ability of this multi-class model was validated based on the independent test set.

### 2.1. Collection of transcriptomic datasets from the public databases

The prefrontal cortex (PFC) was widely regarded as the major region in many clinical studies for the dysfunction of psychiatric disorders. Herein, transcriptomic datasets were collected based on Brodmann areas (BAs) 9, 10, and 46 of PFC tissues in the brain. The keywords “schizophrenia” and “bipolar disorder” were used by searching popular databases, including GEO (Gene Expression Omnibus) and SMRI (Stanley Medical Research Institute). The studies would be collected if meeting the following criteria: (a) cDNA microarray technology was used for the gene expression profiling of *Homo sapiens*; (b) PFC tissues from BA9, BA10, or BA46 were analyzed for gene expression profiling; and (c) the collected dataset consisted of three groups: SCZ patients, BP patients, and healthy controls. In this study, five independent datasets were selected from the public databases. Each collected dataset included a cohort of SCZ patients, a cohort of BP patients, and a cohort of healthy



**Fig. 1.** The detailed flowchart of this study. SCZ: schizophrenia, BP: bipolar disorder, PLS-DA: partial least squares discriminant analysis, GO: gene ontology, SVM: support vector machine.

controls. The detailed information of these datasets (including dataset ID, number of samples, microarray platform, brain region, and the reference in the original publication) is provided in Table 1. All independent datasets were used to identify the gene signature and construct the classification model for SCZ and BP in this study.

## 2.2. Data preprocessing and combining multiple independent studies

The analytical process of data preprocessing and combining multiple studies was in strict conformity with the popular analysis of transcriptomics [20]. The combination of multiple independent datasets was performed using Rv3.4 (<https://www.r-project.org/>). First, the raw data (CEL file) was read, log-transformed, and normalized by the *affy* package, and the parameters of each step were set as default. Then, the probe sets of the genes were mapped to the gene names correspondingly for combination of all datasets [21,22]. If multiple probes could be mapped to the same gene name, the average value would be retained [23]. Herein, five independent studies were combined as a comprehensive dataset for SCZ and BP [24,25]. For this comprehensive dataset, the batch effects among different studies needed to be removed using a batch-effect correction method [26,27]. In this study, the *combat* function using an empirical Bayes framework in the *sva* package was applied to adjust the batch effects in the comprehensive dataset [28,29].

## 2.3. Identification of a gene signature to discriminate SCZ, BP, and healthy subjects

The confounding variables including age, gender, brain pH, post-mortem interval (PMI), and suicide should be considered when identifying gene signatures. To correct the confounding variables, the linear regression using the *lm* function was applied. Each gene was modeled using all confounding factors. After correcting these confounding variables, the gene signature was identified for SCZ, BP, and healthy subjects. Although various feature selection methods have been widely used for identifying gene signatures, most of these methods were designed only for two-class questions. Herein, the gene signature was needed to be discovered in three groups, including SCZ, BP, and healthy subjects. Therefore, partial least squares-discriminant analysis (PLS-DA), a popular feature selection method, was selected to identify the gene signature in multiple groups to differentiate different samples [30,31]. The PLS-DA can be applied for constructing predictive models as well as selecting discriminative variables [32,33]. It was considered a supervised version of PCA (principal component analysis) and applied for dimensionality reduction, feature selection, and classification [34]. The PLS-DA was used to identify features contributing to the differentiation

**Table 1**

Five independent transcriptomic datasets of psychiatric diseases (sorted by sample size). Each dataset contained one cohort of schizophrenia (SCZ), one cohort of bipolar disorder (BP), and one cohort of control (CTRL) samples.

ID	Sample Size (SCZ:BP:CTRL)	Platform	Tissue	Reference
Stanley Bahn GSE92538	99 (34:32:33)	HG-U133A	Frontal (BA46)	<i>BMC Genomics</i> . 7: 70, 2006.
	99 (31:12:56)	HG-U133 Plus 2	Frontal (BA9/46)	<i>PLoS One</i> . 13(7): e0200003, 2018.
GSE62191	88 (29:29:30)	Agilent-014850	Frontal (BA46)	<i>Schizophr Res</i> . 161(2-3):215-21, 2015.
GSE53987	51 (15:17:19)	HG-U133 Plus 2	Frontal (BA46)	<i>PLoS One</i> . 10(3): e0121744, 2015.
Stanley AltarC	33 (11:11:11)	HG-U133A	Frontal (BA10/46)	<i>BMC Genomics</i> . 7: 70, 2006.

in different groups [35]. Recently, this method has been extensively used for feature selection in genomics, transcriptomics, proteomics, metabolomics, and so on [36–38]. In this study, a comprehensive transcriptomic dataset was set as the input data of the PLS-DA for identifying gene signature. As a result, the gene signature among SCZ, BP, and healthy subjects were identified by the cutoff of the Variable Importance in the Projection (VIP > 1.5) in the PLS-DA model [39].

## 2.4. Assessment of the gene signature identified among the three groups

To assess the gene signature, three different measures ( $\alpha$ - $\gamma$ ) were conducted in this study. The three measures were very important for the assessment of the gene signature.

### (α) The level of disease relevance for the gene signature

For complex diseases (such as SCZ and BP), a substantial percentage of genes related with disease were expected to be contained in the gene signature identified in this study. However, because of the measurement variability, a certain percentage of irrelevant genes might be inevitable. In this study, the level of disease relevance for the gene signature was investigated by a comprehensive literature review. The percentage of disease-related genes in the whole gene signature was applied to represent the level of disease relevance.

### (β) The role of the hub genes of the protein-protein interaction network in psychiatric disorders

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database [40] was first applied to construct the PPI (protein-protein interaction) network. A high confidence level (>0.7) was set in the PPI network. Second, the gene signature discovered among SCZ, BP and healthy subjects could be mapped into the PPI network. Third, Cytoscape [41] was applied to visualize the interactions of different features. In this study, the hub genes of SCZ and BP were identified using the high interaction degree ( $\geq 5$ ) in the PPI network. The key role in the development of psychiatric disorders for these hub genes was validated by a comprehensive literature review.

### (γ) The role in psychiatric disorders for the GO terms and pathways in the enrichment analysis

In this study, enrichment analysis was performed based on the gene signature identified among SCZ, BP, and healthy subjects. The Gene Set Enrichment Analysis (GSEA) [42] was applied to perform the enrichment analysis using the hypergeometric test ( $p$  value < 0.05). The significantly overrepresented GO (Gene Ontology) terms and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways were discovered based on the gene signature identified among multiple groups. For these GO terms and KEGG pathways, the important role in the development of psychiatric disorders was revealed by conducting a comprehensive literature review.

## 2.5. A novel multi-class classification model constructed using the gene signature

A classification model with higher consistency is still necessary for accurate diagnoses of SCZ and BP [43]. In this study, the identified gene signature was applied to construct a novel multi-class classification model for SCZ and BP using the support vector machine (SVM). SVM is a supervised machine learning method, which has been a popular tool for the classification of omics data [44]. Compared to other methods, SVM is a powerful tool for recognizing subtle patterns for complex datasets. SVM was applied to identify the training vectors that best discriminated samples in different groups of the training set. The matrix of gene signatures identified based on the comprehensive dataset (Table 1) was set

as the training set in the SVM model to train the best classification model. Moreover, 5-fold cross-validation was used to construct the multi-class classification model. The performance of this multi-class classification model was evaluated by the ROC (receiver operating characteristic) curve and the AUC (area under the ROC curve) value. The AUC value, a measure of the classification ability, ranged from 0 to 1, and the higher the AUC was, the better the performance of the model at distinguishing different classes was.

### 2.6. Validation of the novel multi-class classification model using an independent test set

The classification performance of this novel multi-class model was measured by an independent test set. The dataset GSE87610 [45] was selected as the independent test set for validating the classification model in this study. For this independent set, there were 19 SCZ, 19 BP, and 19 healthy samples of the prefrontal cortex tissues. There were two subsets detected in layer 3 and layer 5 pyramidal neurons from this independent dataset GSE87610. In the SVM model, the training set was the comprehensive dataset (Table 1), and the test sets were the two independent sets from GSE87610. The performance of the multi-class classification model could be accurately validated by these two independent test sets. Therefore, this multi-class model based on machine learning was constructed and validated for psychiatric disorders.

## 3. Results and discussion

### 3.1. Comprehensive transcriptomic data by combining five independent studies

In the beginning, five independent datasets were collected from the public datasets (Table 1). There were two datasets from the Stanley Medical Research Institute (SMRI) database [46], including 99 (34 SCZ, 32 BP, and 33 healthy subjects) and 33 (11 SCZ, 11 BP, and 11 healthy subjects) samples based on the HG-U133A platform, which was labeled Stanley Bahn and Stanley AltarC, respectively. In datasets GSE92538 [47] and GSE53987 [48], there were 99 (31 SCZ, 12 BP, and 56 healthy subjects) and 51 (15 SCZ, 17 BP, and 19 healthy subjects) samples based on the HG-U133 Plus 2 platform, respectively. There were 88 (29 SCZ, 29 BP, and 30 healthy subjects) samples based on the Agilent-014850 platform for the GSE62191 dataset [18]. Each dataset was analyzed after data preprocessing using the R language. After removing the batch effects, these five independent datasets were combined as the most comprehensive data. This combined dataset contained 120 SCZ patients, 101 BP patients, and 149 healthy subjects. In total, there were 11,425 genes for all samples of the three groups in this comprehensive dataset.

### 3.2. Gene signature identified using the combined dataset to differentiate multiple groups

The information on confounding variables including age, gender, brain pH, PMI, and suicide is shown in Supplementary Table S1. There were obvious differences for three groups in age, gender, brain pH, PMI, and suicide using the student's *t*-test. As reported, brain pH was significantly lower in psychotic subjects compared to healthy controls [20]. Because these datasets were collected from public databases, the experimental design was not changed in this study. Moreover, the Pearson correlation coefficient ( $|r|$ ) was applied to assess the association for each gene. There was no gene with a correlation coefficient greater than 0.6, which is generally viewed as the cut-off for a meaningful correlation [49]. Moreover, linear regression was applied to correct these confounding variables, and the corrected matrix was used for identifying gene signatures. Using the comprehensive dataset, the gene signature was identified by the PLS-DA method to differentiate the samples of SCZ, BP, and healthy subjects.

As a result, 341 differentially expressed genes (DEGs) were identified

by the cutoff of VIP >1.5 using the PLS-DA method (Supplementary Fig. S1). The dysregulation of all DEGs between SCZ patients and healthy controls, between BP patients and healthy controls, as well as between SCZ and BP patients is shown in Supplementary Table S2. Detailed information on the top 15 dysregulated genes with higher VIP values is described in Table 2. To visualize the differences among multiple groups, the heatmap of the top 15 dysregulated genes is shown in Fig. 2. Specifically, the heatmap of the comprehensive dataset, the subset detected in layer 3 and layer 5 of the PFC from the independent dataset GSE87610 is displayed in Fig. 2A, B, and Fig. 2C, respectively. As shown in Fig. 3, the boxplots of the top 15 DEGs with higher VIP values were used to demonstrate the expression of genes directly. Taking the gene APOC1 with the highest VIP value (VIP = 2.42) as an example, the expression in the SCZ and BP patients was lower than the expression in the healthy controls, and the expression in SCZ patients was higher than the expression in BP patients. APOC1 gene polymorphisms were reported to be associated with cognitive impairment progression in patients with late-onset Alzheimer's disease [50].

### 3.3. Systematic assessment of the gene signature identified among three groups

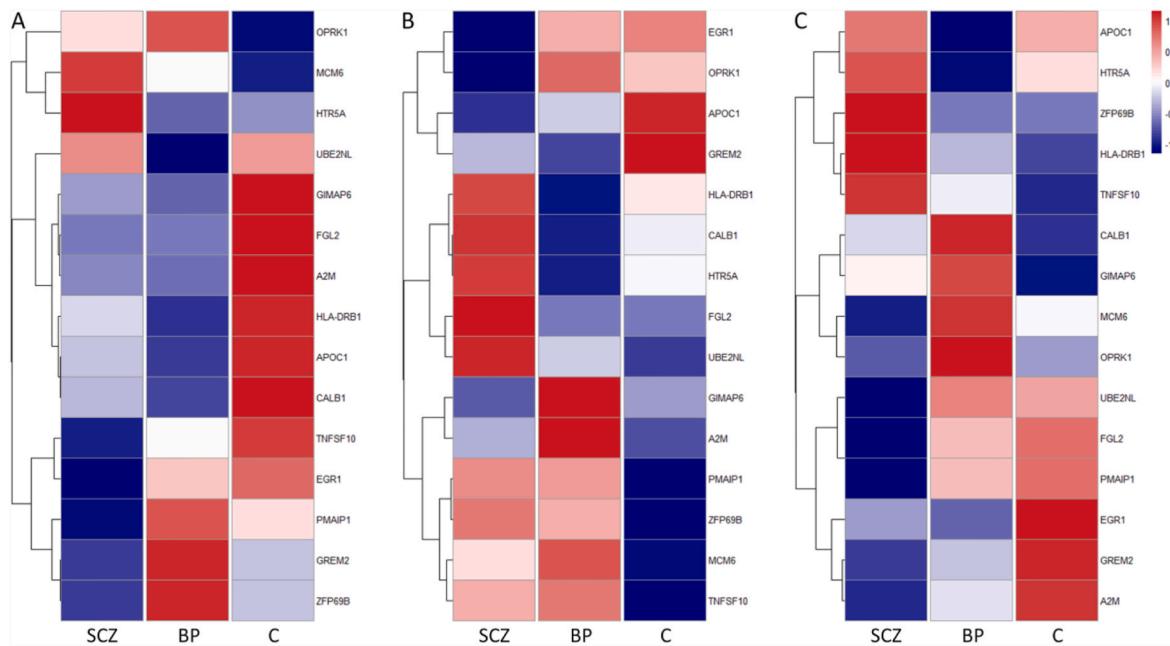
(a) The great disease relevance of the identified gene signature was proven

Herein, the disease relevance of the top 15 genes for differentiating multiple groups was verified through a literature review. The relationship between each gene and psychotic disorder (schizophrenia, bipolar disorder, or cognitive impairment) is shown in Supplementary Table S3. As a result, a great disease relevance (100%) of the top 15 DEGs was discovered. Among these gene signatures, APOC1 gene polymorphisms were reported to be associated with cognitive impairment progression [50]. MCM6 was decreased in SCZ patients [51]. The differential expression of CALB1 was revealed in the nucleus accumbens of SCZ patients [52]. The rhythmicity of FGL2 was discovered in healthy subjects but lost rhythmicity in SCZ patients [53]. GREM2 was the gene associated with BP [54]. rs1800883 in HTR5A was significantly associated with SCZ [55]. HTR5A could play an important role in the pathogenesis of BP [56]. These results revealed that there was a great disease relevance for the gene signature identified among SCZ patients, BP patients, and healthy controls.

**Table 2**

Detailed information on the top 15 dysregulated genes identified by PLS-DA (partial least squares discriminant analysis) with a cutoff of VIP >1.5. SCZ (schizophrenia), BP (bipolar disorder) and CTRL (healthy controls).

Order	Entrez ID	Symbol	VIP	Fold Change between Two Groups		
				SCZ vs. CTRL	BP vs. CTRL	SCZ vs. BP
1	341	APOC1	2.42	-0.09	-0.17	0.08
2	4175	MCM6	2.34	0.11	0.04	0.07
3	793	CALB1	2.26	-0.02	-0.07	0.06
4	10875	FGL2	2.23	-0.11	-0.14	0.03
5	64388	GREM2	2.20	-0.02	0.10	-0.12
6	3361	HTR5A	2.18	0.07	-0.03	0.11
7	3123	HLA-	2.17	-0.14	-0.29	0.14
		DRB1				
8	5366	PMAIP1	2.13	-0.11	0.04	-0.15
9	474344	GIMAP6	2.12	-0.12	-0.14	0.02
10	8743	TNFSF10	2.12	-0.18	-0.11	-0.07
11	2	A2M	2.12	-0.19	-0.18	0.00
12	1958	EGR1	2.11	-0.40	-0.09	-0.31
13	4986	OPRK1	2.02	0.04	0.07	-0.03
14	389898	UBE2NL	2.01	-0.03	-0.12	0.09
15	65243	ZFP69B	1.99	-0.05	0.08	-0.13



**Fig. 2.** Heatmap of the top 15 dysregulated genes identified by partial least squares discriminant analysis (PLS-DA) in the three groups. (A) Heatmap of the comprehensive dataset (Table 1). (B) Heatmap of the subset detected in layer 3 of the prefrontal cortex from this independent dataset GSE87610. (C) Heatmap of the subset detected in layer 5 of the prefrontal cortex from GSE87610. SCZ (schizophrenia), BP (bipolar disorder) and C (control samples).

(β) The hub genes in the PPI network played a key role in psychiatric disorders

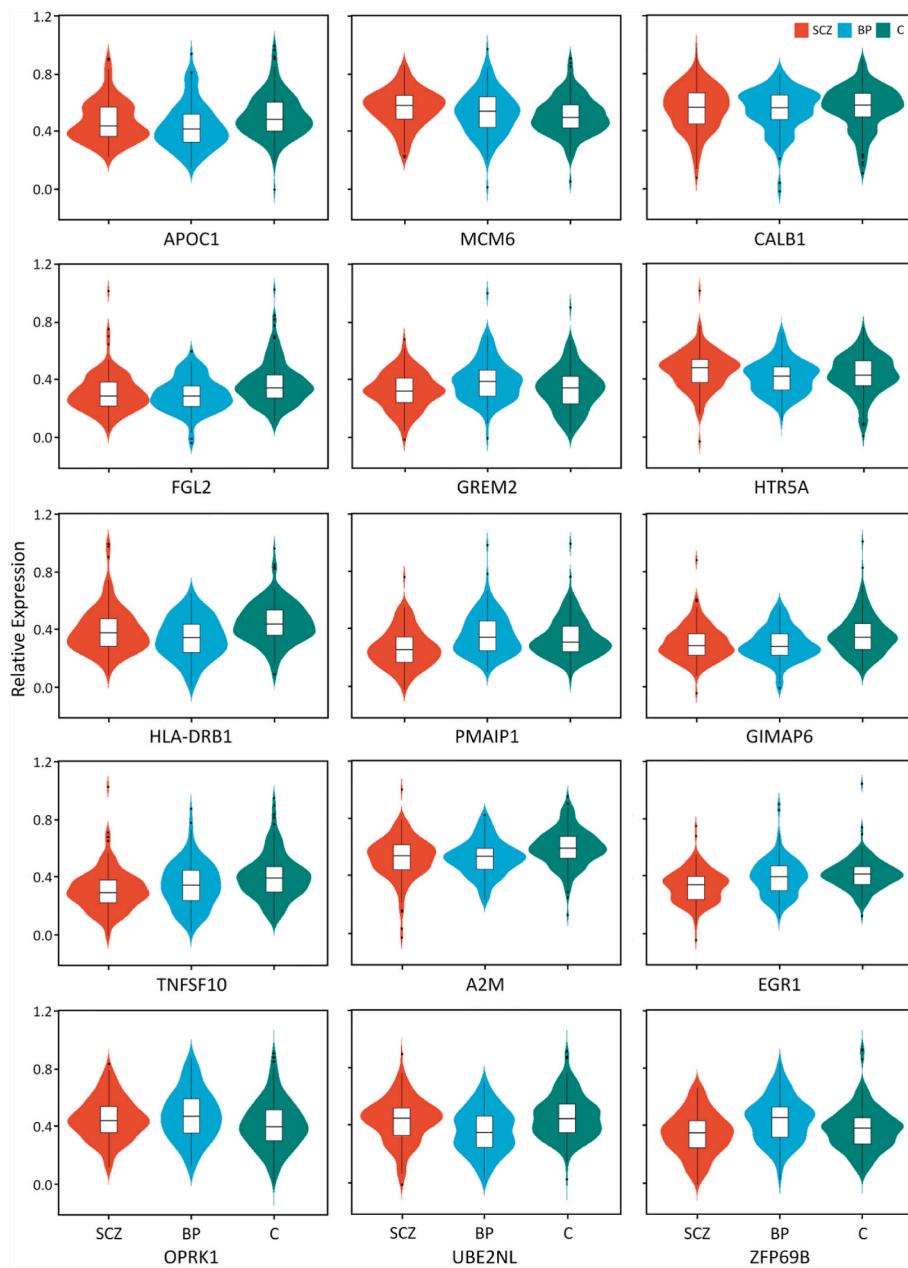
**Construction of PPI networks:** The PPI network was constructed based on the STRING database [40], and MCODE was applied for module analysis of this network [57]. The hub genes were identified and visualized using the *CytoHubba* [58] plugin of *Cytoscape* [41]. The PPI network was applied for identifying hub genes associated with psychiatric disorders in this study. As shown in Fig. 4, the PPI network was constructed using all 341 DEGs identified by the PLS-DA method. And the corresponding modules in this network are shown in Supplementary Fig. S2. The hub genes were ranked by the score value of the degree in this PPI network. The top six hub genes by degree (score value  $\geq 15$ ) included C3, DYNLL1, ADCY3, ITGB2, NMU, and BDKRB1.

**Validation of the hub genes:** As shown in Fig. 5, boxplots were used to validate the differential expression of the hub genes. There were significant changes among the different groups from these boxplots of the expression of DEGs. Based on a comprehensive literature review, the level of C3 in first-episode psychosis was reported to be higher than that in healthy controls [59], and the overactivation of the complement system resulted in neurotoxicity and pathological synapse loss, and progressive dysfunction [60]. Single markers and haplotypes in DYNLL1 were associated with SCZ and BP, and DYNLL1 was a relevant candidate in the pathogenesis of these disorders [61]. The activity of ADCY and GAL was reported to be interrelated, ADCY3 binds CACNA1C, and an SNP (rs1006737) within CACNA1C was associated with BP and SCZ [62]. ITGB2 was identified as one of the modules most highly associated with Alzheimer's disease [63]. NMU is a secreted neuropeptide in the brain, which acts as a neuronal-derived regulator in diverse physiologic processes [64,65]. The antipsychotic drug treatment affected the level of BDKRB1 in blood leukocytes in a first-episode psychosis study [66,67]. From these results, the hub genes identified in the PPI network played a key role in psychiatric disorders, which strongly indicates the reliability of the gene signature identified in this study.

(γ) The enriched GO terms and KEGG pathways played a key role in psychiatric disorders

Using 341 DEGs identified by the cutoff of VIP  $> 1.5$  in PLS-DA model among SCZ, BP, and healthy subjects, GO terms and KEGG pathways were enriched in this study. As shown in Fig. 6A, enrichment analysis of GO terms was performed using gene signatures among multiple groups to identify psychiatric disease-related biological processes. For example, the differences between SCZ and BP were due to genes associated with response to stimulus, which played a key role in the communication of environmental factors to the cells [18]. The detailed information including gene count in overlap and FDR (False discovery rate)  $q$ -value for the top 100 biological process terms was shown in Supplementary Table S4. As demonstrated in Fig. 6B, the molecular functions were enriched using the gene signature among multiple groups. For example, the enzymes coined oligopeptidases could play roles as modulators of neuropeptidergic systems and were also implicated in neurogenesis, neurodevelopment, and brain formation [68]. The detailed information including gene count in overlap and FDR  $q$ -value for top 100 molecular function terms was shown in Supplementary Table S5. As demonstrated in Fig. 6C, several key cell components were enriched in this study. It was reported that the predicted target genes in SCZ were enriched for neurodevelopmental processes including neuron projection development [69]. The detailed information including gene count in overlap and FDR  $q$ -value for top 100 cell component terms was shown in Supplementary Table S6.

Moreover, 13 KEGG pathways were enriched by the gene signature identified by the cutoff of VIP  $> 1.5$  in PLS-DA model (as shown in Fig. 6D and Supplementary Table S7). These pathways included the p53 signaling pathway, neuroactive ligand-receptor interaction, complement and coagulation cascades, lysosome, intestinal immune network for IgA production, mitogen activated protein kinase (MAPK) signaling pathway, leukocyte transendothelial migration, and so on. As reported, the gene encoding p53 was susceptible to SCZ [70], and the p53 signaling pathway was also associated with BP [71]. The analysis of neuroactive ligand-receptor interaction pathways implicated that PGE2 was as a novel mediator of antipsychotic treatment response [72]. The complement and coagulation cascades were the most significant biological processes involved in these pathologies of SCZ and BP [73]. Lysosomal function had a central role in maintaining neuronal homeostasis and was linked to neurodegeneration [74]. The KEGG pathways



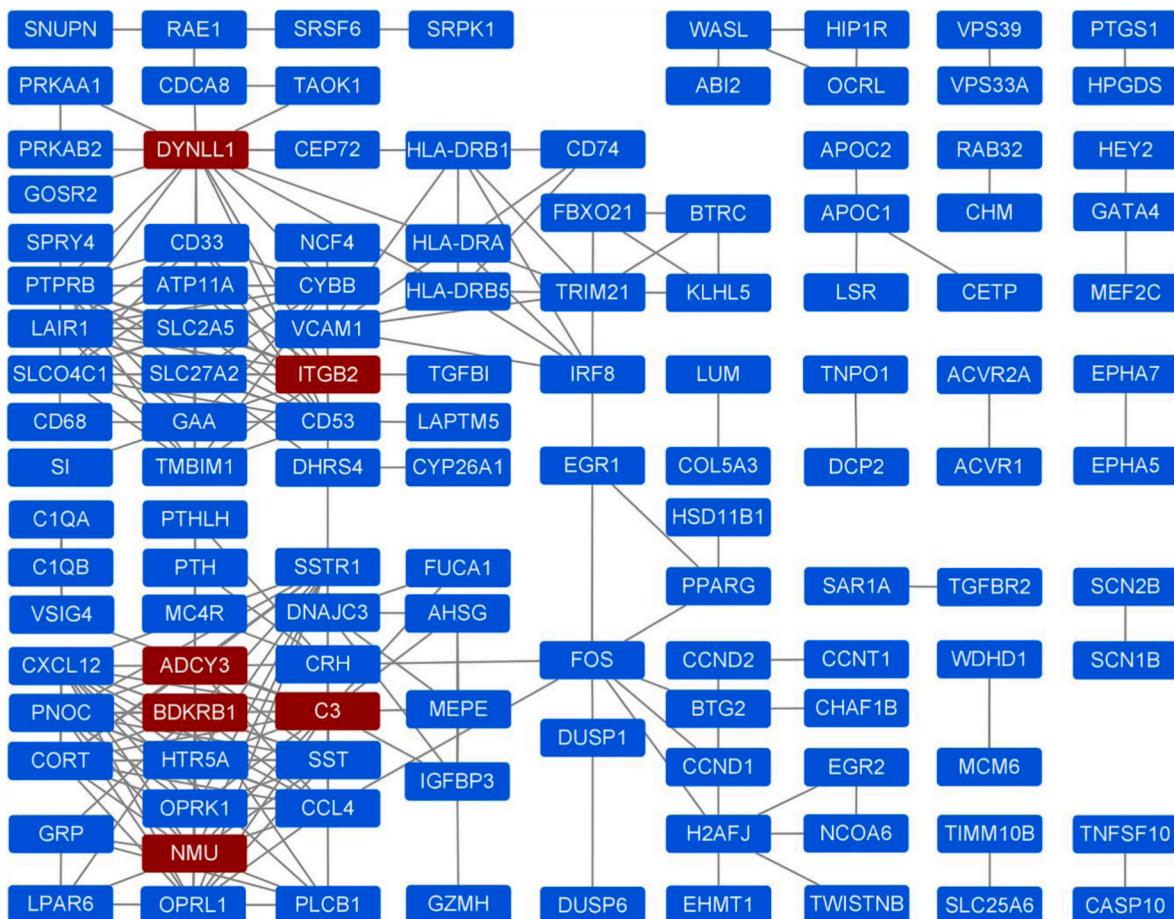
**Fig. 3.** The boxplots of the top 15 genes with the highest VIP values that were differentially expressed in multiple groups. Red, blue, and green indicate the SCZ (schizophrenia), BP (bipolar disorder), and C (healthy controls), respectively.

included several immune-related pathways (such as asthma and intestinal immune network for IgA production), and SCZ was confirmed to be a complex disease involving immune systems [75,76]. Abnormal MAPK signaling pathway activity was associated with the development of SCZ [77]. Lysosome and leukocyte transendothelial migration was enriched with the SCZ-associated miRNA signature [78]. According to this evidence, the GO terms and KEGG pathways enriched using gene signature in this study played a key role in psychiatric disorders.

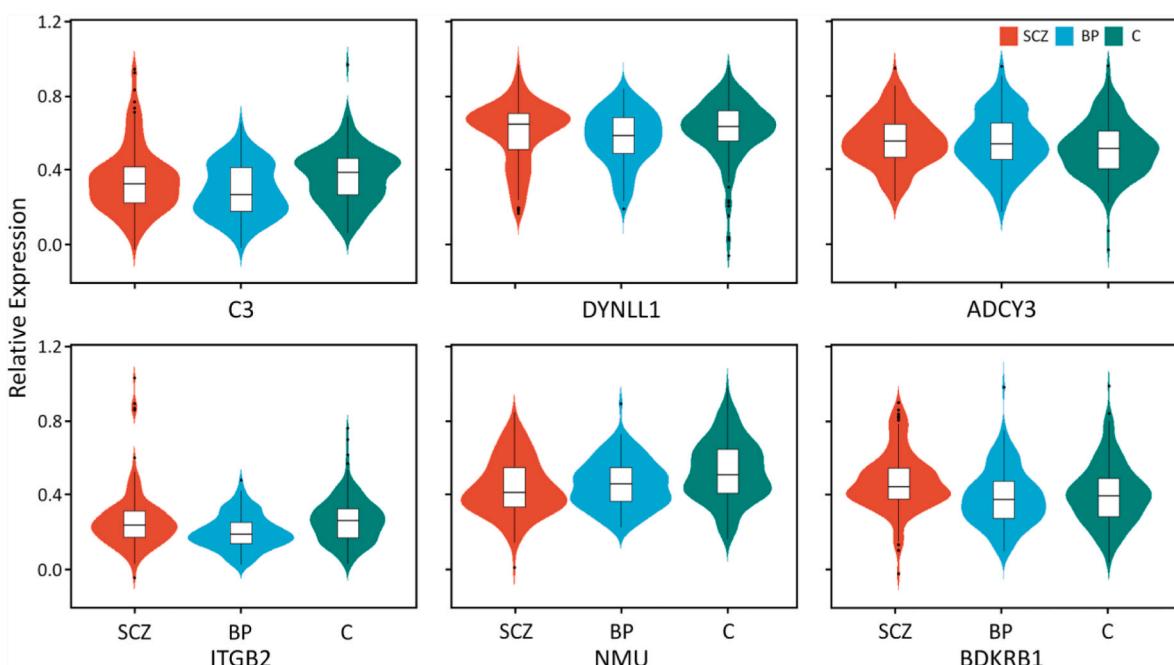
#### 3.4. Construction of the novel multi-class classification model for SCZ and BP

A novel multi-class classification model with the identified gene signature was constructed in this study. The gene signature identified from the comprehensive data (Table 1) was used to discriminate different samples of multiple groups (including SCZ, BP, and healthy subjects). First, the SVM method was applied to construct the model for

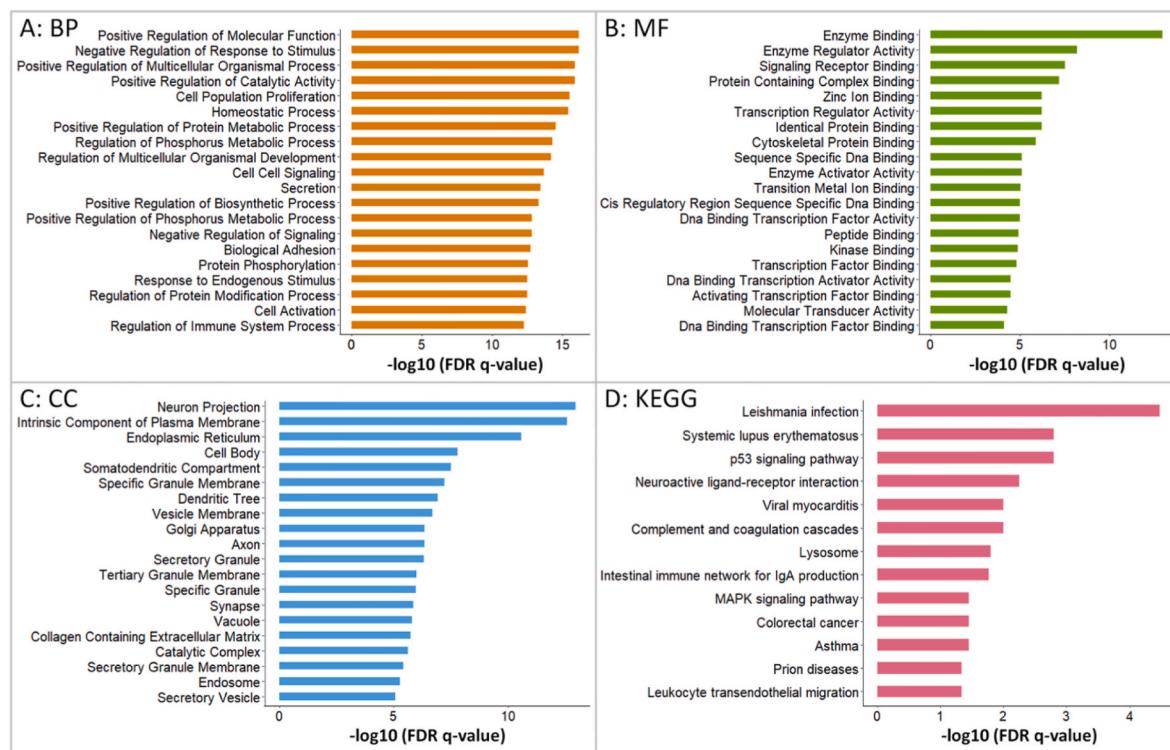
multi-class classification using the gene signature. Second, the classification performance of this multi-class model was evaluated by 5-fold cross-validation. Third, the AUC value was applied to quantify the performance of this classification model. SVM method was applied to construct the classification model for multiple groups and the AUC value could be calculated in this model. To classify multiple groups, the one-versus-rest approach was used in the classification model using the *svm* function of the *e1071* package. One-versus-rest can make binary classification algorithms capable of working as multi-class classification algorithms. This approach mainly split the multi-class data as binary classification data so that the binary classification algorithm can be applied. The only challenge of this method was that a model should be created for every class. In this study, three classes including groups of SCZ, BP, and healthy subjects required three models based on the comprehensive data. The *multi\_roc* function of the *multiROC* package was applied to calculate the specificity, sensitivity, and AUC value in multi-class classifications. The AUC value was obtained using the macro-



**Fig. 4.** The PPI network constructed in this study using the gene signature among schizophrenia, bipolar disorder, and healthy controls. The top 6 genes ranked by degree (score  $\geq 15$ ) in the network are marked in red.



**Fig. 5.** The boxplots of the six hub genes in the PPI (protein-protein interaction) network with the highest degree (score  $\geq 15$ ). Red, blue, and green indicate SCZ (schizophrenia), BP (bipolar disorder), and C (healthy controls), respectively.



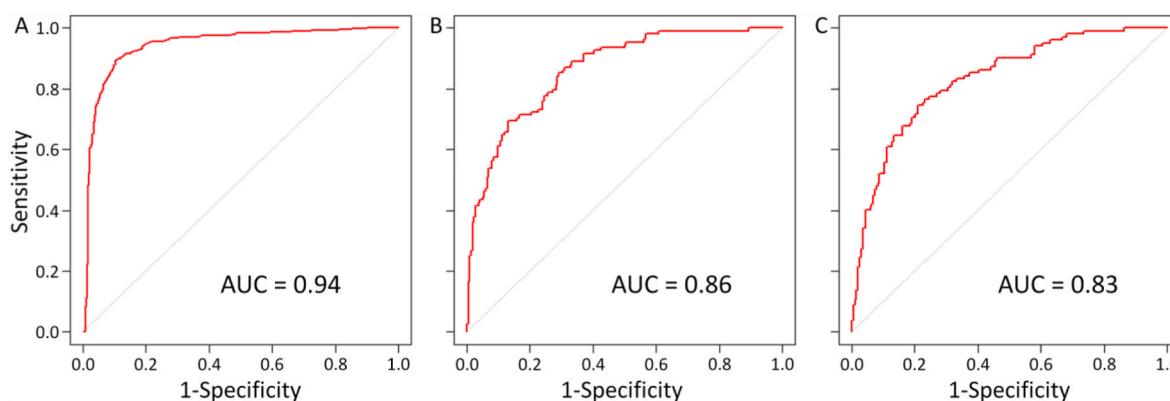
**Fig. 6.** The enrichment analysis based on the gene signature identified using PLS-DA among multiple groups (SCZ patients, BP patients, and healthy controls) in this study. The top 20 (A) biological process (BP) terms, (B) molecular function (MF) terms, and (C) cell component (CC) terms of gene ontology enrichment. (D) The 13 KEGG pathways enriched in this study.

average of AUC values of binary classifications by the one-versus-rest approach.

As shown in Fig. 7A, the AUC was 0.94 in the classification model using 5-fold cross-validation based on the comprehensive dataset (Table 1). Cross-validation was applied to evaluate the performance of machine learning models. Therefore, this classification model was validated as a model with high performance. The model of this study was constructed for classifying samples among multiple groups including SCZ, BP, and healthy subjects. The multi-class classification model added complexity to the modeling task and resulted in 0.94 prediction accuracy for the evaluation set. Therefore, this multi-class classification model was a model with high performance based on 5-fold cross-validation.

### 3.5. Validation of the multi-class classification model using the independent test set

To validate the capacity of this multi-class classification model, the comprehensive dataset (Table 1) was set as the training set, the independent data was as the test set to calculate prediction accuracy. The test sets consisted of two subsets from the independent dataset GSE87610. The first test set was detected from layer 3 of the PFC, and the second test set was detected from layer 5 of the PFC. As displayed in Fig. 7B and C, the AUC values were 0.86 and 0.83 for these two test sets, respectively. The high performance of the independent test sets could accurately reflect the ability of the classification model. Especially for the multi-class dataset, the performance of the classification model was good for discriminating the three groups. A high-performance classification model based on the gene signature was constructed and would be useful



**Fig. 7.** The multi-class classification model constructed for schizophrenia and bipolar disorder using the SVM algorithm. (A) The ROC curve and AUC value for the 5-fold cross validation using the comprehensive dataset. (B) The ROC curve and AUC value for the first independent test set detected in layer 3 of the prefrontal cortex from GSE87610. (C) The ROC curve and AUC value for the second independent test set detected in layer 5 of the prefrontal cortex from GSE87610.

for SCZ patients, BP patients, and healthy subjects. Because of the substantial overlap between SCZ and BP based on genetic studies, it was difficult to categorize these two psychiatric disorders. Therefore, the novel multi-class classification model might be able to discriminate different samples of these two psychiatric disorders from healthy controls.

There are still some limitations in this study. The first limitation is that only microarray datasets were used for identifying gene signatures for SCZ, BP, and healthy subjects. The second limitation is that because it was unable to obtain medication use information for all subjects, it was not able to incorporate this confounding variable into the analysis linear regression. The third limitation is that the prefrontal cortex was collected for constructing a classification model of psychiatric disorders. Although the prefrontal cortex has been widely used in clinical studies, a classification model based on post-mortem brain tissue was not useful for diagnostic purposes. The fourth limitation is that some overfitting would occur in the SVM classification model constructed based on DEGs because these DEGs were identified using the PLS-DA method and the full dataset was already used in this model. Because this study tried to construct a novel classification model for SCZ, BP, and healthy subjects, more improvements were still needed for a high-performance model. In the future, multi-omics data should be used for enhancing the robustness and consistency of gene signatures. Moreover, the biomarkers discovered from blood and urine samples will be more helpful for the clinical diagnosis.

#### 4. Conclusions

In this study, a comprehensive transcriptomic dataset was combined from five independent studies, comprising 120 SCZ patients, 101 BP patients, and 149 healthy controls. Using the PLS-DA method, 341 DEGs were discovered to differentiate the samples of the three groups. Then, the disease relevance of the gene signature was systematically performed. Finally, a novel multi-class classification model was constructed using the SVM method based on the identified gene signature. The classification capacity of this multi-class model was assessed by the independent test sets, which demonstrated a strong classification ability for this model.

#### Declaration of competing interest

None Declared.

#### Acknowledgements

This work was funded by the National Natural Science Foundation of Jiangsu (BK20210597), and the NUPTSF (Grant No. NY220169).

#### Abbreviations

SCZ	schizophrenia
BP	bipolar disorder
PLS-DA	partial least squares discriminant analysis
PFC	prefrontal cortex
DEGs	differentially expressed genes
GO	gene ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
SVM	support vector machine
BAs	Brodmann areas
GEO	Gene Expression Omnibus
SMRI	Stanley Medical Research Institute
PMI	postmortem interval
PCA	principal component analysis
VIP	Variable Importance in the Projection
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
PPI	protein-protein interaction

GSEA	Gene Set Enrichment Analysis
ROC	receiver operating characteristic
AUC	area under the ROC curve
FDR	false discovery rate
MAPK	mitogen activated protein kinase

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.combiomed.2022.105956>.

#### References

- [1] Schizophrenia Working Group of the Psychiatric Genomics, Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes, *Cell* 173 (2018) 1705–1715.
- [2] M. Fenger-Gron, C.H. Vestergaard, A.R. Ribe, S.P. Johnsen, L. Frost, A. Sandbaek, D.S. Davydow, Association between bipolar disorder or schizophrenia and oral anticoagulation use in Danish adults with incident or prevalent atrial fibrillation, *JAMA Netw. Open* 4 (2021), e2110096.
- [3] T.B. Bigdelli, A.H. Fanous, Y. Li, N. Rajeevan, F. Sayward, G. Genovese, R. Gupta, K. Radhakrishnan, A.K. Malhotra, N. Sun, Q. Lu, Y. Hu, B. Li, Q. Chen, S. Mane, P. Miller, K.H. Cheung, R.E. Gur, T.A. Greenwood, D.L. Braff, S. Consortium on the Genetics of, E.D. Achtyes, P.F. Buckley, M.A. Escamilla, D. Lehrer, D.P. Malaspina, S.A. McCarroll, M.H. Rapaport, M.P. Vawter, M.T. Pato, C.N. Pato, I. Genomic Psychiatry Cohort, H. Zhao, T.R. Kosten, M. Brophy, S. Pyarajan, Y. Shi, T. J. O'Leary, T. Gleason, R. Przygrodzki, S. Muralidhar, J.M. Gaziano, P. Million Veteran, G.D. Huang, J. Concato, L.J. Siever, M. Aslan, P.D. Harvey, Genome-wide association studies of schizophrenia and bipolar disorder in a diverse cohort of US veterans, *Schizophr. Bull.* 47 (2021) 517–529.
- [4] R.A. Power, S. Steinberg, G. Björnsdóttir, C.A. Rietveld, A. Abdellaoui, M. M. Nivard, M. Johannesson, T.E. Galesloot, J.J. Hottenga, G. Willemsen, D. Cesarini, D.J. Benjamin, P.K. Magnusson, F. Ullen, H. Tiemeier, A. Hofman, F. J. van Rooij, G.B. Walters, E. Sigurdsson, T.E. Thorgeirsson, A. Ingason, A. Helgason, A. Kong, L.A. Kiemeney, P. Koellinger, D.I. Boomsma, D. Gudbjartsson, H. Stefansson, K. Stefansson, Polygenic risk scores for schizophrenia and bipolar disorder predict creativity, *Nat. Neurosci.* 18 (2015) 953–955.
- [5] T. Wolfers, N.T. Doan, T. Kaufmann, D. Alnaes, T. Moberget, I. Agartz, J. K. Buitelaar, T. Ueland, I. Melle, B. Franke, O.A. Andreassen, C.F. Beckmann, L. T. Westlye, A.F. Marquand, Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models, *JAMA Psychiatr.* 75 (2018) 1146–1155.
- [6] M.J. Gandal, P. Zhang, E. Hadjimichael, R.L. Walker, C. Chen, S. Liu, H. Won, H. van Bakel, M. Varghese, Y. Wang, A.W. Shieh, J. Haney, S. Parhami, J. Belmont, M. Kim, P. Moran Losada, Z. Khan, J. Mleczko, Y. Xia, R. Dai, D. Wang, Y.T. Yang, M. Xu, K. Fish, P.R. Hof, J. Warrell, D. Fitzgerald, K. White, A.E. Jaffe, E.C. Psych, M.A. Peters, M. Gerstein, C. Liu, L.M. Iakoucheva, D. Pinto, D.H. Geschwind, Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder, *Science* 362 (2018), eaat8127.
- [7] Cross-Disorder Group of the Psychiatric Genomics Consortium, Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs, *Nat. Genet.* 45 (2013) 984–994.
- [8] P. Lichtenstein, B.H. Yip, C. Bjork, Y. Pawitan, T.D. Cannon, P.F. Sullivan, C. M. Hultman, Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study, *Lancet* 373 (2009) 234–239.
- [9] M.M. Nothen, V. Nieratschker, S. Cichon, M. Rietschel, New findings in the genetics of major psychoses, *Dialogues Clin. Neurosci.* 12 (2010) 85–93.
- [10] Y. Li, X. Li, J. Hong, Y. Wang, J. Fu, H. Yang, C. Yu, F. Li, J. Hu, W. Xue, Y. Jiang, Y. Chen, F. Zhu, Clinical trials, progression-speed differentiating features and swiftness rule of the innovative targets of first-in-class drugs, *Brief. Bioinform* 21 (2020) 649–662.
- [11] G.D. Pearlson, Etiologic, phenomenologic, and endophenotypic overlap of schizophrenia and bipolar disorder, *Annu. Rev. Clin. Psychol.* 11 (2015) 251–281.
- [12] H.G. Schnack, M. Nieuenhuis, N.E. van Haren, L. Abramovic, T.W. Scheewe, R. M. Brouwer, H.E. Hulshoff Pol, R.S. Kahn, Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects, *Neuroimage* 84 (2014) 299–306.
- [13] D.M. Ruderfer, A.H. Fanous, S. Ripke, A. McQuillin, R.L. Amdur, C. Schizophrenia Working, Group of the Psychiatric Genomics, , C. Bipolar Disorder Working Group of the Psychiatric Genomics, C. Cross-Disorder Working Group of the Psychiatric Genomics, P.V. Gejman, M.C. O'Donovan, O.A. Andreassen, S. Djurovic, C. M. Hultman, J.R. Kelsoe, S. Jamax, M. Landen, M. Leboyer, V. Nimagaonkar, J. Nurnberger, J.W. Smoller, N. Craddock, A. Corvin, P.F. Sullivan, P. Holmans, P. Sklar, K.S. Kendler, Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia, *Mol. Psychiatr.* 19 (2014) 1017–1024.
- [14] W. Xue, T. Fu, S. Deng, F. Yang, J. Yang, F. Zhu, Molecular mechanism for the allosteric inhibition of the human serotonin transporter by antidepressant escitalopram, *ACS Chem. Neurosci.* 13 (2022) 340–351.
- [15] E.Z. Hoseth, F. Krull, I. Dieset, R.H. Morch, S. Hope, E.S. Gardsjord, N.E. Steen, I. Melle, H.R. Brattbakk, V.M. Steen, P. Aukrust, S. Djurovic, O.A. Andreassen,

- T. Ueland, Exploring the Wnt signaling pathway in schizophrenia and bipolar disorder, *Transl. Psychiatry* 8 (2018) 55.
- [16] R. Birnbaum, D.R. Weinberger, Genetic insights into the neurodevelopmental origins of schizophrenia, *Nat. Rev. Neurosci.* 18 (2017) 727–740.
- [17] D. Tkachev, M.L. Mimmack, M.M. Ryan, M. Wayland, T. Freeman, P.B. Jones, M. Starkey, M.J. Webster, R.H. Yolken, S. Bahn, Oligodendrocyte dysfunction in schizophrenia and bipolar disorder, *Lancet* 362 (2003) 798–805.
- [18] A. de Baumont, M. Maschietto, L. Lima, D.M. Carraro, E.H. Olivier, A. Fiorini, L. A. Barreta, J.A. Palha, P. Belmonte-de-Abreu, C.A. Moreira Filho, H. Brentani, Innate immune response is differentially dysregulated between bipolar disease and schizophrenia, *Schizophr. Res.* 161 (2015) 215–221.
- [19] S.G. Fillman, D. Sinclair, S.J. Fung, M.J. Webster, C. Shannon Weickert, Markers of inflammation and stress distinguish subsets of individuals with schizophrenia and bipolar disorder, *Transl. Psychiatry* 4 (2014), e365.
- [20] M. Mistry, J. Gillis, P. Pavlidis, Genome-wide expression profiling of schizophrenia using a large combined cohort, *Mol. Psychiatr.* 18 (2013) 215–225.
- [21] S. Tippmann, Programming tools: adventures with R, *Nature* 517 (2015) 109–110.
- [22] Q. Yang, J. Hong, Y. Li, W. Xue, S. Li, H. Yang, F. Zhu, A novel bioinformatics approach to identify the consistently well-performing normalization strategy for current metabolomic studies, *Briefings Bioinf.* 21 (2020) 2142–2152.
- [23] M. Mistry, J. Gillis, P. Pavlidis, Meta-analysis of gene coexpression networks in the post-mortem prefrontal cortex of patients with schizophrenia and unaffected controls, *BMC Neurosci.* 14 (2013) 105.
- [24] Q. Yang, B. Li, J. Tang, X. Cui, Y. Wang, X. Li, J. Hu, Y. Chen, W. Xue, Y. Lou, Y. Qiu, F. Zhu, Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data, *Briefings Bioinf.* 21 (2020) 1058–1068.
- [25] Q. Yang, Y. Wang, F.C. Li, S. Zhang, Y. Luo, Y. Li, J. Tang, B. Li, Y. Chen, W. Xue, F. Zhu, Identification of the gene signature reflecting schizophrenia's etiology by constructing artificial intelligence-based method of enhanced reproducibility, *CNS Neurosci. Ther.* 25 (2019) 1054–1063.
- [26] Q. Yang, Y. Gong, Construction of the classification model using key genes identified between benign and malignant thyroid nodules from comprehensive transcriptomic data, *Front. Genet.* 12 (2021), 791349.
- [27] Q. Yang, Y. Wang, Y. Zhang, F. Li, W. Xia, Y. Zhou, Y. Qiu, H. Li, F. Zhu, NOREVA: enhanced normalization and evaluation of time-course and multi-class metabolomic data, *Nucleic Acids Res.* 48 (2020) W436–W448.
- [28] W.E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods, *Biostatistics* 8 (2007) 118–127.
- [29] Q. Yang, B. Li, S. Chen, J. Tang, Y. Li, Y. Li, S. Zhang, C. Shi, Y. Zhang, M. Mou, W. Xue, F. Zhu, MMEASE: online meta-analysis of metabolomic data by enhanced metabolite annotation, marker selection and enrichment analysis, *J. Proteomics* 232 (2021), 104023.
- [30] L.C. Lee, C.Y. Liou, A.A. Jemain, Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps, *Analyst* 143 (2018) 3526–3539.
- [31] F. Li, Y. Zhou, Y. Zhang, J. Yin, Y. Qiu, J. Gao, F. Zhu, POSREG: proteomic signature discovered by simultaneously optimizing its reproducibility and generalizability, *Brief. Bioinform.* 23 (2022) bbac040.
- [32] K.A. Le Cao, S. Boitard, P. Besse, Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems, *BMC Bioinf.* 12 (2011) 253.
- [33] W. Xia, L. Zheng, J. Fang, F. Li, Y. Zhou, Z. Zeng, B. Zhang, Z. Li, H. Li, F. Zhu, PfMulDL: a novel strategy enabling multi-class and multi-label protein function annotation by integrating diverse deep learning methods, *Comput. Biol. Med.* 145 (2022), 105465.
- [34] D. Ruiz-Perez, H. Guan, P. Madhivanan, K. Mathee, G. Narasimhan, So you think you can PLS-DA? *BMC Bioinf.* 21 (2020) 2.
- [35] Q. Wu, Z. Zhang, H. Zhu, T. Li, X. Zhu, H. Gao, Z. Yun, Y. Jiang, Comparative volatile compounds and primary metabolites profiling of pitaya fruit peel after ozone treatment, *J. Sci. Food Agric.* 99 (2019) 2610–2621.
- [36] C. Christin, H.C. Hoefsloot, A.K. Smilde, B. Hoekman, F. Suits, R. Bischoff, P. Horvatovich, A critical assessment of feature selection methods for biomarker discovery in clinical proteomics, *Mol. Cell. Proteomics* 12 (2013) 263–276.
- [37] A.L. Boulesteix, K. Strimmer, Partial least squares: a versatile tool for the analysis of high-dimensional genomic data, *Briefings Bioinf.* 8 (2007) 32–44.
- [38] T. Li, Q. Wu, X. Duan, Z. Yun, Y. Jiang, Proteomic and transcriptomic analysis to unravel the influence of high temperature on banana fruit during postharvest storage, *Funct. Integr. Genomics* 19 (2019) 467–486.
- [39] J.R. Belmonte-Sanchez, R. Romero-Gonzalez, F.J. Arreola, J.L.M. Vidal, A. Garrido Frenich, An innovative metabolomic approach for golden rum classification combining ultrahigh-performance liquid chromatography-orbitrap mass spectrometry and chemometric strategies, *J. Agric. Food Chem.* 67 (2019) 1302–1311.
- [40] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K.P. Tsafou, M. Kuhn, P. Bork, L.J. Jensen, C. von Mering, STRING v10: protein-protein interaction networks, integrated over the tree of life, *Nucleic Acids Res.* 43 (2015) D447–D452.
- [41] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.* 13 (2003) 2498–2504.
- [42] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 15545–15550.
- [43] J. Fu, Y. Zhang, Y. Wang, H. Zhang, J. Liu, J. Tang, Q. Yang, H. Sun, W. Qiu, Y. Ma, Z. Li, M. Zheng, F. Zhu, Optimization of metabolomic data processing using NOREVA, *Nat. Protoc.* 17 (2022) 129–151.
- [44] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, *ACM Trans. Intell. Syst. Technol.* 2 (2011) 1–27.
- [45] D. Arion, Z. Huo, J.F. Enwright, J.P. Corradi, G. Tseng, D.A. Lewis, Transcriptome alterations in prefrontal pyramidal cells distinguish schizophrenia from bipolar and major depressive disorders, *Biol. Psychiatr.* 82 (2017) 594–600.
- [46] B.W. Higgs, M. Elashoff, S. Richman, B. Barci, An online database for brain disease research, *BMC Genom.* 7 (2006) 70.
- [47] M.H. Hagenauer, A. Schulmann, J.Z. Li, M.P. Vawter, D.M. Walsh, R.C. Thompson, C.A. Turner, W.E. Bunney, R.M. Myers, J.D. Barchas, A.F. Schatzberg, S.J. Watson, H. Akil, Inference of cell type content from human brain transcriptomic datasets illuminates the effects of age, manner of death, dissection, and psychiatric diagnosis, *PLoS One* 13 (2018), e0200003.
- [48] T.A. Lanz, J.J. Joshi, V. Reinhart, K. Johnson, L.E. Grantham 2nd, D. Volkson, STEP levels are unchanged in pre-frontal cortex and associative striatum in post-mortem human brain samples from subjects with schizophrenia, bipolar disorder and major depressive disorder, *PLoS One* 10 (2015), e0121744.
- [49] J. Perez-Santiago, R. Diez-Alarcia, L.F. Callado, J.X. Zhang, G. Chana, C.H. White, S.J. Glatt, M.T. Tsuang, I.P. Everall, J.J. Meana, C.H. Woell, A combined analysis of microarray gene expression studies of the human prefrontal cortex identifies genes implicated in schizophrenia, *J. Psychiatr. Res.* 46 (2012) 1464–1474.
- [50] Q. Zhou, D. Peng, X. Yuan, Z. Lv, S. Pang, W. Jiang, C. Yang, X. Shi, G. Pang, Y. Yang, H. Xie, W. Zhang, C. Hu, Z. Yang, APOE and APOC1 gene polymorphisms are associated with cognitive impairment progression in Chinese patients with late-onset Alzheimer's disease, *Neural Regen. Res.* 9 (2014) 653–660.
- [51] S. Okazaki, S. Boku, I. Otsuka, K. Mouri, S. Aoyama, K. Shiroiwa, I. Sora, A. Fujita, Y. Shirai, O. Shirakawa, M. Kokai, A. Hishimoto, The cell cycle-related genes as biomarkers for schizophrenia, *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* 70 (2016) 85–91.
- [52] M. Klarer, J.P. Krieger, J. Richetto, U. Weber-Stadlbauer, L. Gunther, C. Winter, M. Arnold, W. Langhans, U. Meyer, Abdominal vagal afferents modulate the brain transcriptome and behaviors relevant to schizophrenia, *J. Neurosci.* 38 (2018) 1634–1647.
- [53] M.L. Seney, K. Cahill, J.F. Enwright 3rd, R.W. Logan, Z. Huo, W. Zong, G. Tseng, C. A. McClung, Diurnal rhythms in gene expression in the prefrontal cortex in schizophrenia, *Nat. Commun.* 10 (2019) 3355.
- [54] P.N. Fiorica, H.E. Wheeler, Transcriptome association studies of neuropsychiatric traits in African Americans implicate PRMT7 in schizophrenia, *PeerJ* 7 (2019), e7778.
- [55] F. Guan, H. Lin, G. Chen, L. Li, T. Chen, X. Liu, J. Han, T. Li, Evaluation of association of common variants in HTR1A and HTR5A with schizophrenia and executive function, *Sci. Rep.* 6 (2016), 38048.
- [56] A. Yosifova, T. Mushirosda, D. Stoianov, R. Vazharova, I. Dimova, S. Karachanak, I. Zaharieva, V. Milanova, N. Madjrova, I. Gerdjikov, T. Tolev, S. Velkova, G. Kirov, M.J. Owen, M.C. O'Donovan, D. Toncheva, Y. Nakamura, Case-control association study of 65 candidate genes revealed a possible association of a SNP of HTR5A to be a factor susceptible to bipolar disease in Bulgarian population, *J. Affect. Disord.* 117 (2009) 87–97.
- [57] G.D. Bader, C.W. Hogue, An automated method for finding molecular complexes in large protein interaction networks, *BMC Bioinf.* 4 (2003) 2.
- [58] C.H. Chin, S.H. Chen, H.H. Wu, C.W. Ho, M.T. Ko, C.Y. Lin, cytoHubba: identifying hub objects and sub-networks from complex interactome, *BMC Syst. Biol.* 8 (2014) S11.
- [59] Y. Chen, Z. Zhao, F. Lin, L. Wang, Z. Lin, W. Yue, Associations between genotype and peripheral complement proteins in first-episode psychosis: evidences from C3 and C4, *Front. Genet.* 12 (2021), 647246.
- [60] M.T. Heneka, R.M. McManus, E. Latz, Inflammasome signalling in brain function and neurodegenerative disease, *Nat. Rev. Neurosci.* 19 (2018) 610–621.
- [61] E. Shink, J. Morissette, R. Sherrington, N. Barden, A genome-wide scan points to a susceptibility locus for bipolar disorder on chromosome 12, *Mol. Psychiatr.* 10 (2005) 545–552.
- [62] N.R. Wray, M.L. Pergadia, D.H. Blackwood, B.W. Penninx, S.D. Gordon, D. R. Nyholt, S. Ripke, D.J. MacIntyre, K.A. McGhee, A.W. Maclean, J.H. Smit, J. J. Hottenga, G. Willemsen, C.M. Middeldorp, E.J. de Geus, C.M. Lewis, P. McGuffin, I.B. Hickie, E.J. van den Oord, J.Z. Liu, S. Macgregor, B.P. McEvoy, E. M. Byrne, S.E. Medland, D.J. Statham, A.K. Henders, A.C. Heath, G. W. Montgomery, N.G. Martin, D.I. Boomsma, P.A. Madden, P.F. Sullivan, Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned, *Mol. Psychiatr.* 17 (2012) 36–48.
- [63] S.M. Neuner, S.E. Heuer, J.G. Zhang, V.M. Philip, C.C. Kaczorowski, Identification of pre-symptomatic gene signatures that predict resilience to cognitive decline in the genetically diverse AD-BXD model, *Front. Genet.* 10 (2019) 35.
- [64] V.G. Martinez, L. O'Driscoll, Neuromedin U: a multifunctional neuropeptide with pleiotropic roles, *Clin. Chem.* 61 (2015) 471–482.
- [65] V. Cardoso, J. Chesne, H. Ribeiro, B. Garcia-Cassani, T. Carvalho, T. Bouchery, K. Shah, N.L. Barbosa-Morais, N. Harris, H. Veiga-Fernandes, Neuronal regulation of type 2 innate lymphoid cells via neuromedin U, *Nature* 549 (2017) 277–281.
- [66] O. Mantere, K. Trontti, J. Garcia-Gonzalez, I. Balcells, S. Saarnio, T. Mantyla, M. Lindgren, T. Kieseppa, T. Raji, J.K. Honkanen, O. Vaarala, I. Hovatta, J. Suvisaari, Immunomodulatory effects of antipsychotic treatment on gene expression in first-episode psychosis, *J. Psychiatr. Res.* 109 (2019) 18–26.
- [67] T. Fu, F. Li, Y. Zhang, J. Yin, W. Qiu, X. Li, X. Liu, W. Xin, C. Wang, L. Yu, J. Gao, Q. Zheng, S. Zeng, F. Zhu, Varidt 2.0: structural variability of drug transporter, *Nucleic Acids Res.* 50 (2022) D1417–D1431.

- [68] B. Rodriguez, J.V. Nani, P.G.C. Almeida, E. Brietzke, R.S. Lee, M.A.F. Hayashi, Neuropeptides and oligopeptidases in schizophrenia, *Neurosci. Biobehav. Rev.* 108 (2020) 679–693.
- [69] W. Warnica, D. Merico, G. Costain, S.E. Alfred, J. Wei, C.R. Marshall, S.W. Scherer, A.S. Bassett, Copy number variable microRNAs in schizophrenia and their neurodevelopmental gene targets, *Biol. Psychiatr.* 77 (2015) 158–166.
- [70] N.C. Allen, S. Bagade, M.B. McQueen, J.P. Ioannidis, F.K. Kavvoura, M.J. Khouri, R.E. Tanzi, L. Bertram, Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database, *Nat. Genet.* 40 (2008) 827–834.
- [71] S. Chang, J. Wang, K. Zhang, J. Wang, Pathway-based analysis for genome-wide association study data of bipolar disorder provides new insights for genetic study, *Protein Cell* 6 (2015) 912–915.
- [72] D.E. Adkins, A.N. Khachane, J.L. McClay, K. Aberg, J. Bukszar, P.F. Sullivan, E. J. van den Oord, SNP-based analysis of neuroactive ligand-receptor interaction pathways implicates PGE2 as a novel mediator of antipsychotic treatment response: data from the CATIE study, *Schizophr. Res.* 135 (2012) 200–201.
- [73] E.C. Santa Cruz, F.D.S. Zandonadi, W. Fontes, A. Sussulini, A pilot study indicating the dysregulation of the complement and coagulation cascades in treated schizophrenia and bipolar disorder patients, *Biochim. Biophys. Acta, Proteins Proteomics* 1869 (2021), 140657.
- [74] A. Navarro-Romero, M. Montpeyo, M. Martinez-Vicente, The emerging role of the lysosome in Parkinson's disease, *Cells* 9 (2020) 2399.
- [75] The International Schizophrenia Consortium, Common polygenic variation contributes to risk of schizophrenia and bipolar disorder, *Nature* 460 (2009) 748–752.
- [76] P. Jia, Z. Zhao, Searching joint association signals in CATIE schizophrenia genome-wide association studies through a refined integrative network approach, *BMC Genom.* 13 (2012) S15.
- [77] A.J. Funk, R.E. McCullumsmith, V. Haroutunian, J.H. Meador-Woodruff, Abnormal activity of the MAPK- and cAMP-associated signaling pathways in frontal cortical areas in postmortem brain in schizophrenia, *Neuropsychopharmacology* 37 (2012) 896–905.
- [78] E. Gardiner, N.J. Beveridge, J.Q. Wu, V. Carr, R.J. Scott, P.A. Tooney, M.J. Cairns, Imprinted DLK1-DIO3 region of 14q32 defines a schizophrenia-associated miRNA signature in peripheral blood mononuclear cells, *Mol. Psychiatr.* 17 (2012) 827–840.