



## OPEN ACCESS

## EDITED BY

Massimo Tusconi,  
University of Cagliari, Italy

## REVIEWED BY

Maria Carmela Padula,  
Ospedale San Carlo, Italy  
Jung Goo Lee,  
Inje University Haeundae Paik Hospital,  
Republic of Korea

## \*CORRESPONDENCE

Huifeng Yang  
✉ 290367464@qq.com

RECEIVED 16 March 2023

ACCEPTED 20 September 2023

PUBLISHED 04 October 2023

## CITATION

Shen J, Feng Y, Lu M, He J and Yang H (2023) Identification of the role of immune-related genes in the diagnosis of bipolar disorder with metabolic syndrome through machine learning and comprehensive bioinformatics analysis. *Front. Psychiatry* 14:1187360. doi: 10.3389/fpsy.2023.1187360

## COPYRIGHT

© 2023 Shen, Feng, Lu, He and Yang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Identification of the role of immune-related genes in the diagnosis of bipolar disorder with metabolic syndrome through machine learning and comprehensive bioinformatics analysis

Jing Shen<sup>1</sup>, Yu Feng<sup>2,3</sup>, Minyan Lu<sup>1</sup>, Jin He<sup>1</sup> and Huifeng Yang<sup>1\*</sup>

<sup>1</sup>The Affiliated Jiangsu Shengze Hospital of Nanjing Medical University, Nanjing, China, <sup>2</sup>Medicine and Health, The University of New South Wales, Kensington, NSW, Australia, <sup>3</sup>Melbourne Medical School, The University of Melbourne, Parkville, VIC, Australia

**Background:** Bipolar disorder and metabolic syndrome are both associated with the expression of immune disorders. The current study aims to find the effective diagnostic candidate genes for bipolar affective disorder with metabolic syndrome.

**Methods:** A validation data set of bipolar disorder and metabolic syndrome was provided by the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) were found utilizing the Limma package, followed by weighted gene co-expression network analysis (WGCNA). Further analyses were performed to identify the key immune-related center genes through function enrichment analysis, followed by machine learning-based techniques for the construction of protein-protein interaction (PPI) network and identification of the Least Absolute Shrinkage and Selection Operator (LASSO) and Random Forest (RF). The receiver operating characteristic (ROC) curve was plotted to diagnose bipolar affective disorder with metabolic syndrome. To investigate the immune cell imbalance in bipolar disorder, the infiltration of the immune cells was developed.

**Results:** There were 2,289 DEGs in bipolar disorder, and 691 module genes in metabolic syndrome were identified. The DEGs of bipolar disorder and metabolic syndrome module genes crossed into 129 genes, so a total of 5 candidate genes were finally selected through machine learning. The ROC curve results-based assessment of the diagnostic value was done. These results suggest that these candidate genes have high diagnostic value.

**Conclusion:** Potential candidate genes for bipolar disorder with metabolic syndrome were found in 5 candidate genes (AP1G2, C1orf54, DMAC2L, RABEPK and ZFAND5), all of which have diagnostic significance.

## KEYWORDS

bipolar disorder, metabolic syndrome, differentially expressed genes, machine learning, immune infiltration bipolar disorder, immune infiltration

## 1. Introduction

Bipolar disorder is a serious mental illness characterized by depression, mania, and mixed development (1). Patients often have persistent residual symptoms, psychosocial function problems, cognitive impairment, and low quality of life, along with other complications (2). Bipolar disorder is often missed or misdiagnosed, which makes the disease statistics much lower than its actual prevalence (3). The disease has also been proven to be a chronic hereditary disease (4), and environmental factors have a significant influence on the development and recurrence of the condition (5). Therefore, genetic analysis of patients will provide more support for future diagnosis and therapy. According to prior research, there are multiple correlation patterns between bipolar disorder and various medical diseases involving multiple organ systems (6).

The term “metabolic syndrome” describes the co-existence of numerous recognized cardiovascular risk factors (obesity, insulin resistance, atherogenic dyslipidemia, and hypertension) (7). The disease has grown to be a significant health risk in the modern world as western habits have become more widespread (8). For example, the overall standardized prevalence rate of metabolic syndrome in China is 24.2% (9), which may be related to the high local drinking rate (the drinking rate of Chinese adults is 33.66%) (10), making it one of the major diseases in China.

The lifetime attempted suicide rate of bipolar disorder is 32.3% (11), and the premature mortality rate is also higher than that of other populations. Brenda W J H Penninx et al. reported that the increased risk of premature death of psychiatric patients might be related to the increased risk of metabolic syndrome (12). This study analyzes the pathogenesis of bipolar disorder with metabolic syndrome from the perspective of the immune system.

As is widely known, mental disorders, including BD, can influence a range of metabolic changes such as blood lipids. In our extensive literature review, we found that most studies have focused on the impact of BD on MS (13–15).

However, what piques our interest is the shared risk factors between BD and MS. We are intrigued by the possibility of a dynamic interplay between them, where each condition might influence the other in a dialectical manner. Our aim is to unravel the underlying mechanisms behind this interaction and explore the feasibility of establishing a predictive diagnostic model for early detection of BD in patients with MS.

## 2. Methodology

### 2.1. Data collection

The bipolar disorder data set GSE5388, and the metabolic syndrome data set GSE98895 were provided by the GEO database<sup>1</sup> (16). The specific process is shown in the detailed data set information is shown in Table 1.

TABLE 1 Detailed data set information.

GSE series	Type	Sample size		Platform
		Bipolar disorder	Control	
GSE5388	mRNA	30	31	GPL95
GSE53987	mRNA	52	55	GPL570
GSE39653	mRNA	8	24	GPL10558
		Metabolic syndrome	Control	
GSE98895	mRNA	20	20	GPL6947

### 2.2. The identification of DEGs

The differential expression screening approach Limma (linear models for microarray data) is based on generalized linear models. Here, we do a differential analysis to derive differential expression using the R package limma (version 3.40.6). Genes that differ between the comparison group and the control group (17). The expression spectrum dataset we obtained is specifically used to perform multiple linear regression using the lmFit function, and then the eBayes function is used to compute moderated t-statistics, moderated F-statistics, and log-odds of differential expression by empirical Bayes moderation of the standard errors toward a common value, and finally to determine the significance of the difference for each gene. Create filters for genes with differential expression: To find genes that are differentially expressed, use the following filters: Set a criterion of  $|\log_2$  Fold Change (FC)| > 1 for bipolar disorder filtering and > 1.5 for metabolic syndrome filtering. Also, use  $P=0.05$  as significance level. Genes with notable changes were found using these tests.

### 2.3. Weighted gene co-expression network analysis and modular gene selection

First, we calculated the MAD (Median Absolute Deviation) of each gene using the gene expression profile (18). We then deleted the top 50% of genes with the smallest MAD, and then excluded outlier genes and samples using the goodSamplesGenes function of the R package WGCNA. Specifically, you can utilize WGCNA to create a scale-free co-expression network. For all pair-wise Genes, the average linkage approach and Pearson's correlation matrices were first used. Then, using the power function  $A_{mn} = |C_{mn}|$  ( $C_{mn}$  = Pearson's correlation between Gene<sub>m</sub> and Gene<sub>n</sub>;  $A_{mn}$  = adjacency between Gene<sub>m</sub> and Gene<sub>n</sub>), a weighted adjacency matrix was created.  $\beta$  was a soft-thresholding parameter that may highlight significant gene-to-gene correlations and punish less significant ones. The Adjacency Was Transformed Into A Topology Overlap Matrix (Tom), Which Could Measure the Network Connectivity of a Gene DEFINED AS the Sum of ITS Adjacency with All Other Genes for Network Gene Ration, and the CorreSponding Disabilities (1-TOM) Were Calculated. After Selecting the Power of 16, The Adjacency Was Transformed Into A Topology Overlap Matrix (Tom). Average linkage hierarchical clustering was carried out using the TOM-based dissimilarity measure with a minimum size (Gene group) of 30 for the Genes dendrogram

<sup>1</sup> <https://www.ncbi.nlm.nih.gov/geo/>

in order to arrange genes with comparable expression profiles into gene modules. The sensitivity should be set to 3; We determined the module's eigengene dissimilarity, selected a cut line for the dendrogram, and combined certain modules in order to further investigate the module. It is important to note that after merging the modules with a distance of less than 0.25, we eventually acquired 10 co-expression modules. The gray module is regarded as a group of genes that cannot be allocated to any module, which is the most crucial aspect. Then, using truncation criteria ( $|MM| > 0.8$ ), we calculated the expression correlation between the gene and the module feature vector to obtain MM. This allowed us to search for hub genes with high connectivity in clinically significant modules.

## 2.4. Gene function enrichment analysis

For gene set function enrichment analysis, KEGG rest API<sup>2</sup> was used to retrieve the latest KEGG pathway gene annotation, which was used to map genes to the background set, and the 'clusterProfiler' R software package (19) was employed for enrichment analysis to obtain the outcomes of gene set enrichment. The minimum gene was set as 5, the maximum gene set as 5,000, and the  $p$  value  $< 0.05$  was taken as statistically significant.

## 2.5. PPI network construction

STRING was employed to develop the protein–protein interaction network (PPI) (20). The minimum required interaction score was set to medium confidence (0.400). PPI vector graphics and protein node degrees were downloaded directly from the website for further analysis.

## 2.6. Machine learning

In this study, 'glmnet' and "randomForest" R package was used to integrate survival status and time and gene expression data, the LASSO cox method and RandomForest was used for regression analysis (21, 22). Moreover, five-fold cross-validation was also set to optimize the model. We set the Lambda value to 0.08.

Data integration: Integrate gene expression, survival time, and survival status data into a single dataset. To make sure that each sample is connected to its matching survival status, survival time, and gene expression values, this may entail combining and structuring the various data tables. LASSO Cox Regression: The R package 'glmnet' was used to do regression analysis using the LASSO Cox approach. Least Absolute Shrinkage and Selection Operator, sometimes known as LASSO, is a regression analysis technique for high-dimensional data. It has the ability to use variable selection to exclude key gene features linked to time and survival status. Random Forest Regression: 'randomForest' is a R package that was used to do the regression analysis. Random Forest is an ensemble-based machine learning approach appropriate for both classification and regression issues. Here, it was used for regression analysis, possibly to evaluate the

intricate connections between various genetic features, time, and survival status. Setting the lambda value: The lambda value in the LASSO regression is set to 0.08. In LASSO, the regularization parameter lambda is employed to regulate the model's feature sparsity. Different numbers of gene characteristics can be chosen to create regression models by changing the Lambda value.

The ROC analysis was carried out using 'pROC' R software package to obtain AUC (23). The CI function of pROC was used to evaluate AUC and confidence intervals to obtain the final AUC result. And use sangerbox to visualize (24).

## 2.7. Immune infiltration analysis

The Cibersort (25) in the R software package was employed to analyze the immune cell infiltration, and the 'corrplot' (26) in the R software package was cited to draw the heat map of infiltrating immune cells.

# 3. Results

## 3.1. DEGs identification

The Limma tool was utilized to find 2,289 DEGs in the bipolar disorder data set, of which 1,096 showed up-regulation and 1,193 showed down-regulation. The heat map and volcano map of bipolar disorder DEGs are shown in Figure 1. From the metabolic syndrome data set, 579 DEGs were chosen, of which 342 showed down-regulation and 255 showed up-regulation. The heat map and volcano map of the metabolic syndrome DEGs are shown in Figure 2.

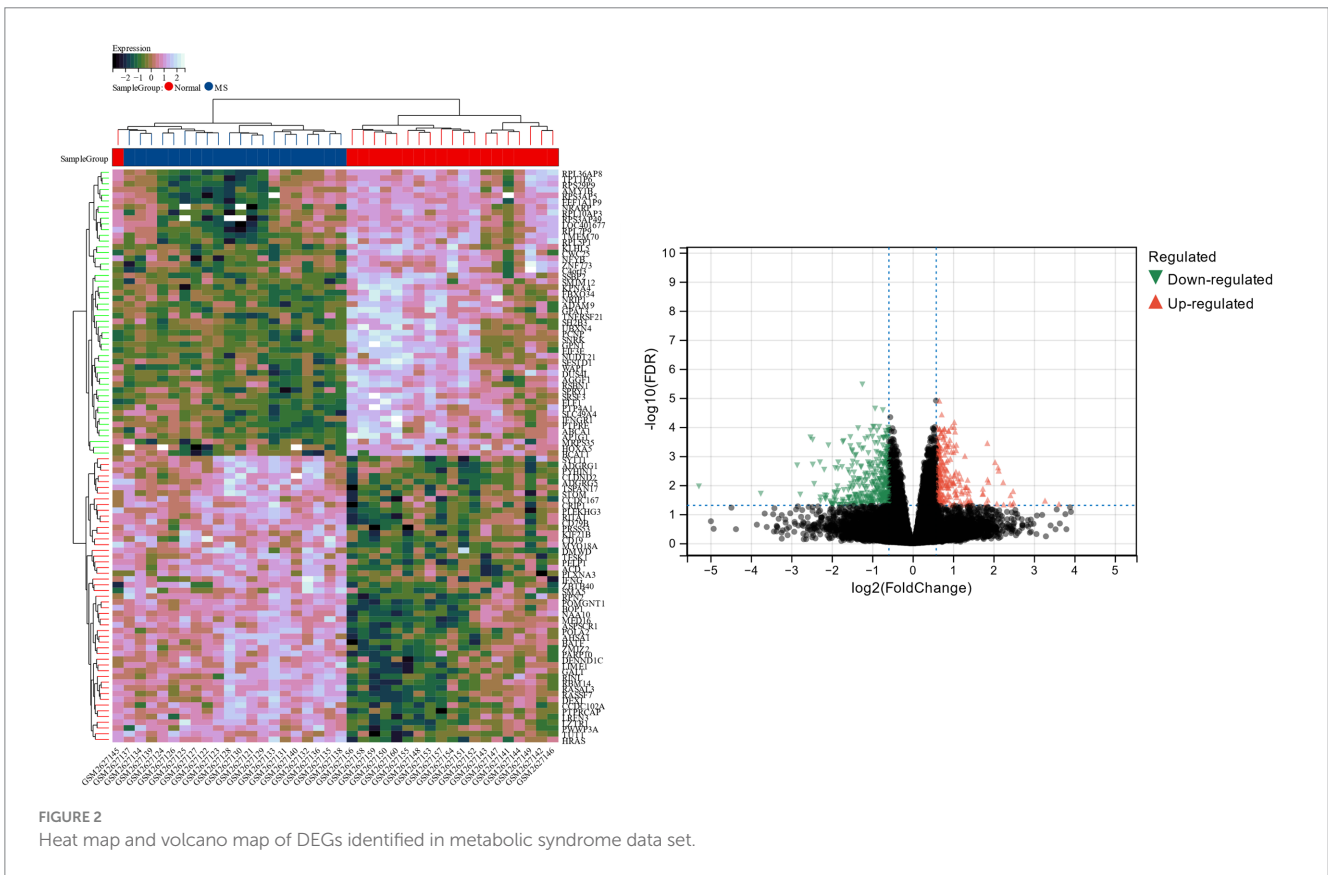
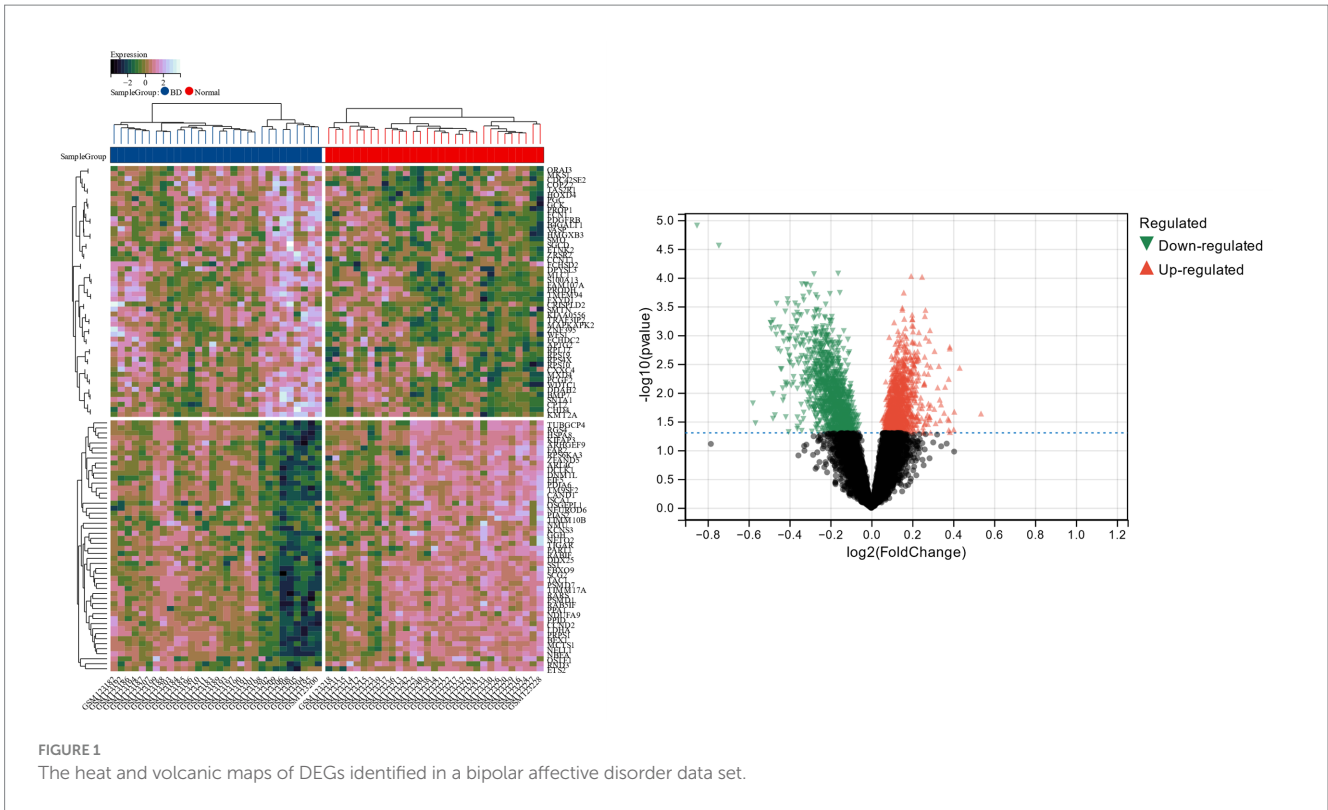
## 3.2. Weighted gene co-expression network analysis and modular gene selection

$\beta$  is a soft threshold parameter that can highlight the high degree of gene-to-gene correlation and penalize a low degree of correlation. Here we chose  $\beta = 14$  as the soft threshold for this study and drew a cluster tree of metabolic syndrome and the control group (Figure 3). On this basis, ten gene co-expression modules (GCMs) were generated (Figures 4A,B), and the correlation between metabolic syndrome and GCM was shown (Figure 4C). The results showed that there was a significant association between metabolic syndrome and the yellow module (691 genes; correlation coefficient = 0.74,  $p = 5.2 \times 10^{-8}$ ). In the metabolic syndrome's yellow module, we also evaluated a relationship between module members and gene significance and found a substantial positive association ( $r = 0.62$ ; Figure 4D).

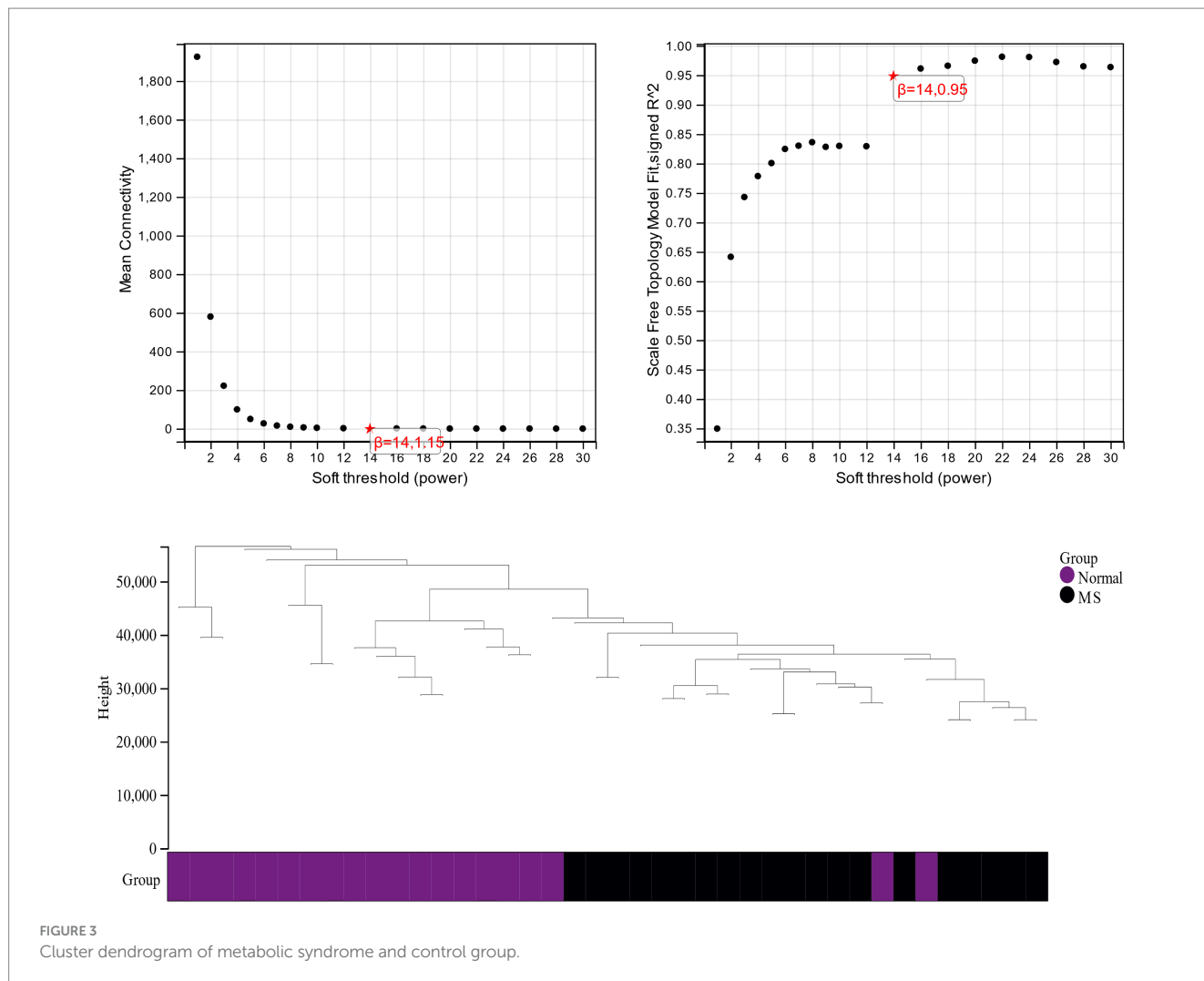
## 3.3. Functional enrichment analysis of metabolic syndrome

According to the research of Zhou et al. (27) GSE98895 is a relatively new data set of metabolic syndrome; therefore, based on the cross genes of the Limma and WGCNA module genes, a functional enrichment analysis was performed. Through the intersection of 579 DEGs and 691 genes in the yellow module, 97 common genes (CGs)

<sup>2</sup> <https://www.kegg.jp/kegg/rest/keggapi.html>







were selected (Figure 5A). CSs were most enriched in “Axon guidance” and “Valine, Leucine and Isoleucine Biosynthesis,” according to KEGG analysis (Figure 5B). GO analysis showed that in terms of cell components (CC), CG was mainly located in the “axon part” and “endoplasmic reticulum-Golgi intermediate compartment” (Figure 5C). The main biological processes (BP) of CGs include “regulation of cellular component organization,” “negative regulation of organelle assemble” and “histone deacetylation” (Figure 5D). Molecular function (MF) showed that the most important items in CGs were “GTPase binding” and “magnesium ion binding” (Figure 5E).

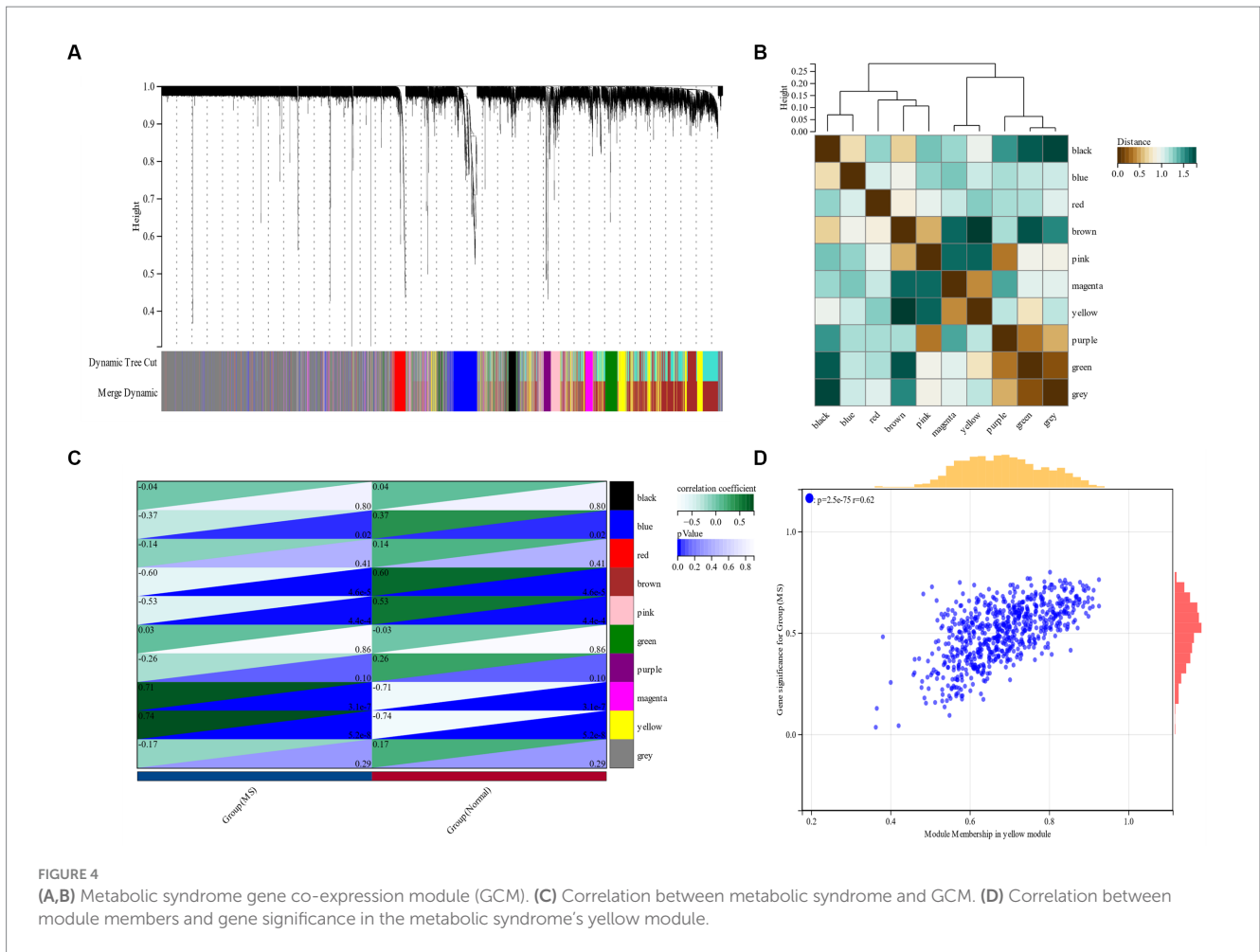
### 3.4. Enrichment analysis of bipolar disorder and metabolic syndrome and node gene identification constructed by PPI network

A total of 129 genes were obtained by crossing the yellow module genes in the DEGs of bipolar disorder and metabolic syndrome (Figure 6A). These genes were primarily enriched in the “mTOR signaling pathway” and “Parkinson’s disease,” according to the KEGG enrichment analysis (Figure 6B). According to GO analysis, CG was

mostly found in “organelle membrane” and “mitochondrion” in terms of cell components (CC; Figure 6C). The main biological processes (BP) of CGs include “proteolysis taking part in cellular protein catabolic process,” “ubiquitin-dependent protein catabolic process,” “modification-dependent protein catabolic process” and “modification-dependent macromolecule catabolic process” (Figure 6D). Molecular function (MF) showed that the most important item in CGs was “enzyme binding” (Figure 6E). A PPI network was constructed, and the analysis showed that 86 genes and the PPI network could interact, and the results were sorted according to the number of nodes (Figures 6F,G).

### 3.5. Identifying candidate genes through machine learning

LASSO regression was used to identify candidate genes, and as per the findings, 14 potential candidate biomarkers (Figures 7A,B) were identified. We also applied RF regression for candidate gene screening, from which ten potential candidate biomarkers were revealed (Figures 7C,D). The results of the two machine learning screens were then intersected and analysed, resulting in five candidate genes



(AP1G2, C1orf54, DMAC2L, RABEPK, ZFAND5; Figure 7E), and subsequent diagnostic value evaluation has been carried out.

Construction of Forest Plot based on five candidate central genes (Figure 8A). ROC curves of 5 potential candidate genes and Forest Plot were established, and the AUC and 95% CI of each item were calculated to assess their specificity and sensitivity. AP1G2 (AUC=0.75, CI=0.87–0.62), C1orf54 (AUC=0.69, CI=0.83–0.56), DMAC2L (AUC=0.66, CI=0.80–0.52), RABEPK (AUC=0.71, CI=0.85–0.58), ZFAND5 (AUC=0.75, CI=0.87–0.62) and Forest Plot (AUC=0.84, CI=0.94–0.75; Figure 8B). The results showed that all candidate genes had high diagnostic value for bipolar disorder with metabolic syndrome.

Further, we validated the diagnostic model on two independent bd datasets, gse39653 (AUC 0.78, CI 0.87–0.70) for the pbmc sample and gse53987 (AUC 0.84, CI 0.98–0.70) for the brain tissue sample (Figure 8B). The results showed that the model also achieved a high diagnostic value in the revalidation group.

### 3.6. Immune cell infiltration analysis

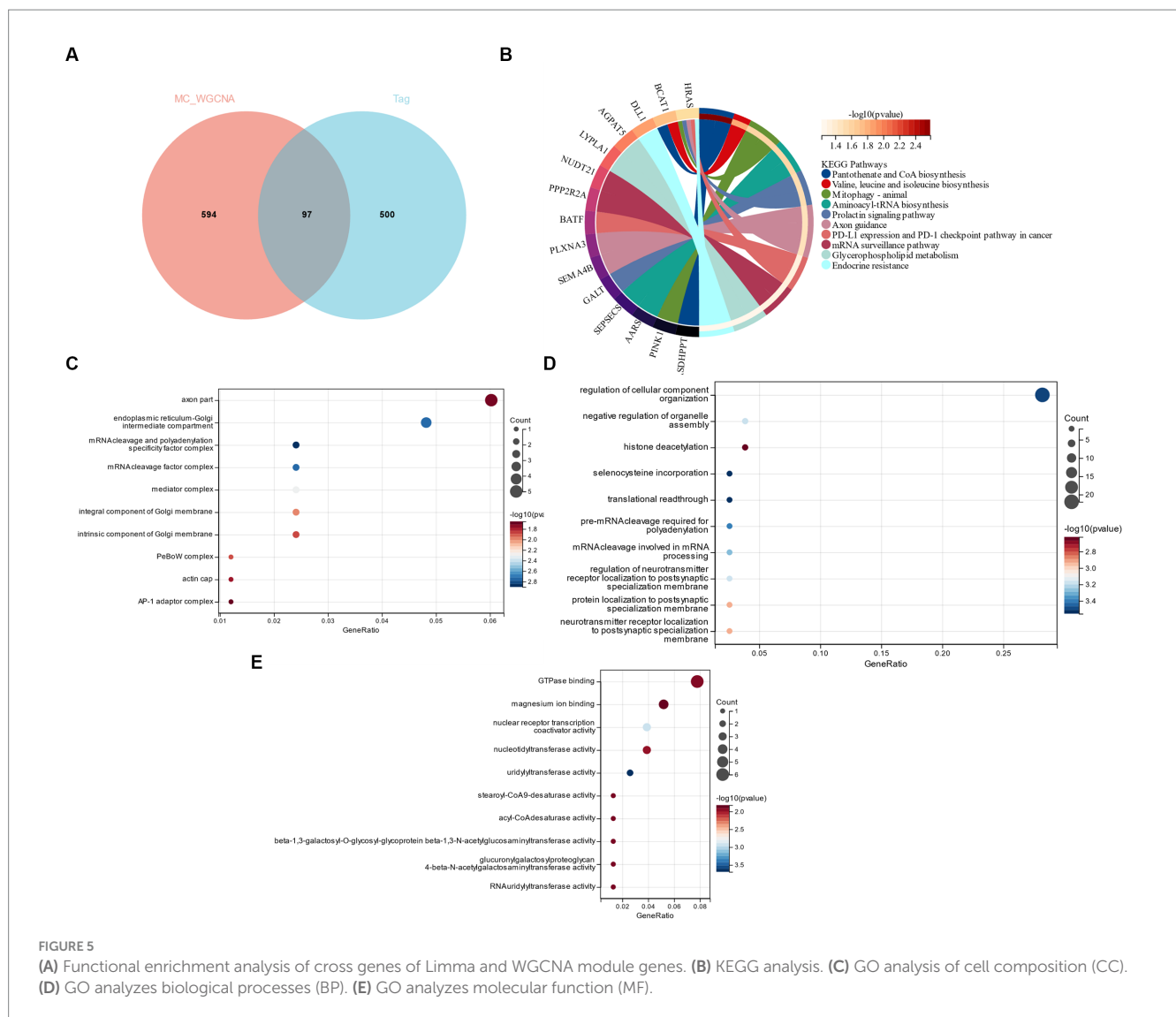
To compare the proportion of each sample's 22 different types of immune cells between the bipolar disorder group and the control group, a bar chart was constructed (Figure 9A). Figure 9B shows that

while the numbers of memory B cells, CD4 resting T cells, monocytes, and resting dendritic cells are lower in individuals with bipolar disorder, the levels of naïve B cells, plasma cells, and resting mast cells are greater. There is a differential infiltration correlation according to the correlation of 22 types of immune cells, excluding naïve T cells and Eosinophils (Figure 9C).

## 4. Discussion

More than 1% of people worldwide suffer from bipolar disorder, and a meta-analysis by Davy Vancampfort et al. found that 32.6% of individuals with severe mental illness also have concomitant metabolic syndrome (28, 29). Patients with bipolar disorder are more likely to have metabolic syndrome, which can negatively affect executive functioning (30). The co-morbid mechanisms of bipolar disorder and metabolic syndrome have not yet been thoroughly studied; thus, this study employed a machine learning approach to identify a link between the two.

This study examined five genes with a potential link to bipolar illness, none of which has been linked to previous research on the condition. AP1G2 (Adaptor Related Protein Complex 1 Subunit Gamma 2) is a Protein Coding gene. Diseases associated with AP1G2 include Tracheoesophageal Fistula With Or Without Esophageal



Atresia and Esophageal Atresia. Among its related pathways are trans-Golgi Network Vesicle Budding and Vesicle-mediated transport. Gene Ontology (GO) annotations related to this gene include binding and obsolete protein transporter activity (31). Additionally, AP1G2 is a linked gene in both Developmental and Epileptic Encephalopathy (32) and Intellectual Developmental Disorder, Autosomal Dominant (33). These two conditions share certain similarities with bipolar disorder, supporting the significance of our findings.

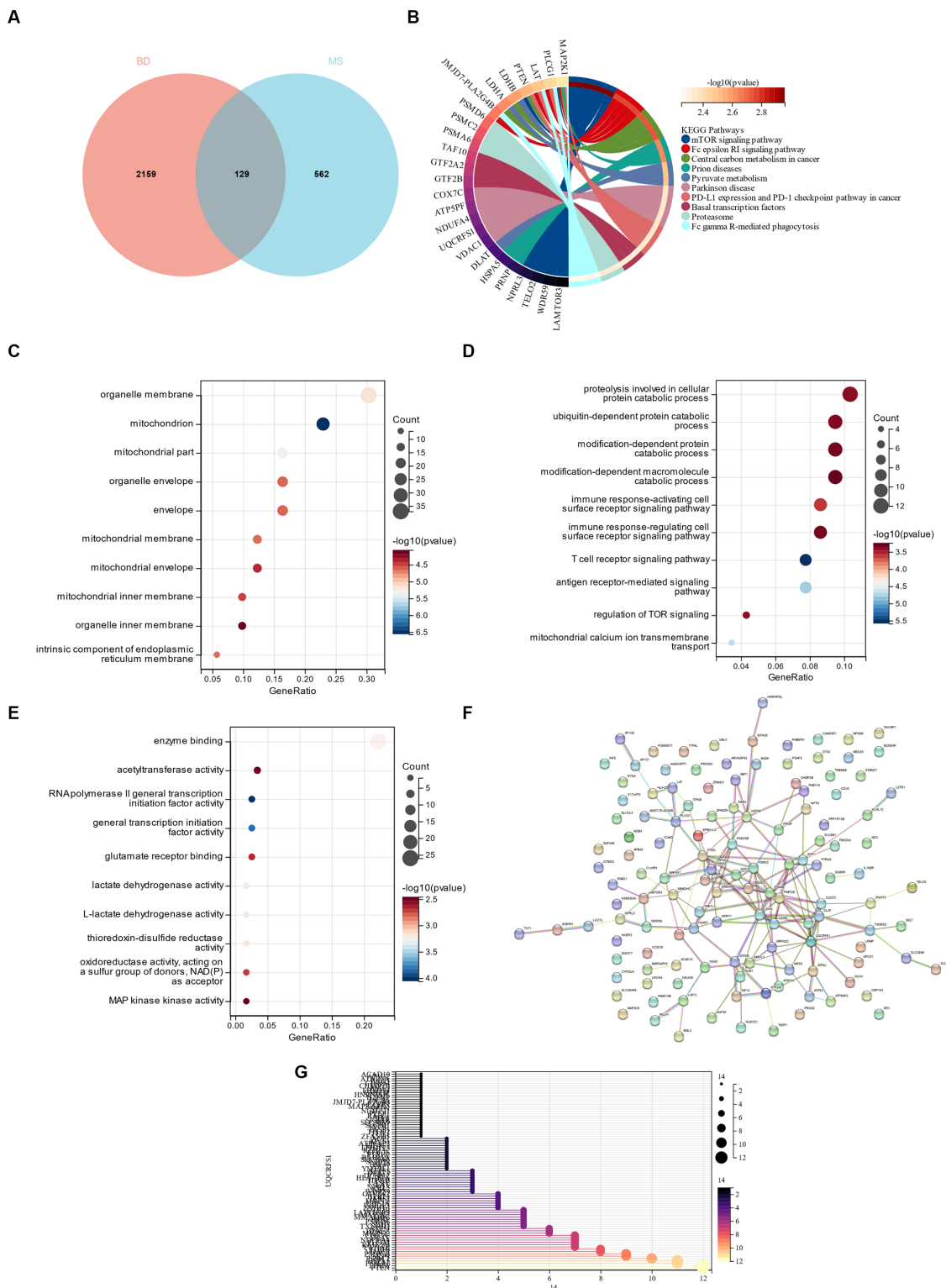
A gene for protein-coding is called RABEPK (Rab9 Effector Protein with Kelch Motifs). Retrograde transport at the Trans-Golgi Network and vesicle-mediated transport are two of its associated processes. Transport is mediated by vesicles and the network (28). We did not search for disorders linked to RABEPK, however Rachel L. Kember's study This gene was discovered to be linked to opioid use disorder (OUD) (34), and previous research has shown a link between OUD and BD (35). Additionally, our candidate gene is correspondingly validated by this.

The proteins encoded by both AP1G2 and RABEPK are associated with the vesicular pathway, which has been a hot topic of research for

several years, and both have shown opposite risk for the development of bd in our study. We are conducting further analysis of this pathway in recent studies, which we believe will help to unravel the mechanisms underlying the prevalence of psychiatric disorders, including bd.

ZFAND5 is a less studied mrna that has been associated with lipoprotein metabolism. zfan5 enhances the stability of ARE-RNA by competing for mRNA binding with TTP, a transcript associated with inflammatory mediators. This is consistent with our analysis in immune infiltration.

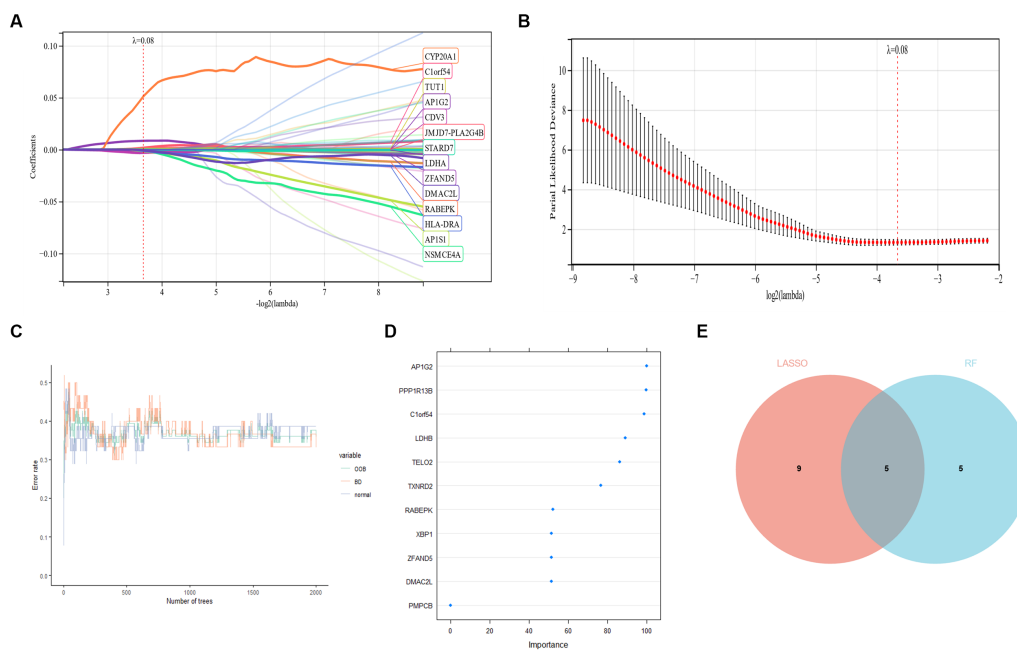
With advances in molecular genetic phenotyping studies, the links between social behavior and genetic variation to protein and receptor function and disease development are increasingly being revealed. The findings of this study shed light on the possibility that metabolic disorders, as an exposure factor, may cause alterations in the genetic markers of qtl, which in turn lead to differential expression of mrna transcripts, which in turn play a role in translational protein function (including alterations in vesicular transport pathways and other metabolic pathways as well as in the immune microenvironment) and finally lead to the outcome of bd. We think this is a very interesting



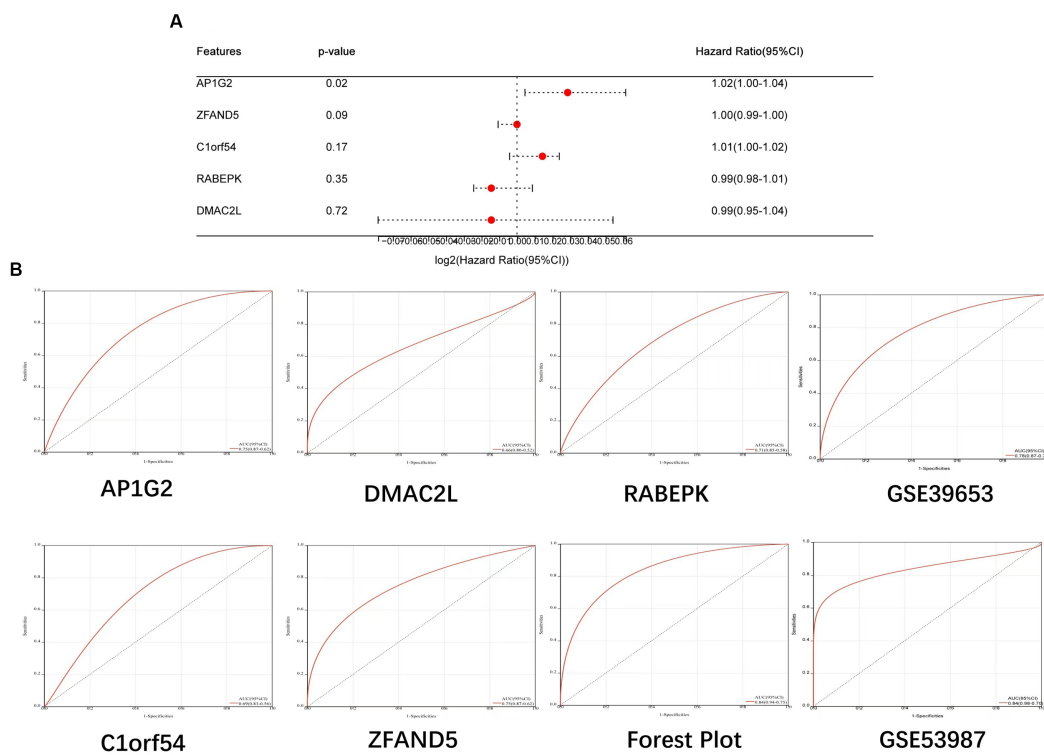
**FIGURE 6** (A) Cross between DEG in bipolar disorder and yellow module gene in metabolic syndrome. (B) KEGG analysis of cross genes. (C) GO analysis of cell composition (CC). (D) GO analyzes BP. (E) GO analysis MF. (F) PPI network interaction diagram. (G) 86 gene nodes in PPI network.

and interesting line of research, and for some time now our team has been working on a Mendelian randomization study analysing neurodegenerative diseases and social behavior. Although this study

has not yet fully elucidated the cause-effect relationship, we believe that this step in the analysis of the mrna and protein levels is important and meaningful.

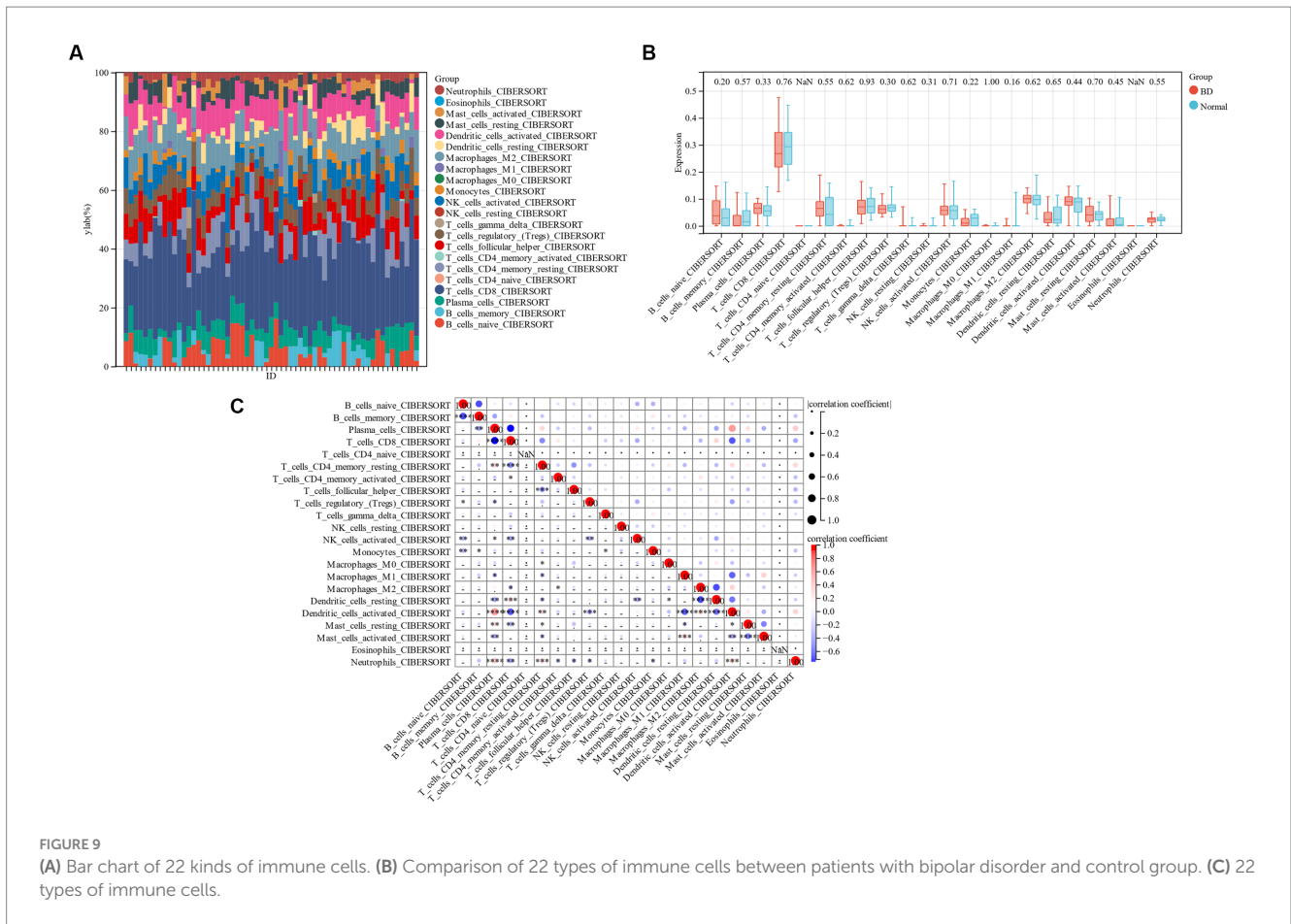


**FIGURE 7** (A,B) LASSO regression candidate gene identifying. (C,D) Applied RF regression for candidate gene screening. (E) Five candidate diagnostic genes are identified via the above two algorithms.



**FIGURE 8** (A) Construction of Forest Plot based on five candidate central genes. (B) ROC curves of 5 potential candidate genes, Forest Plot and Validation Team were established.





In addition, the diagnostic target genes and risk model formulas obtained from this study will be useful for early diagnosis and individualized risk indication for patients suspected of having bd in the metabolic syndrome, and our subsequent studies on genetic markers for qtl will be useful for earlier risk assessment of bd.

### 5. Conclusion

A total of 5 candidate genes were identified, and all of them had diagnostic value, providing potential candidate genes for bipolar disorder with metabolic syndrome.

### Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: <https://www.ncbi.nlm.nih.gov/geo/> with accession numbers GSE5388 and GSE98895.

### Author contributions

YF and JS wrote the main manuscript text. ML, JH, and HY provide experimental help and all authors reviewed the manuscript.

All authors contributed to the article and approved the submitted version.

### Acknowledgments

In the vastness of space and immensity of time, it is my joy to spend a planet and an epoch with Maggie.

### Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

### Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Müller JK, Leweke FM. Bipolar disorder: clinical overview. *Med Monatsschr Pharm.* (2016) 39:363–39.
- Bonnín CDM, Reinares M, Martínez-Arán A, Jiménez E, Sánchez-Moreno J, Solé B, et al. Improving functioning, quality of life, and well-being in patients with bipolar disorder. *Int J Neuropsychopharmacol.* (2019) 22:467–77. doi: 10.1093/ijnp/pyz018
- Sajatovic M. Bipolar disorder: disease burden. *Am J Manag Care.* (2005) 11:S80–4.
- Legrand A, Ifimovici A, Khayachi A, Chaumette B. Epigenetics in bipolar disorder: a critical review of the literature. *Psychiatr Genet.* (2021) 31:1–12. doi: 10.1097/ypg.0000000000000267
- Pitchot W, Scantamburlo G, Anseau M, Souery D. Bipolar disorder: a multifactorial disease. *Rev Med Liege.* (2012) 67:366–73.
- Sinha A, Shariq A, Said K, Sharma A, Jeffrey Newport D, Salloum IM. Medical comorbidities in bipolar disorder. *Curr Psychiatry Rep.* (2018) 20:36. doi: 10.1007/s11920-018-0897-8
- Huang PL. A comprehensive definition for metabolic syndrome. *Dis Model Mech.* (2009) 2:231–7. doi: 10.1242/dmm.001180
- Saklayen MG. The global epidemic of the metabolic syndrome. *Curr Hypertens Rep.* (2018) 20:12. doi: 10.1007/s11906-018-0812-z
- Li Y, Zhao L, Yu D, Wang Z, Ding G. Metabolic syndrome prevalence and its risk factors among adults in China: a nationally representative cross-sectional study. *PLoS One.* (2018) 13:e0199293. doi: 10.1371/journal.pone.0199293
- Li Y, Zhao L, Yu D, Fang H, Yu W, Wang J, et al. Association between drinking and metabolic syndrome among adults in China. *Wei Sheng Yan Jiu.* (2019) 48:531–6.
- D'Ambrosio V, Salvi V, Bogetto F, Maina G. Serum lipids, metabolic syndrome and lifetime suicide attempts in patients with bipolar disorder. *Prog Neuro-Psychopharmacol Biol Psychiatry.* (2012) 37:136–40. doi: 10.1016/j.pnpbp.2011.12.009
- Penninx B, Lange SMM. Metabolic syndrome in psychiatric patients: overview, mechanisms, and implications. *Dialogues Clin Neurosci.* (2018) 20:63–73. doi: 10.31887/DCNS.2018.20.1/bpenninx
- Wu Q, Zhang X, Liu Y, Wang Y. Prevalence and risk factors of comorbid obesity in Chinese patients with bipolar disorder. *Diabetes Metab Syndr Obes.* (2023) 16:1459–69. doi: 10.2147/dmso.S404127
- Tsubata N, Kuroki A, Tsujimura H, Takamasu M, N II, Okamoto T. Pilot and feasibility studies of a lifestyle modification program based on the health belief model to prevent the lifestyle-related diseases in patients with mental illness. *Healthcare.* (2023) 11:1690. doi: 10.3390/healthcare11121690
- Tam To BRoy R, Melikian N, Gaughran FP, O'Gallagher K. Coronary artery disease in patients with severe mental illness. *Interv Cardiol.* (2023) 18:e16. doi: 10.15420/icr.2022.31
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* (2013) 41:D991–5. doi: 10.1093/nar/gks1193
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* (2015) 43:e47. doi: 10.1093/nar/gkv007
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* (2008) 9:559. doi: 10.1186/1471-2105-9-559
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* (2012) 16:284–7. doi: 10.1089/omi.2011.0118
- Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* (2021) 49:D605–12. doi: 10.1093/nar/gkaa1074
- Zhang M, Zhu K, Pu H, Wang Z, Zhao H, Zhang J, et al. An immune-related signature predicts survival in patients with lung adenocarcinoma. *Front Oncol.* (2019) 9:1314. doi: 10.3389/fonc.2019.01314
- Alderden J, Pepper GA, Wilson A, Whitney JD, Richardson S, Butcher R, et al. Predicting pressure injury in critical care patients: a machine-learning model. *Am J Crit Care.* (2018) 27:461–8. doi: 10.4037/ajcc2018525
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* (2011) 12:77. doi: 10.1186/1471-2105-12-77
- Shen W, Song Z, Zhong X, Huang M, Shen D, Gao P, et al. Sangerbox: a comprehensive, interaction-friendly clinical bioinformatics analysis platform. *iMeta.* (2022) 1:e36. doi: 10.1002/imt2.36
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* (2015) 12:453–7. doi: 10.1038/nmeth.3337
- Hu K. Become competent within one day in generating boxplots and violin plots for a novice without prior R experience. *Methods Protoc.* (2020) 3:64. doi: 10.3390/mps3040064
- Zhou Y, Shi W, Zhao D, Xiao S, Wang K, Wang J. Identification of immune-associated genes in diagnosing aortic valve calcification with metabolic syndrome by integrated bioinformatics analysis and machine learning. *Front Immunol.* (2022) 13:937886. doi: 10.3389/fimmu.2022.937886
- Grande I, Berk M, Birmaher B, Vieta E. Bipolar disorder. *Lancet.* (2016) 387:1561–72. doi: 10.1016/s0140-6736(15)00241-x
- Vancampfort D, Stubbs B, Mitchell AJ, De Hert M, Wampers M, Ward PB, et al. Risk of metabolic syndrome and its components in people with schizophrenia and related psychotic disorders, bipolar disorder and major depressive disorder: a systematic review and meta-analysis. *World Psychiatry.* (2015) 14:339–47. doi: 10.1002/wps.20252
- Dalkner N, Bengesser SA, Birner A, Fellendorf FT, Fleischmann E, Großschädl K, et al. Metabolic syndrome impairs executive function in bipolar disorder. *Front Neurosci.* (2021) 15:717824. doi: 10.3389/fnins.2021.717824
- Safran M, Rosen N, Twik M, BarShir R, Stein TI, Dahary D, et al. The GeneCards suite In: I Abugessaisa and T Kasukawa, editors. *Practical guide to life science databases.* Singapore: Springer Nature Singapore (2021). 27–56.
- Rappaport N, Twik M, Plaschkes I, Nudel R, Iny Stein T, Levitt J, et al. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res.* (2016) 45:D877–87. doi: 10.1093/nar/gkw1012
- Dell'Angelica EC, Bonifacino JS. Coatopathies: genetic disorders of protein coats. *Annu Rev Cell Dev Biol.* (2019) 35:131–68. doi: 10.1146/annurev-cellbio-100818-125234
- Kember RL, Vickers-Smith R, Xu H, Toikumo S, Niarchou M, Zhou H, et al. Cross-ancestry meta-analysis of opioid use disorder uncovers novel loci with predominant effects in brain regions associated with addiction. *Nat Neurosci.* (2022) 25:1279–87. doi: 10.1038/s41593-022-01160-z
- Livne O, Sinai O, Lev-Ran S. Shared psychotic disorder associated with bipolar disorder in the primary case in the context of opioid misuse. *J Psychiatr Pract.* (2022) 28:259–64. doi: 10.1097/prs.0000000000000628

## Glossary

BD	Bipolar disorder
MS	Metabolic syndrome
GEO	Gene Expression Omnibus
DEGs	Differentially expressed gene
WGCNA	Weighted gene co-expression network analysis
PPI	Protein–protein interaction
LASSO	Least absolute shrinkage and selection operator
RF	Random Forest
ROC	Receiver operating characteristic
MAD	Medium Absolute Development
GCMs	Co-expression modules
GO	Gene Ontology
ODD	Opioid use disorder
CC	Cell components
BP	Biological processes
MF	Molecular function