BMC Psychiatry

RESEARCH

Open Access

Integrative analysis via bioinformatics and machine learning identifies SERPING1 as a biomarker candidate for major depressive disorder

Xiaokui Yuan[1] and Tong Wang[1*]

Abstract

Background Major Depressive Disorder (MDD) represents a grave mental affliction characterised by intricate pathological mechanisms and an elevated susceptibility to neurodegeneration. This study aims to integrate machine learning and bioinformatics analysis to identify potential biomarkers related to the pathogenesis of MDD and to elucidate the genes that have a causal association with this disease.

Methods The training dataset was formed by merging two GEO datasets (GSE52790 and GSE98793), while the validation dataset included GSE44593 and GSE54564. Bioinformatics methods such as WGCNA, PPI, SMR, and 113 machine learning models (113 different algorithms) were employed in this study.

Results In the course of this study, a total of 20 hub genes were identified. Through machine learning screening, the Lasso + RF model surfaced as the preeminent diagnostic model, registering an AUC value of 0.807 for the SERPING1 gene. Colocalisation analysis unveiled a pronounced upregulation in the expression levels of the pivotal gene, SERPING1, among patients afflicted with MDD. Moreover, Summary-data-based Mendelian Randomisation (SMR) analysis elucidated a substantial positive correlation between SERPING1 and an incresaed risk of MDD, thereby suggested its prospective function in impeding the pathological progression of MDD via the modulation of pertinent pathways.

Conclusions By amalgamating bioinformatics analysis, machine learning, colocalisation analysis, and SMR analysis, this study designated SERPING1 as a potential cardinal gene and biomarker implicated in MDD pathogenesis. These discoveries accentuate the clinical applicability of SERPING1 as a diagnostic biomarker for MDD.

Clinical trial number Not applicable.

Keywords Major depressive disorder, Biomarker, Bioinformatics, Machine learning, SMR

*Correspondence:
Tong Wang
1791147443@qq.com
[1]The Fourth People's Hospital of Chengdu/The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, University of Electronic Science and Technology of China, Chengdu, China

Introduction

Major depressive disorder (MDD) is a severe mental health condition with significant implications for individuals and society as a whole [1, 2]. Epidemiological data indicate that in the United States, the lifetime prevalence of MDD is approximately 17% in men and 30% in women, with a 9% prevalence in the general adult population. Despite the rising global incidence of depression, treatment rates remain alarmingly low, with only 9.5% of patients receiving any form of treatment and a mere 0.5% achieving full treatment adherence [3]. Moreover, the prevalence of depression is also increasing among adolescents and students, presenting new challenges for disease prevention and control [4].

Currently, the clinical diagnosis of MDD primarily relies on comprehensive clinical history-taking and structured or semi-structured patient interviews, supplemented by standardized rating scales (e.g., the Hamilton Depression Rating Scale) and necessary imaging examinations to rule out organic lesions and quantify symptom severity [5]. However, these methods remain confined to phenomenological descriptions and lack specific biomarkers to assist in objective diagnosis [6]. At present, there are no widely recognized specific biomarkers applied in clinical practice, limiting the sensitivity and specificity of diagnosis and hindering early identification and personalized intervention [7]. Given the increasingly severe epidemiological trends of MDD, the development of novel biomarkers holds significant clinical importance for enhancing the objectivity and precision of clinical diagnosis.

In recent years, the substantial expansion of public databases and the rapid advancement of bioinformatics technologies have provided a novel research avenue for systematically elucidating the molecular pathogenesis of diseases and precisely identifying key molecular targets [8, 9]. The integrated application of multi-omics data and cutting-edge machine learning algorithms has enabled the rapid and accurate extraction of critical molecular information closely associated with disease onset and progression from vast biological datasets, thereby significantly deepening our understanding of the aetiology and disease progression [10]. Additionally, research in physiological signal processing and multimodal fusion technology has also provided new ideas for the diagnosis and classification of diseases such as MDD. For instance, the multimodal large language learning model (LLM) based on electroencephalogram (EEG) and physiological signals, constructed by Shen et al., can be used for the identification and intervention of mental disorders (such as MF2-Net and WDANet et al.) [11–14]. It can also greatly assist in clinical diagnosis and disease management.

In this study, we employed a multi-faceted analytical approach, including differential gene screening, Weighted Gene Co-expression Network Analysis (WGCNA), and the Summary-data-based Mendelian Randomization (SMR) method, to thoroughly investigate the pathological mechanisms of MDD [15]. Our primary objective was to precisely identify key genes directly implicated in disease pathogenesis and systematically evaluate their potential as biomarkers and therapeutic targets. Initially, we comprehensively collected and meticulously analysed RNA sequencing datasets related to MDD from the GEO database, successfully screening a series of differentially expressed genes (DEGs) through gene expression differential analysis. Subsequently, the WGCNA approach was utilised to further pinpoint the most disease-relevant candidate genes [16]. Concurrently, a protein-protein interaction (PPI) network was constructed to precisely identify the hub genes within it. To develop a highly accurate diagnostic model, we employed 113 machine learning algorithms to conduct a comprehensive evaluation of the diagnostic performance of the candidate genes, accompanied by Receiver Operating Characteristic (ROC) curve analysis [17]. Furthermore, we systematically validated the causal relationship between the key genes and MDD risk by integrating colocalisation analysis with the SMR method.

This study innovatively combined bioinformatics analysis with causal inference methods, not only clearly elucidating the central role of the SERPING1 gene in the pathological progression of MDD but also robustly confirming its strong association with an increased risk of developing the disorder. These significant findings provided a fresh perspective for deepening our understanding of pathogenesis and lay a solid theoretical foundation for the development of early diagnostic tools based on key genes and the formulation of clinical intervention strategies.

Methods and materials

Integration of GEO datasets and analysis of DEGs

RNA-sequencing datasets pertinent to MDD (GSE52790 and GSE98793) were sourced from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The study cohorts encompassed 74 normal control specimens and 140 MDD specimens, with data acquisition finalized on December 5, 2025. Initial data normalization was executed employing the "limma" and "sva" packages within the R programming environment (as depicted in Figure S1). Following this, differential expression analysis was carried out using the "limma" and "DESeq2" packages. This analytical process led to the identification of 1,847 DEGs. For the purpose of visualizing the analytical outcomes, the "pheatmap" and "ggplot2" packages were utilized. Main parameter settings of differential expression analysis: log2 fold change (log2FC) > 0.1 and false discovery rate (FDR) < 0.05.

Weighted gene co-expression network analysis of DEGs

The integrated dataset underwent WGCNA via the application of the "limma" and "WGCNA" packages within the R software environment (Main parameter settings: MEDissThres = 0.25, minModuleSize = 60.). Through this analytical approach, disease-associated gene modules were successfully identified, from which a gene set demonstrating the utmost relevance to the disease was extracted. Subsequently, the intersection between this disease-related gene set, comprising 299 genes, and the set of DEGs (1,847 genes in total) was determined. This intersection analysis yielded 170 pivotal genes. To visually represent the overlapping gene subsets, a Venn diagram was generated utilizing R.

KEGG and GO enrichment analyses

To gain deeper insights into the potential biological roles of the 170 overlapping genes, we performed GO enrichment analysis. This analysis covered three fundamental domains: biological processes (BP), cellular components (CC), and molecular functions (MF), describing the activities occurring at the molecular level. Simultaneously, we utilized the KEGG (https://www.genome.jp/kegg/) to dissect the signaling pathways potentially implicated in the overlapping targets. A p-value < 0.05 was considered statistically significant.

Construction of PPI network and identification of hub genes

A protein-protein interaction (PPI) network was meticulously constructed for the 170 overlapping genes by leveraging the STRING database (https://string-db.org/). In this process, Homo sapiens was specified as the target species, and an interaction score threshold of 0.9 (highest confidence). The generated network was subsequently imported into Cytoscape for visualization and scoring. The cytoHubba plugin integrated within Cytoscape was employed to pinpoint hub genes. The top 20 scoring targets were designated as hub genes for subsequent analyses, and the PPI network was visualised accordingly.

Construction of machine learning diagnostic model and ROC curve

To identify the most effective diagnostic model, we systematically trained the 20 identified hub genes using a comprehensive panel of 113 distinct machine learning diagnostic models (Detailed information of 113 algorithms was listed in Table S2). These "113 machine learning models" refer to 113 distinct algorithms used to screen for key gene sets that can distinguish between diseases and healthy conditions. This ensemble encompassed a diverse range of model types, including glmBoost + SVM, Ridge regression, Lasso + SVM, et al. The training dataset was amalgamated from two GEO datasets (GSE52790 and GSE98793), and during the calculation of the internal AUC of the training data, cross-validation was conducted after feature selection and scaling (as shown in the train module in Fig. 3). In addition, two independent datasets (GSE44593 and GSE54564) were used for external CV and model evaluation [18–21]. As illustrated in the GSE44593 module and GSE54564 module in Fig. 3. All statistical analyses and visualisations were conducted using R, incorporating packages such as "mixOmics", "survcomp", and "ComplexHeatmap". The Lasso + RF model emerged as the optimal diagnostic model, with the specific genes included in the model detailed in Table S1. Besides, we conducted the DeLong test on the top 5 machine learning models to conduct a comparison among these models. Subsequently, differential expression bar plots for the genes within the optimal model were generated using the "reshape2", "ggpubr", and "PerformanceAnalytics" packages in R. ROC curves for the genes in the optimal model were constructed utilising the "glmnet" and "pROC" packages.

Colocalisation analysis

Exposure data were retrieved from the GWAS database (https://gwas.mrcieu.ac.uk/), while outcome data (finngen_R12_F5_DEPRESSION_DYSTHYMIA.gz) were obtained from the FinnGen database (https://www.finngen.fi/en/access_results) [22]. Bayesian colocalisation analysis was subsequently performed to assess the consistency of causal variants between the exposure and outcome. This analysis evaluated five competing causal hypotheses: H0 (no association), H1 (exposure-specific effect), H2 (outcome-specific association), H3 (pleiotropy with variant heterogeneity), and H4 (shared causal variant). A posterior probability for H4 (PPH4) exceeding the threshold of 0.7 was considered statistically significant, indicating colocalization through a shared pathogenic mechanism [23]. The analysis was conducted using the "COLOC" package.

SMR analysis

The SMR platform was implemented on a Linux operating system for analytical purposes. Heterogeneity of experimental variables was evaluated using the HEIDI test. For causal inference analysis, eQTLs were utilized as exposure variables, supplemented with GWAS results obtained from the Finn database. The analysis was conducted using default SMR parameters, incorporating false discovery rate (FDR)-corrected multiplicity adjustment ($\alpha < 0.05$) alongside HEIDI-based pleiotropy screening (empirical p-value < 0.05).

Results

Identified DEGs related to MDD

We initially analyzed RNA-sequencing datasets from the Gene Expression Omnibus database to identify DEGs between MDD and control tissues. As shown in Fig. 1A and B, distinct gene expression profiles were observed between MDD and control groups. A total of 1,847 DEGs were identified, with a log2 fold change (log2FC) > 0.1 and false discovery rate (FDR) < 0.05 (Fig. 1A and B) [24]. Among these, 799 genes exhibited upregulation, while 1,048 genes showed down-regulation in MDD-affected tissues (Fig. 1B).

WGCNA identified key modules among DEGs that exhibited close associations with MDD

As shown in Fig. 1C and D, the MEyellow, MEblue, MEgrey, and MEbrown modules demonstrated significant positive correlations with MDD ($R > 0.5$, $p < 0.05$). Additionally, the distribution of gene significance values showed significant variation across these modules ($p < 0.05$). Subsequent intra-module analysis revealed a strong positive correlation between module membership and gene significance in the MEgrey module ($R = 0.7$, $p < 0.05$) (Fig. 1E). Consequently, the 299 genes comprising the MEgrey module were designated as MDD-related genes for subsequent analyses.

Substantial overlap between key modules indicates a close relationship between the biological functions and pathways associated with MDD

To identify shared genetic components, the intersection between the WGCNA modules and the DEGs with significant positive correlations was determined. A total of 170 genes were identified as common to both the MEgrey module of WGCNA and the DEG set (Fig. 1F). Subsequently, these intersecting genes underwent GO and KEGG enrichment analyses. GO enrichment analysis revealed that the 170 genes were significantly enriched for biological processes including defense response to virus, secretory granule lumen, immune receptor activity, cell killing, and cytoplasmic vesicle lumen (Fig. 2A). KEGG pathway analysis demonstrated that these genes were predominantly involved in pathways such as phagosome formation, antigen processing, and transcriptional misregulation in cancer signaling (Fig. 2B).

Integrate the analysis of the PPI network with machine learning to select the optimal diagnostic model for MDD

The Degree algorithm in Cytoscape was employed to construct an interactive network of genetically dominant PPIs, identifying the 20 most influential key genes. Machine learning-based integration of the expression profiles of these 20 genes was performed to filter key genes associated with MDD (Figs. 2C and D and 3). In this study, a total of 113 predictive classification machine learning models were applied to both training and testing datasets using a cross-validation framework. Among these, the Lasso + RF algorithm achieved the highest AUC score of 0.73 in both the training and testing groups, thereby being identified as the most effective classification model (Fig. 3). Consequently, a set of characteristic genes (as detailed in Table S1) were identified using the Lasso + RF algorithm. Furthermore, the DeLong test for the first 5 machine learning models revealed that the Lasso + RF model was significantly superior to the other models ($p = 0.023$) (Table S3). This further indicated that the Lasso + RF model was the optimal machine learning model. The expression levels of these genes in the best diagnostic model were compared between the MDD group and the control group (Fig. 4A). Additionally, ROC curve analysis of the model genes (Fig. 4B) revealed that the areas under the curves for most genes (including MMP8, SERPING1, SIGLEC1, CXCL10, MPO, and IFI27) exceeded 0.75, suggesting their potential utility in distinguishing healthy individuals from MDD patients.

The elevated level of SERPING1 is positively correlated with the risk of the MDD

To investigate the potential of genetic biomarkers associated with genetic variation, the SMR analysis was conducted using eQTL data and GWAS data from the FinnGen database, followed by HEIDI test analysis (Fig. 5). Additionally, after applying Bonferroni correction and performing colocalization analysis, the SERPING1 gene was identified as being significantly associated with MDD risk (Fig. 6). Subsequently, we performed an intersection analysis between the genes selected from the optimal diagnostic model and those identified through SMR analysis, ultimately identifying the intersecting gene SERPING1 (Fig. 7A, B). Furthermore, the scatter plot generated from the SMR analysis demonstrated a negative correlation between SERPING1 expression and MDD risk ($P < 0.05$) (Fig. 7C). Consequently, elevated levels of SERPING1 gene expression were found to be associated with an increased risk of developing MDD.

Discussion

This work sought to integrate and apply multiple bioinformatics methodologies to investigate the pathological mechanisms underlying MDD and identify potential biomarkers. By employing a comprehensive analytical framework that incorporated differential gene expression screening, WGCNA, machine learning algorithms, and SMR analysis, we systematically identified SERPING1 as a pivotal molecular in MDD pathogenesis. The diagnostic potential of SERPING1 for MDD was further characterized through ROC curve analysis.

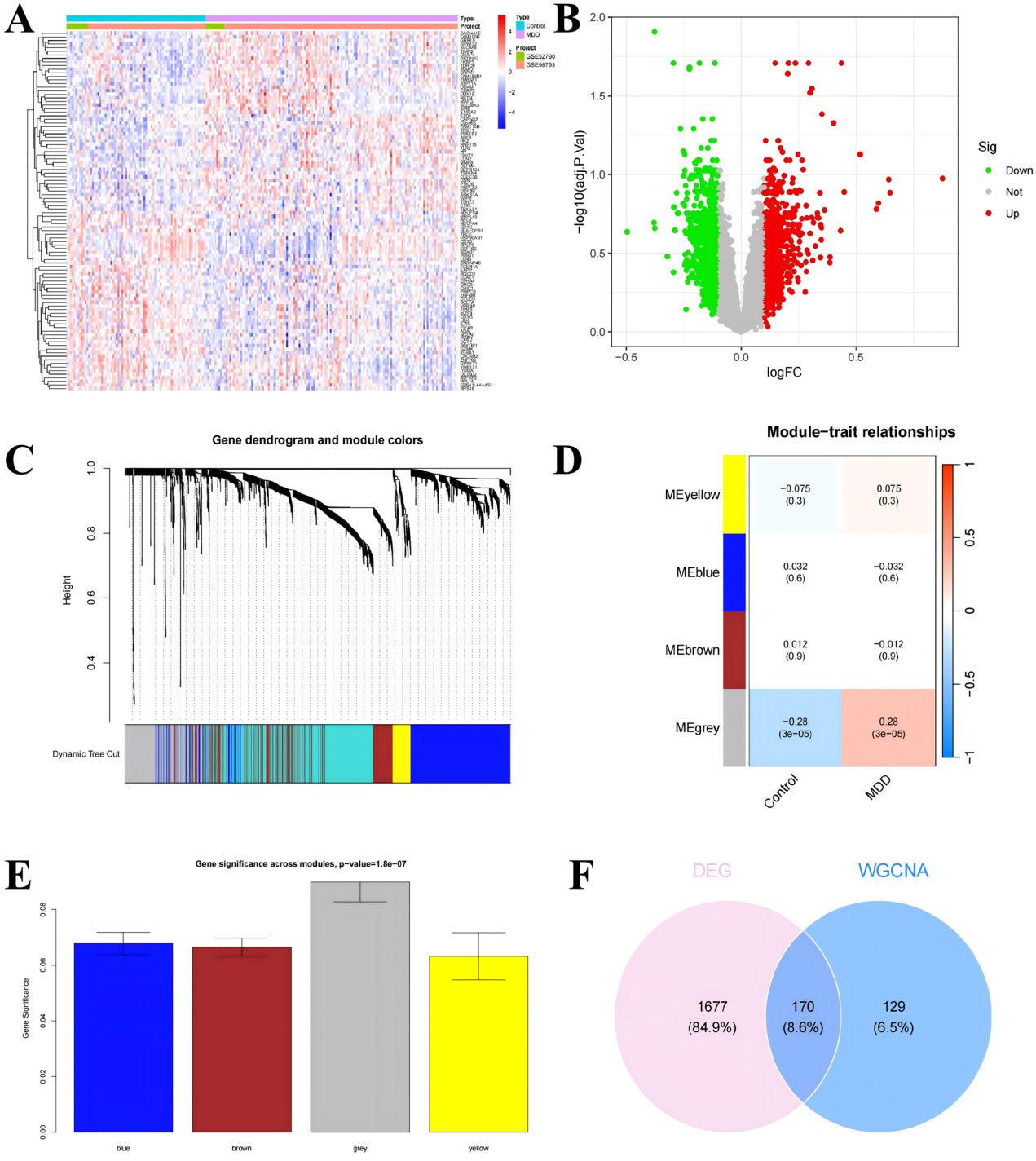

Fig. 1 Identification of pivotal genes differentiating patients with Major Depressive Disorder from healthy controls. (**A**) Heatmap depicting the differentially expressed genes (DEGs) between the healthy control group and the MDD group. (**B**) Cluster heatmap illustrating the expression patterns of DEGs in both the MDD group and the control group; the outcomes derived from differential expression analysis and WGCNA are employed to elucidate the molecular underpinnings of Major Depressive Disorder. (**C**) Threshold of WGCNA analysis. (**D**) and (**E**) Visual representation of the relationships between gene modules and phenotypic traits. Each cell within the diagrams presents the respective correlation coefficient along with its associated p-value. (**F**) Venn diagram showcasing the overlap between DEGs and genes within co-expressed modules

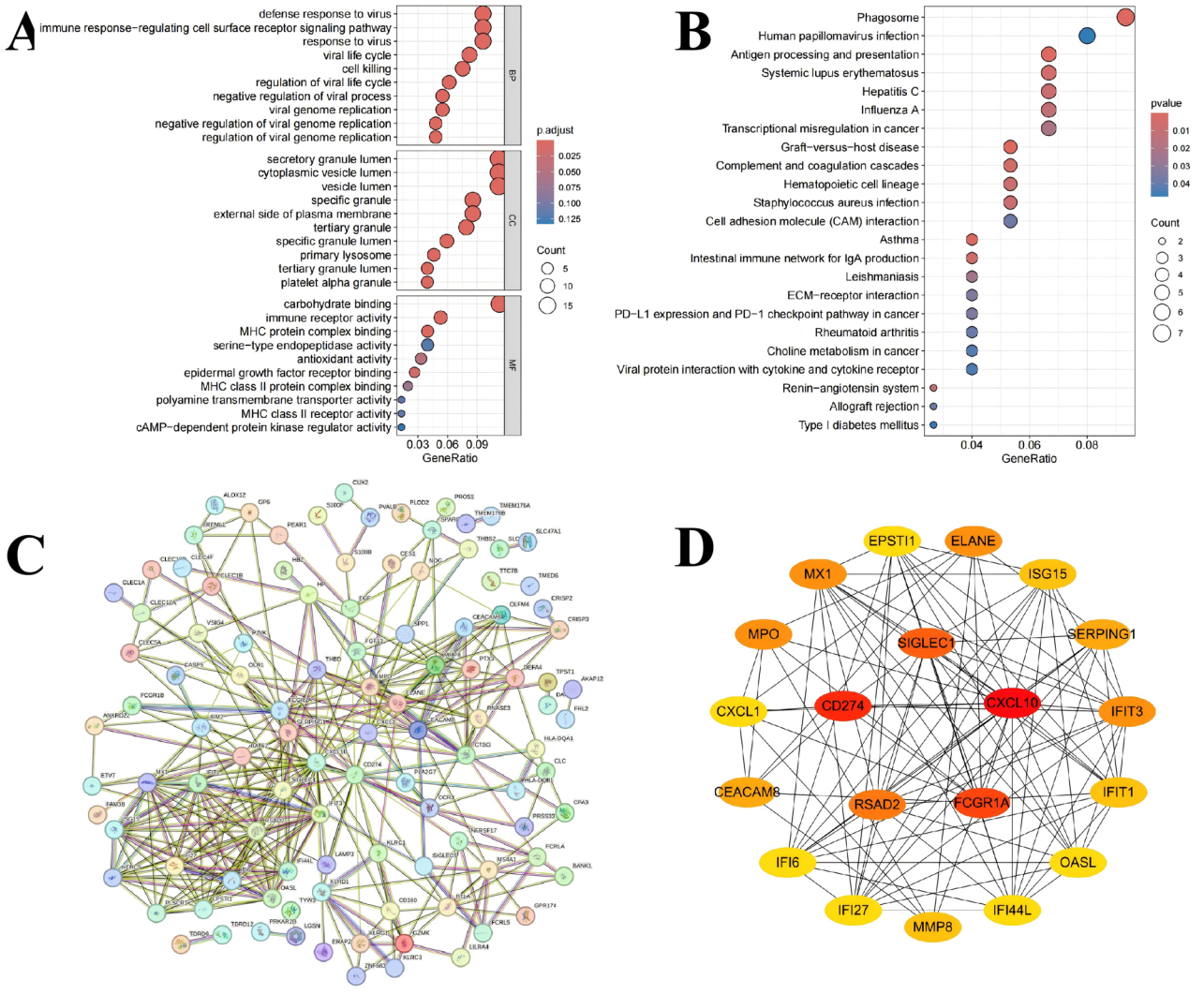

Fig. 2 Enrichment analysis of DEGs coupled with subsequent screening for hub genes from the pool of differential genes. (**A**) Enrichment analysis based on Gene Ontology (GO). (**B**) Enrichment analysis utilizing the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. (**C**) Protein-protein interaction (PPI) network constructed from overlapping genes found in both WGCNA modules and significantly up-regulated DEG modules. (**D**) Identification of the top 20 pivotal genes within the PPI network

The integrated bioinformatics methodologies employed in this study were designed to simultaneously capitalize on their respective significant advantages. First, this modular screening approach, which combines WGCNA, PPI analysis, and SMR technology, mitigates the inherent randomness associated with single-gene analysis. Furthermore, a rigorous cross-validation framework was established through the implementation of 113 distinct machine learning models, thereby enhancing diagnostic specificity. Additionally, by utilizing genetic variations as instrumental variables, confounding environmental factors were effectively controlled, enabling the establishment of robust causal inference between genetic loci and disease manifestations. The comprehensive application of these diverse analytical techniques effectively addresses the limitations inherent in any single methodology while providing novel insights into the pathophysiological mechanisms of MDD.

MDD (as a prevalent psychiatric condition) is characterized by a complex pathophysiology involving dysregulated molecular networks [25]. Extensive research has demonstrated that the pathogenesis of MDD is closely associated with dysfunctions in neurotransmitter systems and impairments in neurotrophic factor signaling pathways et al. [26]. The intricate interplay among these pathological mechanisms presents substantial challenges for the early diagnosis of MDD. The findings from this study not only advance our comprehension of the molecular underpinnings of MDD but also furnish a critical foundation for the development of novel diagnostic approaches and therapeutic targets.

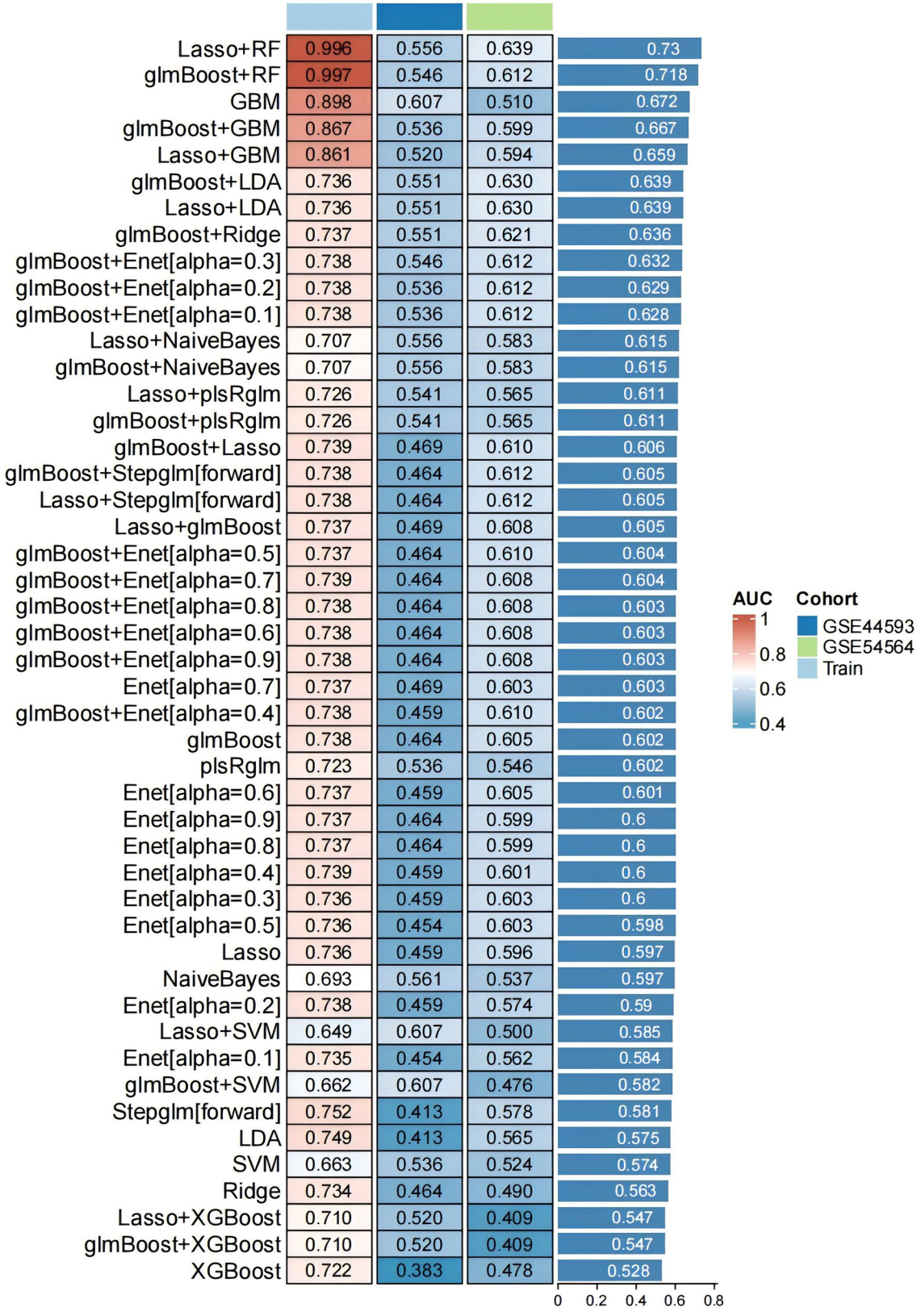

Fig. 3 A total of 113 combinations of machine learning algorithms were assessed utilizing a 10-fold cross-validation approach

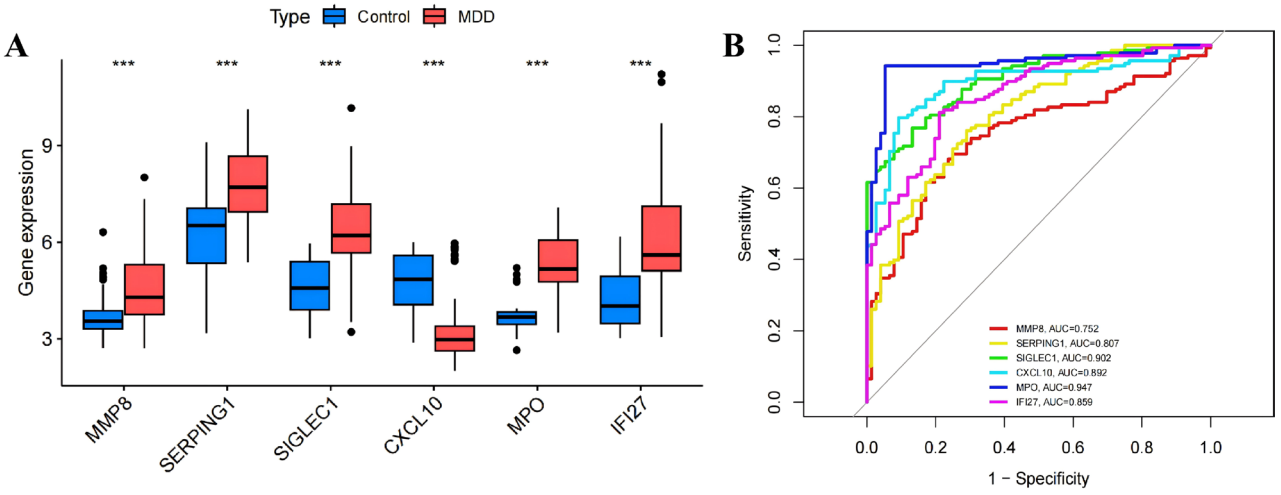

Fig. 4 ROC curve analysis and expression profiles of characteristic genes. (**A**) Receiver operating characteristic (ROC) curves for six characteristic genes across the training set, test sets, and external validation set. (**B**) Expression levels of the 6 characteristic genes in the external validation sets

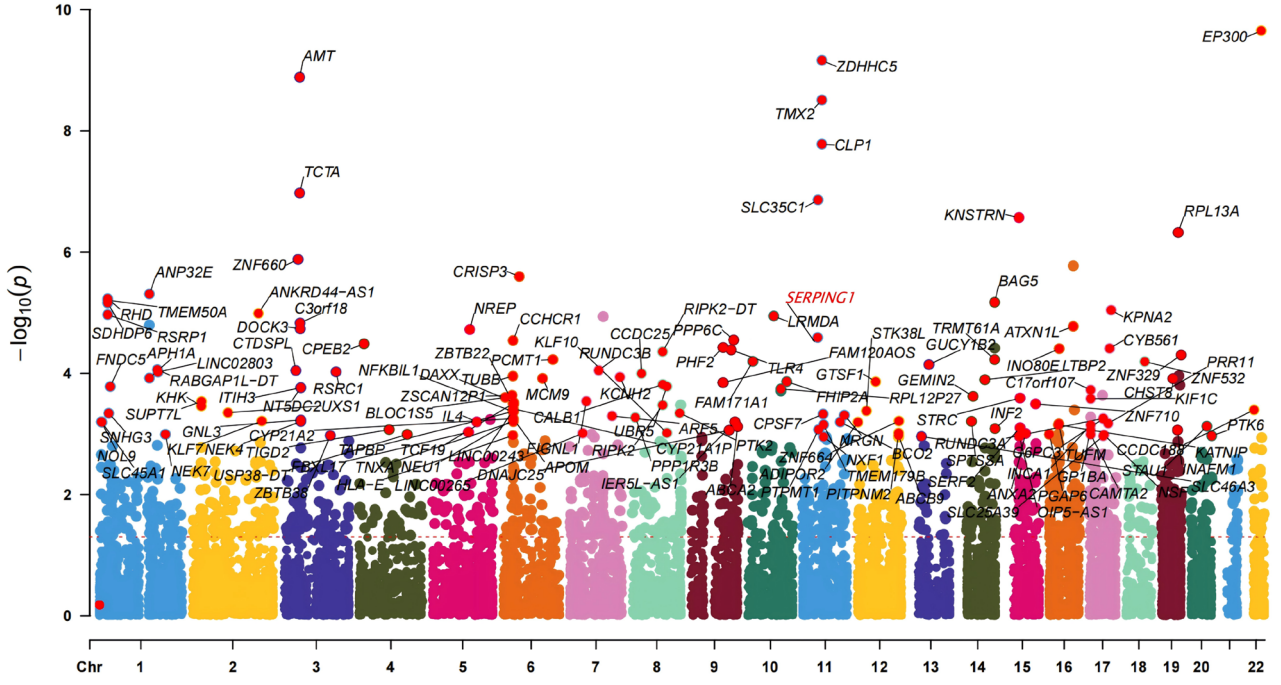

Fig. 5 SMR analysis based on eQTL data and large-scale GWAS data (Finngen database). Manhattan plot showing genes associated with Major Depressive Disorder

The SERPING1 gene, located on chromosome 11, encodes C1 esterase inhibitor (C1-INH), a pivotal serine protease inhibitor that plays a critical regulatory role in the complement system, kinin system, and coagulation cascade [27–29]. The protein it encodes regulates immune homeostasis by modulating serine proteases such as C1s and C1r [30, 31]. In neurodegenerative diseases, oxidative stress and neuroinflammation are central pathogenic mechanisms. SERPING1 may participate in disease progression by modulating the complement system and thereby influencing neuroinflammatory responses. SERPING1 may participate in disease development by modulating the complement system, thereby influencing neurotoxic inflammatory responses [32]. Research by Amal Bouzid et al. identified LCN2, OLFM4, SERPING1, TCN1, and THBS1 as potential diagnostic biomarkers for MDD, capable of predicting depressive episodes [33]. The SERPING1 gene may also be a risk gene for schizophrenia [28].

Within the central nervous system, Serping1 expression is localised to astrocytes [32, 34]. In acute central nervous system injury and chronic neurodegenerative diseases,

A

ID	Symbol	PP.H0	PP.H1	PP.H2	PP.H3	PP.H4
eqtl-a-ENSG00000149131	SERPING1	6.82E-248	0.037334528	1.25E-247	0.067567178	0.895098294
eqtl-a-ENSG00000005381	MPO	6.05E-288	0.985676976	3.84E-290	0.0062443	0.008078724
eqtl-a-ENSG00000088827	SIGLEC1	0.75952392	0.166805218	0.055490075	0.012180626	0.006000161
eqtl-a-ENSG00000118113	MMP8	2.06E-21	0.984128068	2.08E-23	0.009912649	0.005959283
eqtl-a-ENSG00000169245	CXCL10	1.81E-113	0.98908866	1.43E-115	0.007817064	0.003094276
eqtl-a-ENSG00000165949	IFI27	2.48E-167	0.989558397	1.85E-169	0.007360172	0.003081431

B

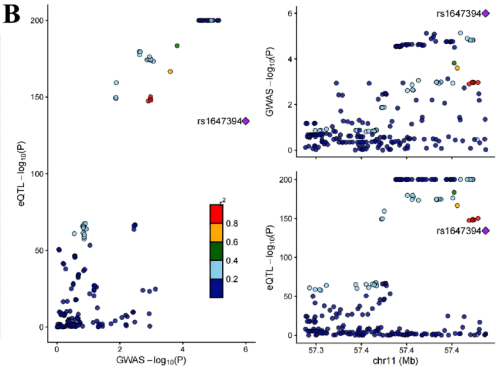

Fig. 6 The co-localization analysis of genes from the optimal diagnostic model and Major Depressive Disorder. (**A**) Results of co-localization analysis between genes in the optimal diagnostic model (derived from eQTL analysis) and Major Depressive Disorder (based on GWAS data). (**B**) Genomic map illustrating the co-localized regions between SERPING1 and Major Depressive Disorder

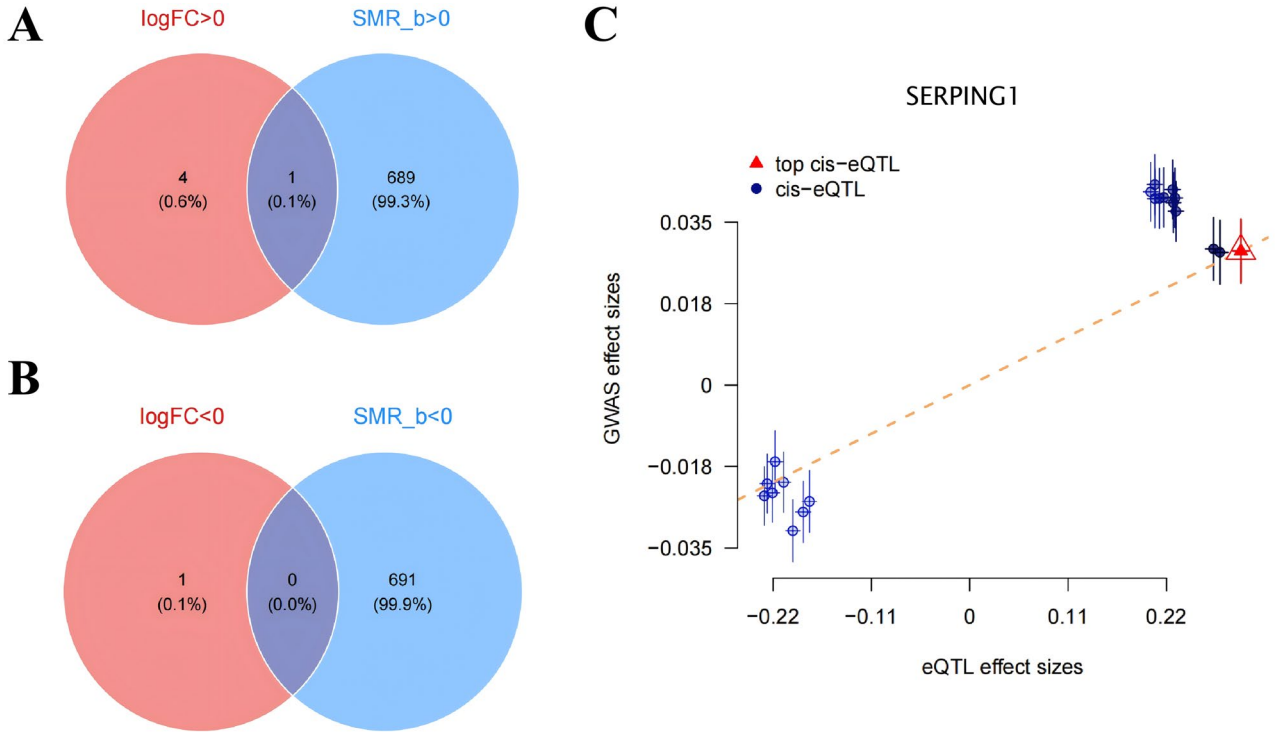

Fig. 7 Identification of key genes from diagnostic model genes and SMR analysis. (**A**) Overlap between up-regulated model genes and high-risk genes. (**B**) Overlap between down-regulated model genes and low-risk genes. (**C**) Scatter plot from SMR analysis demonstrating a positive association between SERPING1 expression and the risk of Major Depressive Disorder

neuronal damage coincides with a reactive astrocyte response in the human body [35]. Furthermore, astrocyte dysfunction, following dopaminergic neuronal loss and α-synuclein accumulation, increases the expression of proteins including SerpinG1, complement component C3, PSMB8, and GBP2 [36]. Additionally, the SERPING1 gene has been associated with immune cell infiltration and may drive the hypoxic phenotype in necrotic glioblastoma, thereby influencing hypoxia-induced glioma stem cell properties [37].

While this study demonstrates the robust diagnostic performance of SERPING1 (AUC = 0.807), its specificity as a biomarker requires further validation. Future research should incorporate larger and more diverse datasets to clarify the clinical utility of SERPING1 in the early diagnosis of MDD. However, our findings are subject to several limitations. First, the restricted sample size of the MDD cohort derived from the GEO database may limit the generalizability of the results. Second, the observations are primarily derived from computational

bioinformatics analyses and thus require validation through direct clinical and experimental studies.

Conclusion

This study, employing an integrated approach combining bioinformatic analysis with multiple machine learning algorithms, identified significant underexpression of the SERPING1 gene in MDD, suggesting its potential utility as a diagnostic biomarker. Furthermore, SMR analysis utilizing eQTL data demonstrated a significant positive correlation between SERPING1 expression levels and MDD risk. These findings highlight the potential of SERPING1 as a biomarker for MDD, offering novel perspectives for early diagnosis and advancing related research avenues.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12888-026-07956-8.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Acknowledgements
We sincerely express our gratitude to the Fourth People's Hospital of Chengdu.

Author contributions
All authors made a significant contribution to the work. Xiaokui Yuan and Tong Wang: Data curation, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. All authors have contributed to and have approved the final manuscript.

Funding
This work received no funding.

Data availability
All data generated or analyzed during this study are included in this published article.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 23 December 2025 / Accepted: 4 March 2026
Published online: 07 March 2026

References

1. Shorter JR, Pasman JA, Kurvits S, Jangmo A, Naamanka J, Harder A, Hagen E, Kowalec K, Frilander N, Zetterberg R, et al. Genome-wide association analyses identify distinct genetic architectures for early-onset and late-onset depression. Nat Genet. 2025.
2. Pillinger T, Arumuham A, McCutcheon RA, D'Ambrosio E, Basdanis G, Branco M, Carr R, Finelli V, Furukawa TA, Gee S, et al. The effects of antidepressants on cardiometabolic and other physiological parameters: a systematic review and network meta-analysis. Lancet. 2025;406(10515):2063–77.
3. Dwyer JB, Aftab A, Radhakrishnan R, Widge A, Rodriguez CI, Carpenter LL, Nemeroff CB, McDonald WM, Kalin NH. Hormonal Treatments for Major Depressive Disorder: State of the Art. Am J Psychiatry. 2020;177(8):686–705.
4. Casella CB, Kessler RC, Mneimneh Z, Merikangas KR, Polanczyk GV, Salum GA. Emotional distress in adolescents in 2018 and 2022: a comparison of cross-sectional national probabilistic samples from six countries. Lancet Child Adolesc Health. 2026;10(1):39–48.
5. Quataert S, Martin D, Anderson P, Giebink GS, Henrichsen J, Leinonen M, Granoff DM, Russell H, Siber G, Faden H, et al. A multi-laboratory evaluation of an enzyme-linked immunoassay quantitating human antibodies to Streptococcus pneumoniae polysaccharides. Immunol Invest. 2001;30(3):191–207.
6. Kennis M, Gerritsen L, van Dalen M, Williams A, Cuijpers P, Bockting C. Prospective biomarkers of major depressive disorder: a systematic review and meta-analysis. Mol Psychiatry. 2020;25(2):321–38.
7. Dadkhah M, Jafarzadehgharehziaaddin M, Molaei S, Akbari M, Gholizadeh N, Fathi F. Major depressive disorder: Biomarkers and biosensors. Clin Chim Acta. 2023;547:117437.
8. Yu N, Li W, Sun X, Hu J, Zou Q, Liu Z, Zhang D, Zhang W, Gao R. soFusion: facilitating tissue structure identification via spatial multi-omics data fusion. Brief Bioinform. 2025;26(5).
9. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20(8):467–84.
10. Hu Y, Liu Y, Mao X, Jia C, Ferguson JF, Xue C, Reilly MP, Li H, Li M. PennSeq: accurate isoform-specific gene expression quantification in RNA-Seq by modeling non-uniform read distribution. Nucleic Acids Res. 2014;42(3):e20.
11. Shen J, You L, Ma Y, Zhao Z, Liang H, Zhang Y, Hu B. UA-DAAN: An Uncertainty-Aware Dynamic Adversarial Adaptation Network for EEG-Based Depression Recognition. Affect Comput IEEE Trans on. 2025;16(3):2130–41.
12. Shen J, Zhang X, Wang G, Ding Z, Hu B. An Improved Empirical Mode Decomposition of Electroencephalogram Signals for Depression Detection. Affective Computing, IEEE Trans on (T-AFFC). 2022, 13(1).
13. Shen J, Zhang Y, Liang H, Zhao Z, Dong Q, Qian K, Zhang X, Hu B. Exploring the Intrinsic Features of EEG Signals via Empirical Mode Decomposition for Depression Recognition. IEEE Trans Neural Syst Rehabil Eng. 2023;31:356–65.
14. Shen J, Zhu K, Ma R, et al. EmoSavior: Depression recognition and intervention via multimodal physiological signals and large language models. Inform Fusion. 2025: 103772.
15. Sun X, Chen B, Qi Y, Wei M, Chen W, Wu X, Wang Q, Li J, Lei X, Luo G. Multi-omics Mendelian randomization integrating GWAS, eQTL and pQTL data revealed GSTM4 as a potential drug target for migraine. J Headache Pain. 2024;25(1):117.
16. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
17. Wu Z, Yu B, He Q, Huang C. LCN2 drives ferroptosis-associated ischemia-reperfusion injury after renal transplantation: integrated machine learning and in vivo validation. Apoptosis. 2025;30(11–12):3026–43.
18. Chen Z, Tang X, Gu C, Zou S. Investigating biomarkers of mitochondrial and aging-related genes in major depressive disorder through bioinformatics analysis. Front Psychiatry. 2025;16:1653998.
19. Wang H, Dou S, Wang C, Gao W, Cheng B, Yan F. Identification and Experimental Validation of Parkinson's Disease with Major Depressive Disorder Common Genes. Mol Neurobiol. 2023;60(10):6092–108.
20. Wu Z, Cai Z, Shi H, Huang X, Cai M, Yuan K, Huang P, Shi G, Yan T, Li Z. Effective biomarkers and therapeutic targets of nerve-immunity interaction in the treatment of depression: an integrated investigation of the miRNA-mRNA regulatory networks. Aging. 2022;14(8):3569–96.
21. Chen N, Luan Y. Specific expression and common potential therapeutic drugs in different brain regions of major depressive disorder patients: bioinformatics analysis. J Affect Disord. 2025;382:478–87.
22. Xie L, Bai L, Wang T, Deng S. Shared genetic architecture between anxiety, depression and erectile dysfunction: a genome-wide cross-trait analysis. Transl Androl Urol. 2025;14(9):2522–36.
23. Hashimoto M, Kojima Y, Sakamoto T, Ozato Y, Nakano Y, Abe T, Hosoda K, Saito H, Higuchi S, Hisamatsu Y, et al. Spatial and single-cell colocalisation analysis reveals MDK-mediated immunosuppressive environment with regulatory T cells in colorectal carcinogenesis. EBioMedicine. 2024;103:105102.

24. Sun P, Feng S, Yu H, Wang X, Fang Y. Two hub genes of bipolar disorder, a bioinformatics study based on the GEO database. IBRO Neurosci Rep. 2024;17:122–30.
25. Rappeneau V, Wilmes L, Touma C. Molecular correlates of mitochondrial dysfunctions in major depression: Evidence from clinical and rodent studies. Mol Cell Neurosci. 2020;109:103555.
26. Chan WL, Olbrich S, Jiang X, Zhang H, Ip CT, Brunovsky M. Subanesthetic ketamine alters EEG signal complexity: Implications for treatment stratification in depression. J Affect Disord. 2025;386:119477.
27. Haslund D, Ryø LB, Seidelin Majidi S, Rose I, Skipper KA, Fryland T, Bohn AB, Koch C, Thomsen MK, Palarasah Y, et al. Dominant-negative SERPING1 variants cause intracellular retention of C1 inhibitor in hereditary angioedema. J Clin Invest. 2019;129(1):388–405.
28. Maaser A, Forstner AJ, Strohmaier J, Hecker J, Ludwig KU, Sivalingam S, Streit F, Degenhardt F, Witt SH, Reinbold CS, et al. Exome sequencing in large, multiplex bipolar disorder families from Cuba. PLoS ONE. 2018;13(10):e0205895.
29. Chen S, Li LY, Wu ZM, Liu Y, Li FF, Huang K, Wang YX, Chen QJ, Wang XQ, Shen WF, et al. SerpinG1: A Novel Biomarker Associated With Poor Coronary Collateral in Patients With Stable Coronary Disease and Chronic Total Occlusion. J Am Heart Assoc. 2022;11(24):e027614.
30. Ponard D, Gaboriaud C, Charignon D, Ghannam A, Wagenaar-Bos IGA, Roem D, López-Lera A, López-Trascasa M, Tosi M, Drouet C. SERPING1 mutation update: Mutation spectrum and C1 Inhibitor phenotypes. Hum Mutat. 2020;41(1):38–57.
31. Grymová T, Grodecká L, Souček P, Freiberger T. SERPING1 exon 3 splicing variants using alternative acceptor splice sites. Mol Immunol. 2019;107:91–6.
32. Stoll AC, Kemp CJ, Patterson JR, Howe JW, Steece-Collier K, Luk KC, Sortwell CE, Benskey MJ. Neuroinflammatory gene expression profiles of reactive glia in the substantia nigra suggest a multidimensional immune response to alpha synuclein inclusions. Neurobiol Dis. 2024;191:106411.
33. Bouzid A, Almidani A, Zubrikhina M, Kamzanova A, Ilce BY, Zholdassova M, Yusuf AM, Bhamidimarri PM, AlHaj HA, Kustubayeva A, et al. Integrative bioinformatics and artificial intelligence analyses of transcriptomics data identified genes associated with major depressive disorders including NRG1. Neurobiol Stress. 2023;26:100555.
34. Peng X, Ju J, Li Z, Liu J, Jia X, Wang J, Ren J, Gao F. C3/C3aR Bridges Spinal Astrocyte-Microglia Crosstalk and Accelerates Neuroinflammation in Morphine-Tolerant Rats. CNS Neurosci Ther. 2025;31(1):e70216.
35. Das S, Li Z, Noori A, Hyman BT, Serrano-Pozo A. Meta-analysis of mouse transcriptomic studies supports a context-dependent astrocyte reaction in acute CNS injury versus neurodegeneration. J Neuroinflammation. 2020;17(1):227.
36. Weiss F, Hughes L, Fu Y, Bardy C, Halliday GM, Dzamko N. Astrocytes contribute to toll-like receptor 2-mediated neurodegeneration and alpha-synuclein pathology in a human midbrain Parkinson's model. Transl Neurodegener. 2024;13(1):62.
37. Xiao K, Tan J, Yuan J, Peng G, Long W, Su J, Xiao Y, Xiao Q, Wu C, Qin C, et al. Prognostic value and immune cell infiltration of hypoxic phenotype-related gene signatures in glioblastoma microenvironment. J Cell Mol Med. 2020;24(22):13235–47.

Publisher's note