Journal of Translational Medicine

RESEARCH

Open Access

From inflammation to depression: key biomarkers for IBD-related major depressive disorder

Chaoqun Hu[1†], Mei Ge[2†], Yan Liu[1], Wei Tan[1], Yingzhi Zhang[1], Min Zou[1], Lingya Xiang[1], Xiaomei Song[1*] and Hong Guo[1*]

Abstract

Background Inflammatory bowel disease (IBD) is a chronic, inflammatory, and autoimmune disorder, and its incidence of comorbid with major depressive disorder (MDD) is significantly higher than the general population. However, many patients lack proper recognition and necessary psychological health treatments. We aimed to identify potential biomarkers and mechanisms involved in the development of IBD comorbid with MDD (IBD-MDD).

Methods We utilized IBD and MDD-related datasets from the GEO database for differential gene expression analysis, protein-protein interaction (PPI) and pathway enrichment analysis, random forest algorithm, LASSO regression analysis, and construction of a disease prediction model. We assessed the accuracy of the model using ROC curve, explored potential mechanisms through immune infiltration analysis, and validated candidate biomarkers using peripheral blood samples from patients in our center's cohort.

Results We identified 484 IBD-related secreted proteins and 142 key module genes associated with MDD. PPI analysis revealed two crucial modules primarily involved in inflammation and immune regulation. We identified four diagnostic genes (HGF, SPARC, ADAM12, and MMP8) from the 21 shared genes between IBD-related secreted proteins and MDD key module genes, constructed a nomogram model and confirmed its accuracy using ROC curve from an external independent dataset. Immune infiltration analysis revealed significant associations between the four diagnostic genes, and cellular immune dysregulation in MDD. Finally, we validated the expression patterns of the four diagnostic genes in our cohort.

Conclusions Our study discovered four candidate biomarkers for IBD-MDD, providing new insights for the diagnosis and therapeutic intervention of serum-based IBD comorbid with MDD.

Keywords Inflammatory bowel disease, Major depressive disorder, Immune cell infiltration, Diagnostic value, Serum secretory proteins

†Chaoqun Hu and Mei Ge contributed equally to this work.

*Correspondence:
Xiaomei Song
sxm1127@cqu.edu.cn

Hong Guo
hguo_cgh2021@163.com
[1]Department of Gastroenterology, Chongqing General Hospital, Chongqing University, Chongqing, China
[2]Department of Breast and Thyroid Surgery, Chongqing General Hospital, Chongqing University, Chongqing, China

Introduction

Inflammatory bowel disease (IBD) is a chronic non-specific inflammatory disorder that affects the ileum, colon, and rectum, primarily occurring in young and middle-aged adults [1]. The incidence of IBD is increasing globally, with a prevalence of more than 0.5% in Western countries [2]. The development of IBD is closely related to genetic susceptibility, environmental factors, intestinal microbiota, and the immune system [3, 4]. Due to the chronic nature of the disease, IBD requires lifelong treatment, which imposes significant psychological stress on patients and severely impacts their quality of life. Extensive research has been conducted on the connection between IBD and the nervous system. Studies have shown that the prevalence of mood disorders, such as depression and anxiety, is significantly higher in IBD patients compared to the general population, yet most patients do not receive the necessary psychological treatment they require [5]. A systematic review found that the incidence of depression in active IBD patients was as high as 34.7%, while it was around 19.9% in non-active IBD patients [6]. IBD patients with depressive symptoms face increased risks of disease exacerbation, treatment escalation, hospitalization, emergency department visits, and surgery. Similarly, individuals with depression are often diagnosed with IBD, and their mental symptoms correlate with adverse disease activity in IBD [7, 8]. Therefore, there is a common occurrence of comorbidity between mental disorders and gastrointestinal diseases. The concept of the "brain-gut axis" proposed in recent years partially explains the association between these comorbidities, highlighting the complex bidirectional regulation between the brain and the gastrointestinal tract, involving mechanisms such as inflammatory immune responses, the autonomic nervous system, the enteric nervous system, and the role of the gut microbiota and its metabolites [9]. The coexistence of depression in IBD patients further reduces their quality of life and increases healthcare costs and the global burden of disease.

According to the International Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10) and the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), major depressive disorder (MDD) is defined as a persistent period of two weeks or more characterized by depressed mood or loss of interest or pleasure, along with several other symptoms such as psychomotor agitation or retardation, decreased concentration, feelings of worthlessness or guilt, thoughts of death or suicide [10]. The clinical diagnosis and assessment of depression primarily rely on screening scales, interviews, and a comprehensive evaluation of the patient's clinical symptoms, lacking objective biological markers. Commonly used depression scales include the Hamilton Depression Rating Scale (HAMD), Beck Depression Inventory (BDI), among others [11]. Therefore, diagnosing depression depends on the experience of clinicians and the cooperation of patients, leading to a high rate of misdiagnosis. Furthermore, functional magnetic resonance imaging, positron emission tomography, and electroencephalography are often used as auxiliary methods in the clinical diagnosis of MDD [12]. However, these clinical examinations are costly, and their accuracy is still a matter of debate. Therefore, the search for suitable biological markers for MDD remains a focus of biological research, aiming to detect depressive symptoms early in IBD patients and provide appropriate medical interventions.

In recent decades, with the development of high-throughput sequencing technologies such as RNA-seq and microarrays, the availability of large-scale datasets and significant improvements in machine learning software and hardware have led to increased attention to precision psychiatry. Precision psychiatry aims to establish models for individual prediction, providing new means and approaches for the diagnosis of depression and improving the detection rate of MDD. Currently, the main machine learning algorithms used in diagnostic prediction include random forest (RF), naive Bayes (NB), logistic regression (LR), Gaussian mixed model (GMM), support vector machine (SVM), and decision tree (DT) [13]. Although there is a growing consensus on the best practices for precision psychiatry and machine learning, there are still some issues in MDD biomarker research that could affect its performance in real-world prediction. For example, methodological flaws in model validation and a lack of external validation may lead to an overestimation of predictive performance in many studies. Therefore, we conducted dual validation by first validating the MDD dataset in a validation cohort after establishing the diagnostic model, followed by external validation using peripheral blood samples from IBD patients with and without MDD in our center, thereby improving the predictive performance of the biomarkers. In addition, we explored the mechanisms of IBD-related MDD from the perspectives of pathway enrichment and immune cell infiltration.

Methods

Data collection and processing

We retrieved four transcriptomic datasets from the GEO database (https://www.ncbi.nlm.nih.gov/geo/), including GSE117993, GSE126124, GSE98793, and GSE39653. The details of these datasets are shown in Table 1. The IBD-related intestinal tissue expression dataset (GSE117993, GSE126124) contained 76 control samples and 192 intestinal samples from IBD patients, with 131 cases of CD and 61 cases of UC. The IBD patient peripheral blood mononuclear cell dataset GSE126124 included

Table 1 Comparison of basic demographic characteristics between IBD and IBD-MDD

Item	IBD($n=20$)	IBD-MDD($n=20$)	Test	P
Age (years. mean ± SD)	18.50(17.5–37.00)	46.00(27.50–58.50)	*z*=-1.232	0.218
Gender, (% Female)	15%	75%	$x^2=14.545$	< 0.001
BMI	21.27 ± 5.48	22.00(19.00-23.05)	*z=-0.271*	0.787
HAMD-17 total score	2.00(1.00-2.75)	19.90 ± 2.43	*z*=-5.455	< 0.001
HGB (g/L)	131.33 ± 22.70	109.80 ± 30.09	*t* = 1.953	0.058
ESR (mm/h)	10.17 ± 3.49	23.80 ± 14.34	*t*=-1.057	0.298
CRP (mg/L)	3.53(0.96–23.90)	3.22 ± 3.01	*z=-1.258*	0.208
WBC (10^9/L)	6.55 ± 1.82	6.73 ± 2.11	*t* = 1.608	0.116
NEUT (10^9/L)	4.64 ± 1.69	4.24 ± 1.54	*t* = 2.066	0.046
EO (10^9/L)	0.14 ± 0.08	0.14 ± 0.80	*t* = 0.803	0.427
PLT (10^9/L)	290.67 ± 26.37	372.80 ± 177.13	*t*=-0.404	0.689
ALB (g/L)	44.57 ± 5.22	42.10 ± 5.34	*t* = 1.167	0.250
FCP (ug/g)	444.08 ± 438.00	544.76 ± 591.33	*t* = 0.164	0.872
25-OH-VD, (nmol/L)	18.14 ± 5.53	12.85 ± 7.71	*t* = 1.452	0.158

Data was shown as number (percentage), Mean ± SD or Median (IQR). Chi-square test, Mann-Whitney test, and independent samples t-test were used in the analysis

39 control samples and 59 IBD patient peripheral blood samples. The MDD patient peripheral blood dataset GSE98793 was used as the training set, containing 64 control samples and 128 MDD patient peripheral blood samples. The MDD patient peripheral blood microarray dataset GSE39653 was used as the validation set, including 24 control samples and 21 MDD patient peripheral blood samples. Subsequently, we used the *sva* package in R software (Version 1.01) for dataset merging, batch correction, and normalization [14], and the limma package for screening differentially expressed genes (DEGs) between the two groups [15], with a screening criteria of *P* values < 0.05 and |logFC| > 0.585. Finally, the *pheatmap* package and *ggplot2* package were used to visualize the expression patterns of DEGs, generating heatmaps and volcano plots, respectively.

Secreted protein genes

We downloaded 3,946 secreted protein-coding genes from the Human Protein Atlas database (https://www.proteinatlas.org/).

Weighted gene co-expression network analysis (WGCNA)

WGCNA can identify co-expressed gene modules and explore the relationships between gene sets and phenotypes, which is useful for studying core genes in gene expression data [16]. In this study, we used the WGCNA package to construct a scale-free co-expression gene network based on the MDD peripheral blood dataset GSE98793. We selected the top 25% differentially expressed genes for WGCNA analysis, set the fitting index $R^2=0.9$, and the soft threshold $\beta=10$ to ensure the co-expression gene network maintained scale-free topology. After dynamic pruning and TOM hierarchical clustering analysis (minimodules = 100), the most important disease-related modules were identified for further analysis.

PPI and MCODE analysis

To investigate the interactions between IBD-related secreted proteins and important MDD genes, we constructed the IBD and MDD-related PPI networks based on the STRING database (https://www.string-db.org) [17] with a medium confidence score of > 0.4. We used Cytoscape software for PPI network visualization and applied the MCODE plugin in Cytoscape to calculate the network modules with the following parameters: Degree Cutoff: 2, Haircut, node score cutoff: 0.2, k-core 2, Max Depth 100. The top 2 scored modules were selected for further analysis.

Functional enrichment analysis

To elucidate the biological functions and pathways of the candidate genes related to IBD and MDD, we used the *clusterProfiler* package for GO and KEGG enrichment analysis, and the ggplot2 package for visualizing the enrichment results [18].

Machine learning

To further study the potential candidate genes for diagnosing IBD-related MDD, we used the glmnet package for the LASSO logistic regression method [19] and the randomForest package for the RF algorithm [20]. These two machine learning methods were used to narrow down the range of candidate biomarkers. Finally, the intersection of the results from the two algorithms was used as the candidate genes for diagnosis. In the training and validation sets, the pROC package was used to plot the ROC curves and calculate the AUC values of the feature genes.

Nomogram construction and model evaluation

To determine the importance of candidate genes in the diagnosis of IBD-related MDD, we used the rms package to plot the nomograms of the feature genes [21]. The nomogram consists of a "points" and a "total points" scale, the latter of which shows the total score of all genes. The nomogram is an important tool for clinical prediction of IBD-related MDD. The clinical prediction efficiency of the nomogram was then evaluated using

calibration curves and decision curve analysis (DCA). The ROC curve was used to evaluate the performance of the nomogram in predicting IBD-related MDD, and the AUC value > 0.7 was considered clinically significant. Finally, the external MDD dataset GSE39653 was used to validate the predictive efficiency of the nomogram.

Validation of feature gene expression in MDD and IBD

The feature genes obtained from the LASSO regression analysis and RF analysis were validated for their expression in MDD and IBD datasets, respectively. The ggplot2 package was used to compare and visualize the expression levels of the feature genes, and the consistently up-regulated or down-regulated genes in IBD and MDD patients were identified.

GSEA

In the MDD dataset GSE39653, we used the org.Hs.eg.db and clusterProfiler packages to perform gene set enrichment analysis for each feature gene, to compare the differential signaling pathways between MDD and healthy control groups. The Enrichplot package was used to display the top 10 activated and inhibited pathways in MDD.

Immune infiltration analysis

In the MDD dataset GSE39653, we used the CIBERSORT package [22] to evaluate the abundance of infiltrating immune cells in the disease. The Wilcoxon test was used to compare the differences in the proportions of 22 immune cell types between MDD and healthy control samples, with $p < 0.05$ considered statistically significant. The ggplot2 package was used to visualize the infiltration of the 22 immune cells. Spearman's rank correlation coefficients were used to analyze the correlation between the expression of diagnostic biomarkers and the infiltration of immune cells, with $p < 0.05$ considered statistically significant.

Collection of IBD and IBD-MDD patient samples

This study included IBD patients who visited the Department of Gastroenterology at Chongqing General Hospital from January 2024 to June 2024. The control group consisted of 20 IBD patients, and the experimental group consisted of 20 IBD patients with comorbid MDD, for a total of 40 patients. IBD-MDD patientswere diagnosed according to the DSM-5 criteria, and the severity of depressive symptoms was assessed using the 17-item Hamilton Depression Rating Scale (HAMD-17). Patients with HAMD-17 > 17 were enrolled into the IBD-MDD group. Patients with other psychiatric disorders or who had taken antidepressants within 2 weeks were excluded.

Evaluation of the diagnostic model in the external cohort

According to the manufacturer's protocol, the serum levels of HGF, SPARC, ADAM12, and MMP8 were measured in the serum samples of IBD patients with or without comorbid MDD using ELISA kits (Fine Test, Wuhan, China).

Statistical analysis

Statistical analysis was performed using SPSS software (version 26.0, Inc., USA). The normality of the data was assessed using the Shapiro-Wilk test. Continuous variables were expressed as mean ± standard deviation or median (interquartile range). Categorical variables were presented as n (%). Group differences were determined using chi-square test, Mann-Whitney test, and independent samples t-test. $P < 0.05$ was considered statistically significant.

Results

Demographic characteristics and expression levels of routine serum and fecal biomarkers

There were no significant statistical differences in the average age and Body Mass Index (BMI) between IBD with and without comorbid MDD groups included in this study. However, the proportion of females in the IBD-MDD group was significantly higher than in the IBD group. Except for neutrophil absolute count (NEUT), which was significantly higher in the IBD group compared to the IBD-MDD group, the remaining routine serum biomarkers such as hemoglobin (HGB), erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), white blood cell count (WBC), eosinophil count (EO), platelets (PLT), albumin (ALB), and 25-hydroxyvitamin D (25-OH-VD) showed no statistically significant differences between the two groups. Fecal calprotectin (FCP) also exhibited no statistically significant differences between the two groups (Table 1).

Screening of differentially expressed genes (DEGs) in IBD and identification of secreted proteins

The analysis workflow of this study is shown in Fig. 1. Two raw data sets were collected from the GEO database, including samples from IBD patients and healthy controls. After sample merging, batch correction, and normalization, a total of 268 intestinal tissue samples (76 healthy controls vs. 192 IBD) and 98 peripheral blood samples (59 healthy controls vs. 39 IBD) were obtained. The limma package was used to screen for DEGs between the two groups, with the criteria of $p < 0.5$ and $|\log2FC| > 0.585$. The results showed that there were 1,403 DEGs in the intestinal tissue samples, including 830 upregulated and 573 downregulated genes. In the peripheral blood samples, 143 DEGs were identified, including 110 upregulated and 33 downregulated genes. Volcano plots

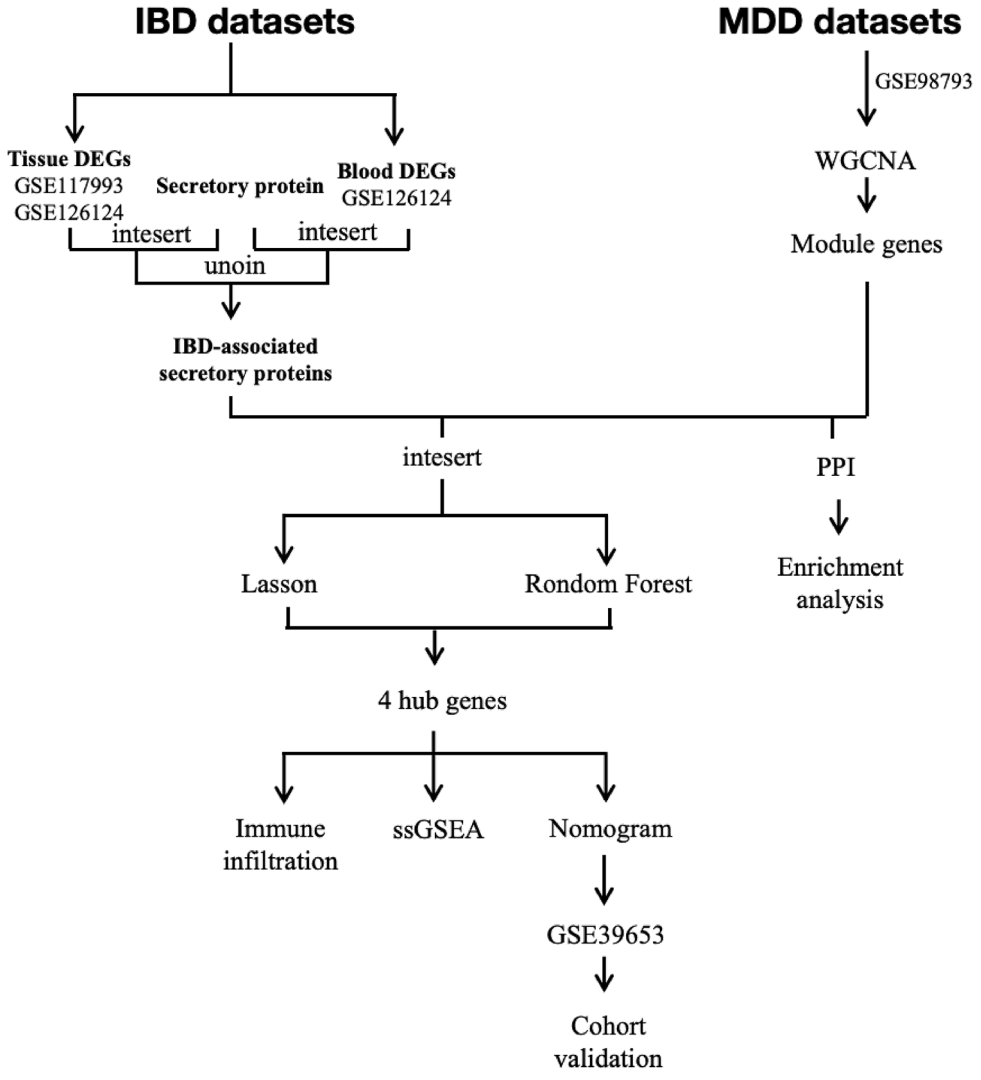

Fig. 1 Flow chart of this study

and heatmaps were used to visualize all the DEGs in the intestinal tissue and peripheral blood mononuclear cells (PBMC) of IBD (Fig. 2A-B).

To further study the secreted proteins in IBD, 3,946 secreted protein-coding genes were downloaded from the Human Protein Atlas database and intersected with the IBD DEG datasets. This resulted in the identification of 463 secreted proteins related to IBD intestinal tissue and 37 secreted proteins related to IBD-PBMC, totaling 484 IBD-related secreted protein genes (Fig. 2C).

Construction of a weighted gene co-expression network and identification of key modules in MDD

The MDD peripheral blood dataset GSE98793 (64 healthy controls + 128 MDD) was used to screen 142 MDD-related DEGs ($p < 0.5$ and $|\log2FC| > 0.585$), including 75 upregulated and 67 downregulated genes. Volcano plots and heatmaps were used to visualize all the DEGs (Supplementary Fig. 1A-B).

To further explore the key genes in MDD, weighted gene co-expression network analysis (WGCNA) was performed to identify the most relevant gene modules in the MDD samples. A soft threshold of $\beta = 10$ was used to ensure the average connectivity and scale-free topology of the co-expression network (Fig. 3A-B). Twelve different colored co-expression gene modules were obtained (Fig. 3C), and correlation analysis showed that the gray module had the highest correlation with MDD ($r = 0.26$, $p = 3e-04$) (Fig. 3D). Therefore, the 430 key genes in the gray module were selected for further analysis (Fig. 3E).

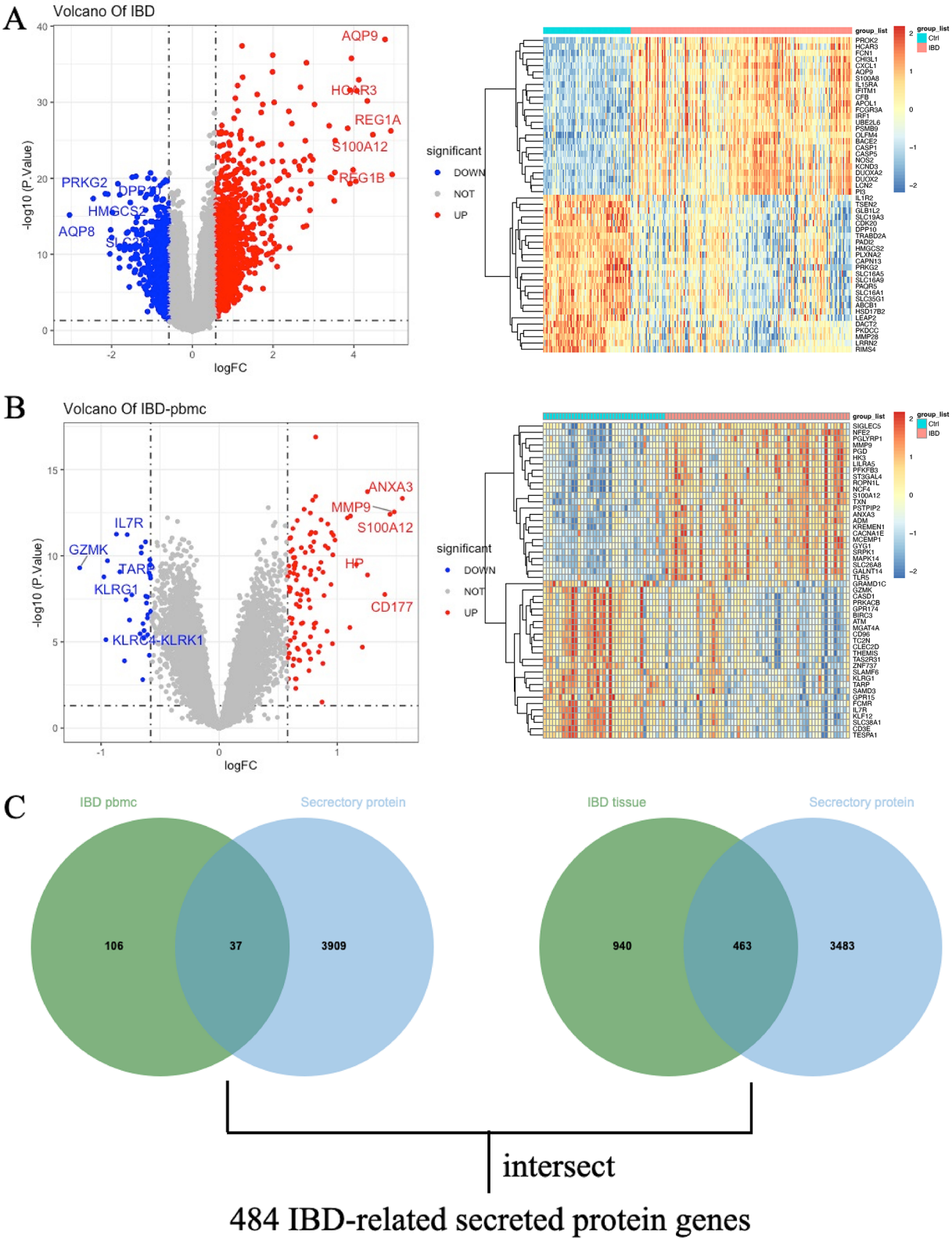

Fig. 2 Integration and differential expression analysis of the IBD dataset. **A**. Volcano plot of DEGs in IBD intestinal tissue samples and heatmap of the top 50 upregulated and downregulated DEGs. **B**. Volcano plot of DEGs in IBD peripheral blood samples and heatmap of the top 50 upregulated and downregulated DEGs. Upregulated genes are represented by red dots, while downregulated genes are represented by blue dots. **C**. Venn diagrams showing the intersection of intestinal tissue and peripheral blood samples with secreted protein genes, resulting in 463 IBD-related secreted protein genes and 37 IBD-PBMC-related secreted protein genes. Further intersection yielded 484 IBD-related secreted protein genes

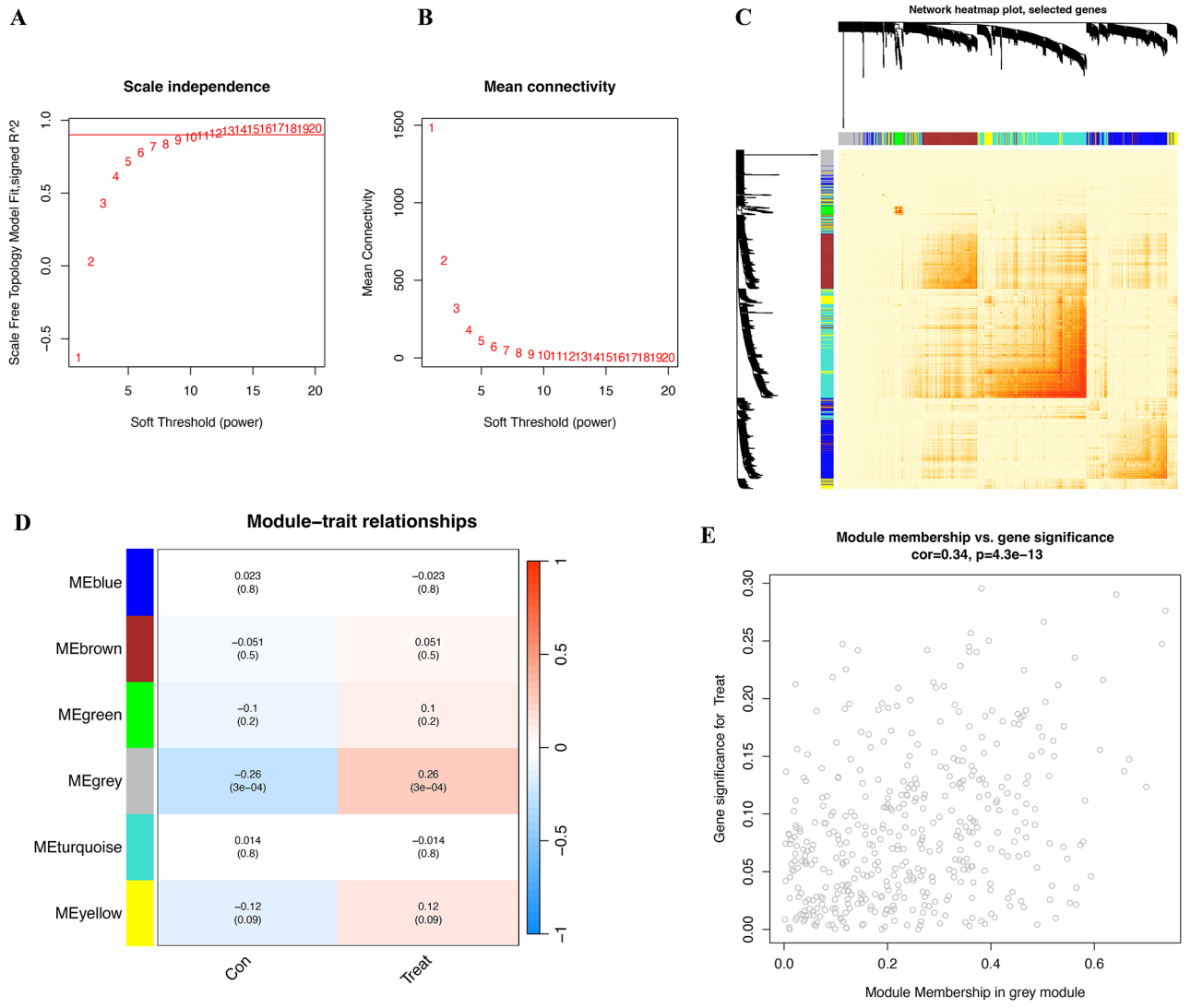

Fig. 3 Identification of key module genes in the MDD dataset using WGCNA. **A**, **B**. Determination of the optimal β value using a scale-free topology model, with β = 10 selected as the soft threshold based on average connectivity and scale independence. **C**. Gene dendrogram and module-feature gene network heatmap. **D**. Heatmap revealing the relationship between module-feature genes and MDD status. The gray module, which exhibited the highest correlation coefficient with MDD, was identified as the key module for MDD. **E**. Correlation plot of gray module members and gene significance in the gray module

Protein-protein interaction network and functional enrichment of IBD-related secreted proteins and MDD pathogenic genes

Clinical studies have found that IBD patients are more likely to develop MDD, suggesting a potential causal relationship between the two. To further explore the pathogenic genes shared by IBD and MDD, the STRING database was used to perform protein-protein interaction analysis between 484 IBD-related secreted proteins and 430 key genes in the MDD pathogenic module. Cytoscape MCODE identified 27 important modules, and the top two most important modules contained 93 genes that were identified as IBD-related MDD pathogenic genes (Fig. 4A).

To better understand the function and mechanisms of these pathogenic genes, the 93 IBD-related MDD pathogenic genes from the top two important modules were imported into the DAVID online database for functional enrichment analysis. Gene Ontology (GO) term analysis showed that the pathogenic genes were mainly located in the extracellular region and were secreted proteins. Biological process (BP) and molecular function (MF) analyses suggested that the pathogenic genes were associated with granulocyte chemotaxis and migration, receptor binding, and receptor activity (Fig. 4B). KEGG pathway analysis indicated that the IBD-related MDD pathogenic genes were closely related to cytokine-cytokine receptor interaction, chemokine signaling pathway, IL-17

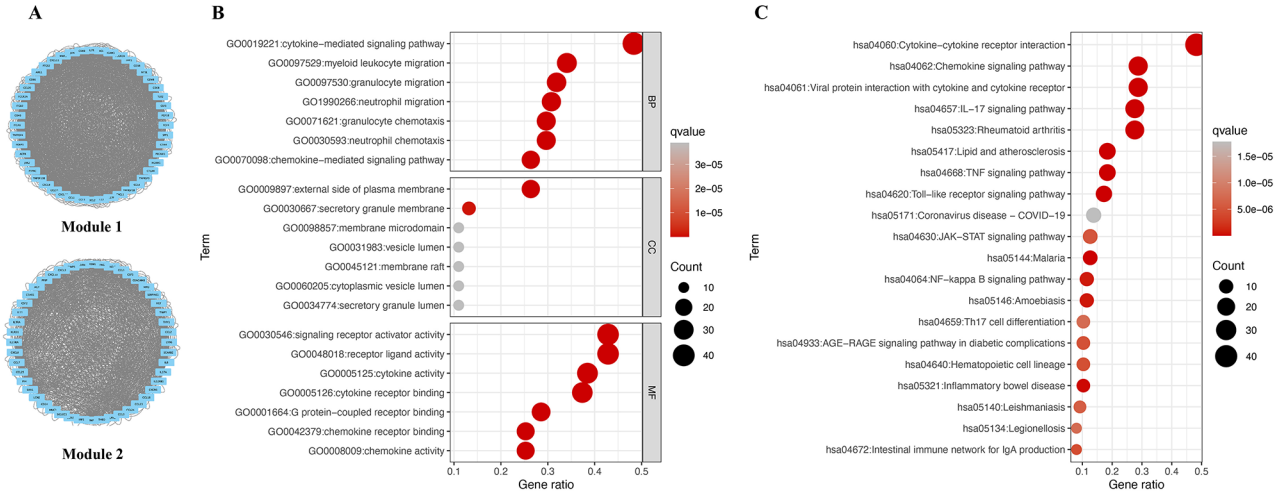

Fig. 4 PPI analysis of IBD-related secreted proteins and key genes in MDD. **A**. PPI network of the top-scoring modules' genes based on MCODE analysis in Cytoscape. **B**, **C**. Bubble plots showing the results of GO and KEGG enrichment analysis

signaling pathway, TNF signaling pathway, Toll-like receptor signaling pathway, JAK-STAT signaling pathway, NF-kappa B signaling pathway, Th17 cell differentiation, and the intestinal immune network for IgA production (Fig. 4C). These results suggest that the secreted proteins in IBD may affect the ligand-receptor binding in MDD, thereby influencing immune regulation and inflammatory response.

Machine learning screening of diagnostic biomarker genes

Considering the potential key role of IBD-related secreted proteins in the development and progression of MDD, the intersection of IBD-related secreted proteins and MDD key disease genes resulted in 21 candidate genes. These 21 candidate genes were used to construct a disease diagnostic model for IBD-related MDD, which could help distinguish IBD patients with or without comorbid MDD (Fig. 5A).

The random forest (RF) machine learning algorithm was applied to rank the 21 candidate genes based on their mean decrease in Gini index, and genes with MeanDecreaseGini > 2 were extracted. Except for CD44, the remaining 20 candidate genes were identified as potential pathogenic genes that significantly impact the diagnosis of IBD-related MDD (Fig. 5B-C). Additionally, the LASSO regression algorithm identified 4 potential pathogenic genes (HGF, SPARC, ADAM12, and MMP8) from the 21 candidate genes (Fig. 5D-E). By taking the intersection of the candidate genes identified by LASSO and RF, 4 common feature genes were finally obtained (Fig. 5F).

In the MDD dataset (GSE98793), compared to the control group, MDD patients showed significantly higher expression levels of HGF and MMP8 ($P<0.05$), while SPARC and ADAM12 had an increasing trend but without statistical significance (Fig. 5G). Furthermore, in the IBD dataset, compared to the control group, IBD patients exhibited significantly higher expression levels of HGF, MMP8, SPARC, and ADAM12 ($P<0.05$) (Fig. 5H). Therefore, it can be concluded that MDD and IBD patients both have upregulated expression of HGF and MMP8.

The prediction model of IBD-related MDD

To better diagnose and predict IBD-related MDD, we performed logistic regression analysis on 4 feature genes, and then built a nomogram based on this (Fig. 6A). By summing the scores corresponding to each feature gene, we obtained a total score that corresponds to the disease risk of IBD-related MDD. The calibration curve suggests that the predictive probability of the nomogram model is nearly the same as that of the ideal model (Fig. 6B). The decision curve analysis (DCA) plot indicates that the nomogram model is beneficial for the diagnosis of IBD-related MDD (Fig. 6C). ROC curve analysis was used to evaluate the area under the curve (AUC) of the nomogram model to determine its sensitivity and specificity in diagnosing IBD-related MDD (Fig. 6D), and the AUC value was 0.724 in the internal dataset GSE98793. Furthermore, in the external dataset GSE39653, the nomogram showed an AUC value of 0.714 for MDD patients (Fig. 6E), suggesting that the nomogram model has good diagnostic performance for IBD-related MDD.

Single-gene enrichment analysis and immune infiltration analysis

To elucidate the role of the 4 feature genes in the MDD dataset GSE98793, we performed GSEA analysis, which showed that the KEGG enrichment was mainly in neutrophil extracellular trap formation, NOD-like receptor

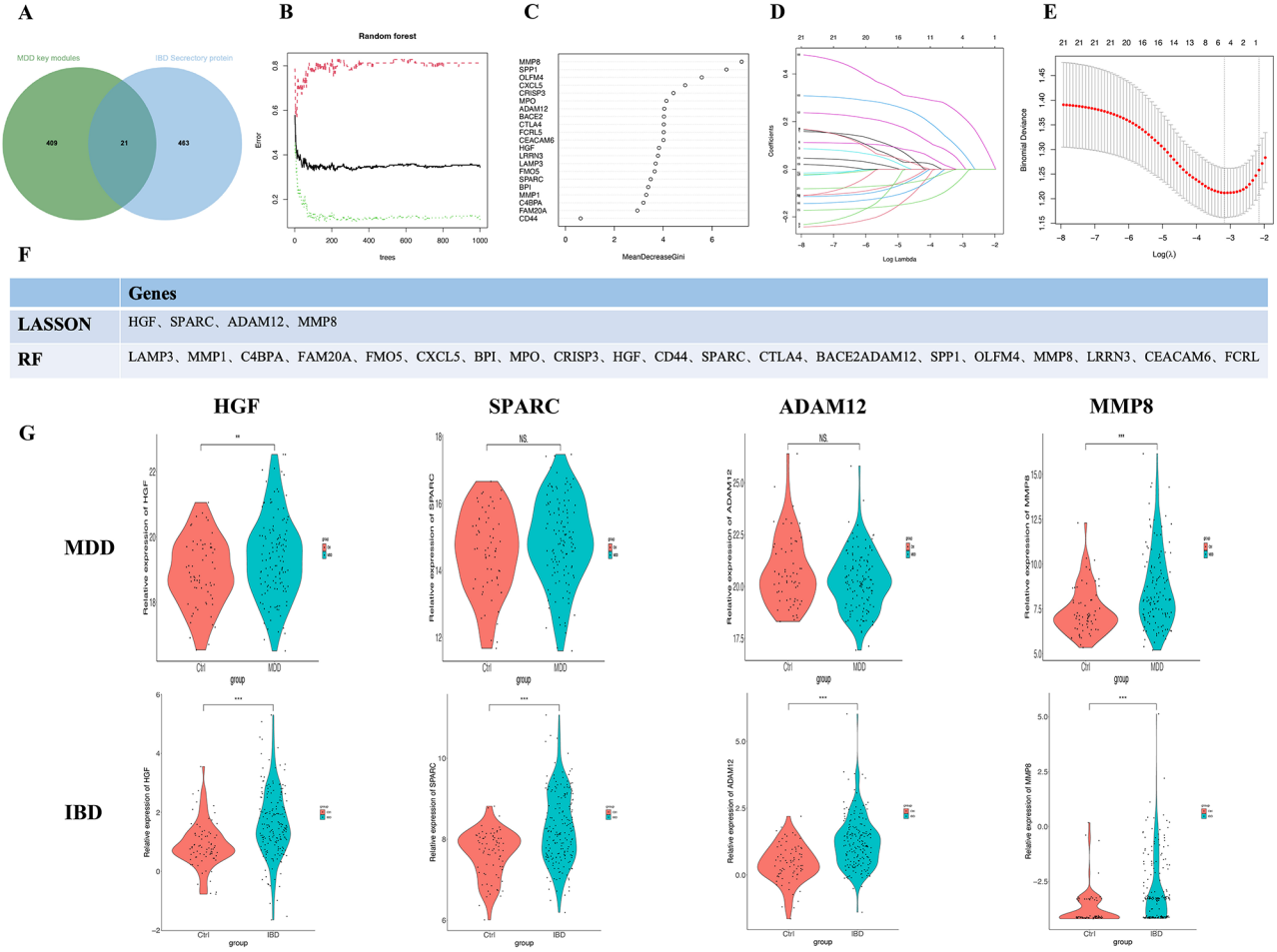

Fig. 5 Identification of potential diagnostic biomarkers for IBD-related MDD using machine learning approaches. **A**. Venn diagram showing the overlap of 21 common genes between IBD-related secreted proteins and key genes in MDD. **B**, **C**. MeanDecreaseGini plots of the 21 genes in MDD using the Random Forest (RF) algorithm. **D**, **E**. Identification of the minimum value and lambda value for diagnostic biomarker selection using the LASSO logistic regression algorithm. **F**. Intersection of candidate genes selected by LASSO and RF algorithms, resulting in four potential pathogenic feature genes: HGF, SPARC, ADAM12, and MMP8. **G**, **H**. Expression levels of the four feature genes in the MDD dataset (GSE98793) and the IBD dataset

signaling pathway, Toll-like receptor signaling pathway, Th1, Th2 and Th17 cell differentiation, and chemokine signaling pathway, suggesting that the 4 feature genes are involved in the immune regulation of MDD (Fig. 7A).

To further study the relationship between immune regulation and diagnostic markers in MDD, we used the CIBERSORT algorithm to infer the characteristics of immune cells. Analysis of 22 infiltrating immune cells in the MDD dataset GSE98793 found that there were only significant differences in 2 infiltrating immune cells, with MDD showing a higher proportion of monocytes and a lower proportion of helper T cells compared to the control group (Fig. 7B). Furthermore, we further explored the relationship between the expression of the 4 feature genes and the proportion of different infiltrating immune cell types. MMP8 was positively correlated with monocytes, neutrophils, resting NK cells, and $CD4^+$ memory T cells, but negatively correlated with activated NK cells, CD4 naive T cells, and regulatory T cells; HGF was positively correlated with M2 macrophages, but negatively correlated with CD8+T cells; ADAM12 was negatively correlated with mast cells (Fig. 7C).

External validation of the four diagnostic markers

To further confirm the accuracy of the above integrated bioinformatics analysis, we recruited IBD with and without comorbid MDD patients from our hospital as an external validation cohort, and detected the levels of the 4 candidate diagnostic markers in the patient's serum by ELISA. The results showed that the serum levels of HGF and MMP8 were significantly elevated in IBD patients with comorbid MDD, SPARC and ADAM12 showed an increasing trend (Fig. 8A). Meanwhile, we developed a predictive model for IBD-related MDD based on our cohort (Fig. 8B). ROC curve analysis of the efficacy of each feature gene and the predictive model showed that

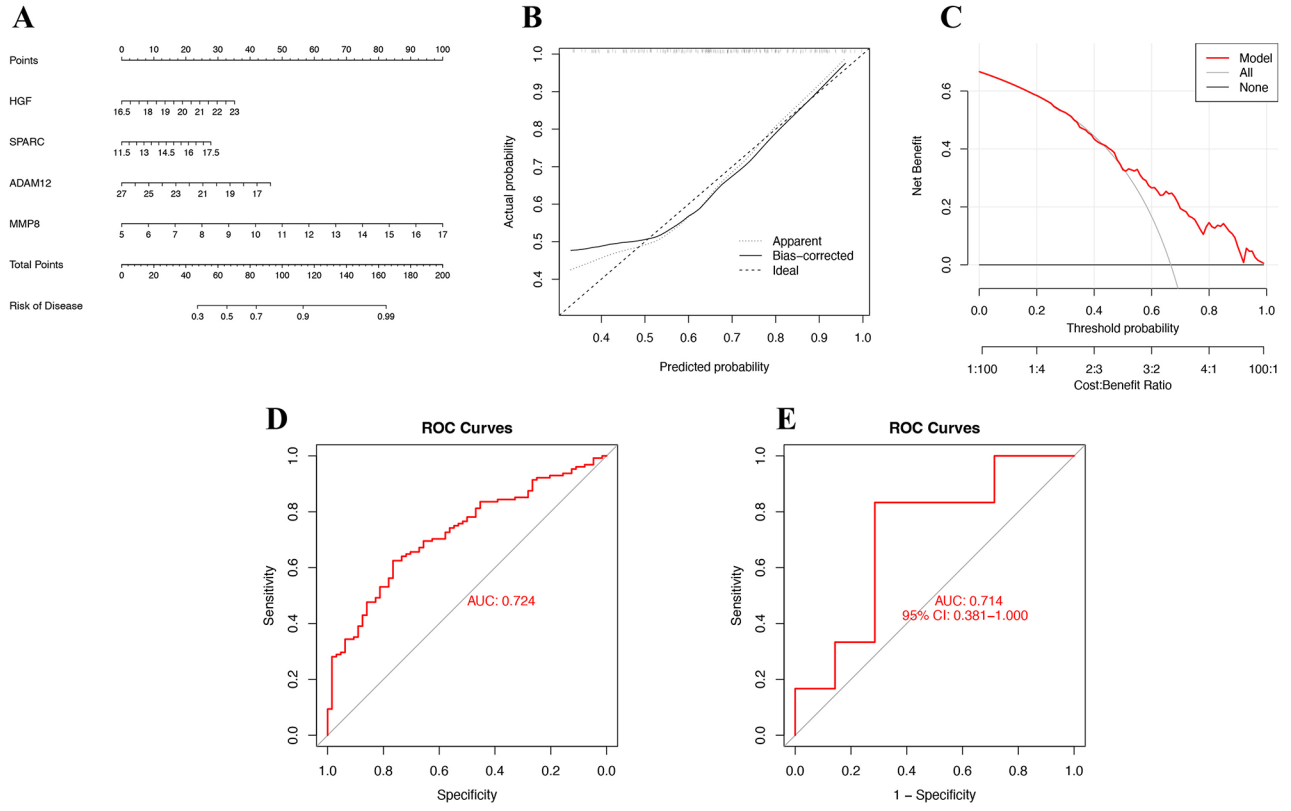

Fig. 6 Development and evaluation of the diagnostic nomogram model. **A**. nomogram constructed based on diagnostic biomarkers. **B**. Calibration curve of the nomogram model predicting MDD in IBD-related MDD. The dashed line labeled "Ideal" represents the standard curve, indicating perfect prediction by an ideal model. The dashed line labeled "Apparent" represents the uncalibrated prediction curve, while the solid line labeled "Bias-Corrected" represents the calibrated prediction curve. **C**. Decision curve analysis (DCA) of the nomogram model. The black line labeled "None" represents the net benefit assuming no patients have MDD. The gray line labeled "All" represents the net benefit assuming all patients have MDD, while the red line labeled "Model" represents the net benefit assuming the identification of IBD-related MDD based on the nomogram model. **D**. Receiver Operating Characteristic (ROC) curve of the diagnostic performance of the nomogram model for predicting MDD in the internal dataset from the GEO database. **E**. ROC curve of the diagnostic performance of the nomogram model for predicting MDD in the external dataset from the GEO database

the AUC values of HGF and MMP8 were >0.8, suggesting they have strong clinical diagnostic value (Fig. 8C). Furthermore, the AUC value of the joint predictive model based on the 4 diagnostic markers (AUC=0.87, 95%CI: 0.738–0.975) was higher than that of each feature gene, indicating that the nomogram may have strong diagnostic value for IBD-related MDD (Fig. 8D).

Discussion

Depression and anxiety are common comorbidities in IBD; however, there is currently a lack of large-scale, high-quality, prospective population cohort studies to elucidate the true prevalence, diagnosis, temporal relationship, and specific mechanisms of IBD-related MDD. Exploring objective predictive biomarkers from a biological perspective could provide opportunities for earlier and more effective interventions or prevention for these patients [23]. For IBD patients with comorbid MDD, healthcare providers need to consider comorbidity as part of the background of IBD and recognize the importance of addressing mental health in the clinical outcomes of IBD. A prospective study published in 2021 found that, compared to the previous 12 months, the number of visits to the emergency department for any reason decreased within 12 months after intervention for IBD patients at risk of mental health disorders who received psychological interventions [24]. In addition to benefiting mental illness and quality of life, psychological interventions also impact the course of IBD itself. Therefore, examining quantitative biomarkers that are bidirectionally associated with IBD and MDD can provide a more objective assessment of patients' disease status. Integrative bioinformatics analysis has become increasingly common in exploring new genes, potential diagnostic and prognostic biomarkers, underlying mechanisms, and therapeutic targets, providing valuable insights into various diseases [21]. Through various integrative bioinformatics methods, this study has for the first time identified shared pathogenic genes encoding secreted proteins between IBD and MDD. Protein-protein

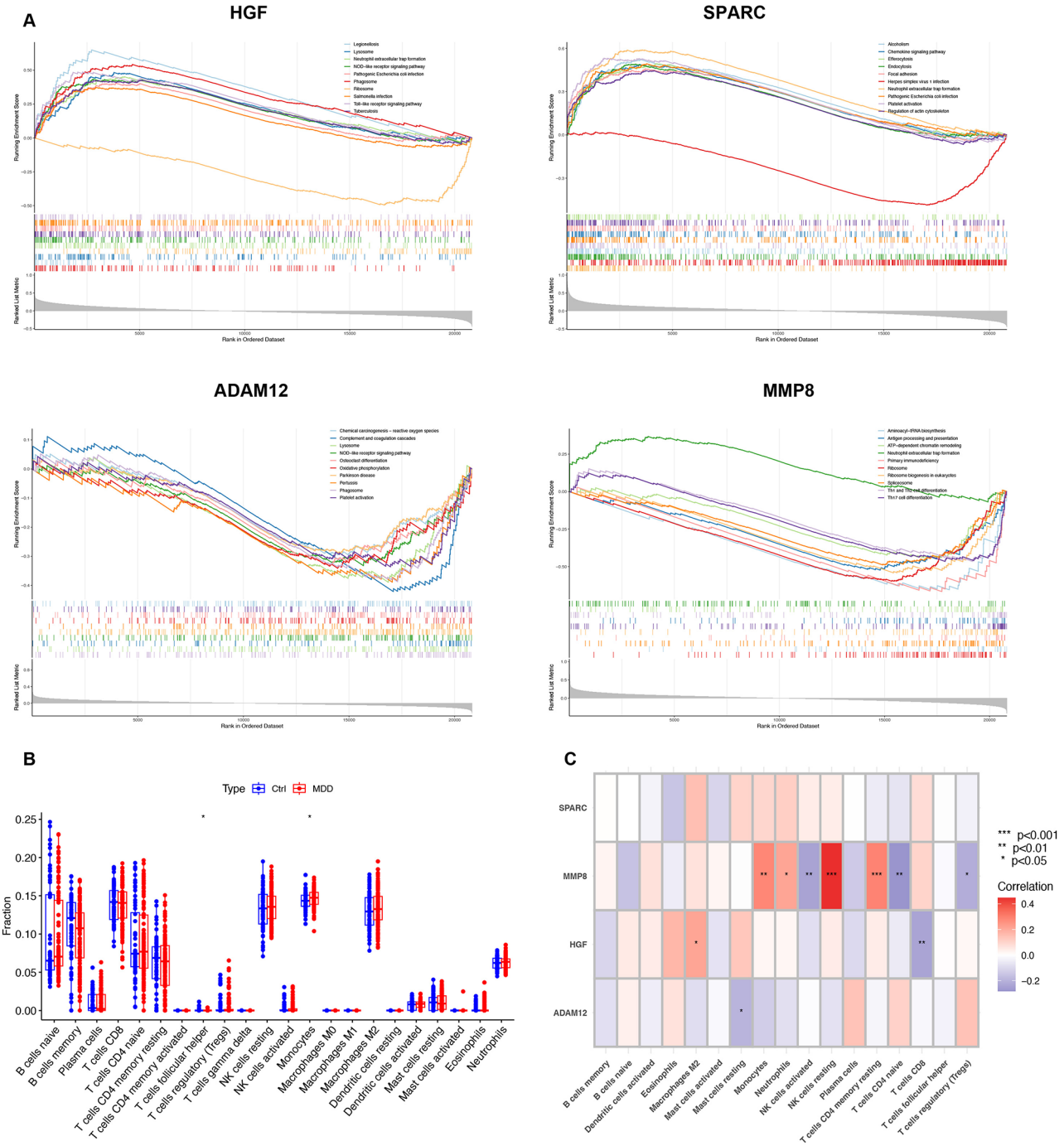

Fig. 7 Single-gene enrichment analysis and immune infiltration analysis of candidate biomarkers. **A**. GSEA analysis and KEGG enrichment of the four feature genes. **B**. Violin plots comparing the 22 immune cell types between the MDD group and the control group. **C**. Heatmap revealing the correlation of immune cell infiltration above the $p < 0.05$ threshold

interaction analysis and pathway enrichment analysis revealed potential pathogenic mechanisms of IBD-related MDD, including inflammatory and immune processes, as well as signaling pathways such as the IL-17 signaling pathway, TNF signaling pathway, JAK-STAT signaling pathway, NF-kappa B signaling pathway, and intestinal immune network for IgA production, including Th17 cell differentiation. Depression is an inflammatory disorder characterized by cell-mediated immune activation. The NF-κB signaling pathway is considered crucial for the development of major depressive disorder, regulating numerous genes associated with inflammation and neuroplasticity, thereby affecting neurogenesis [25]. Th17 cells are involved in the gut-brain axis to mediate

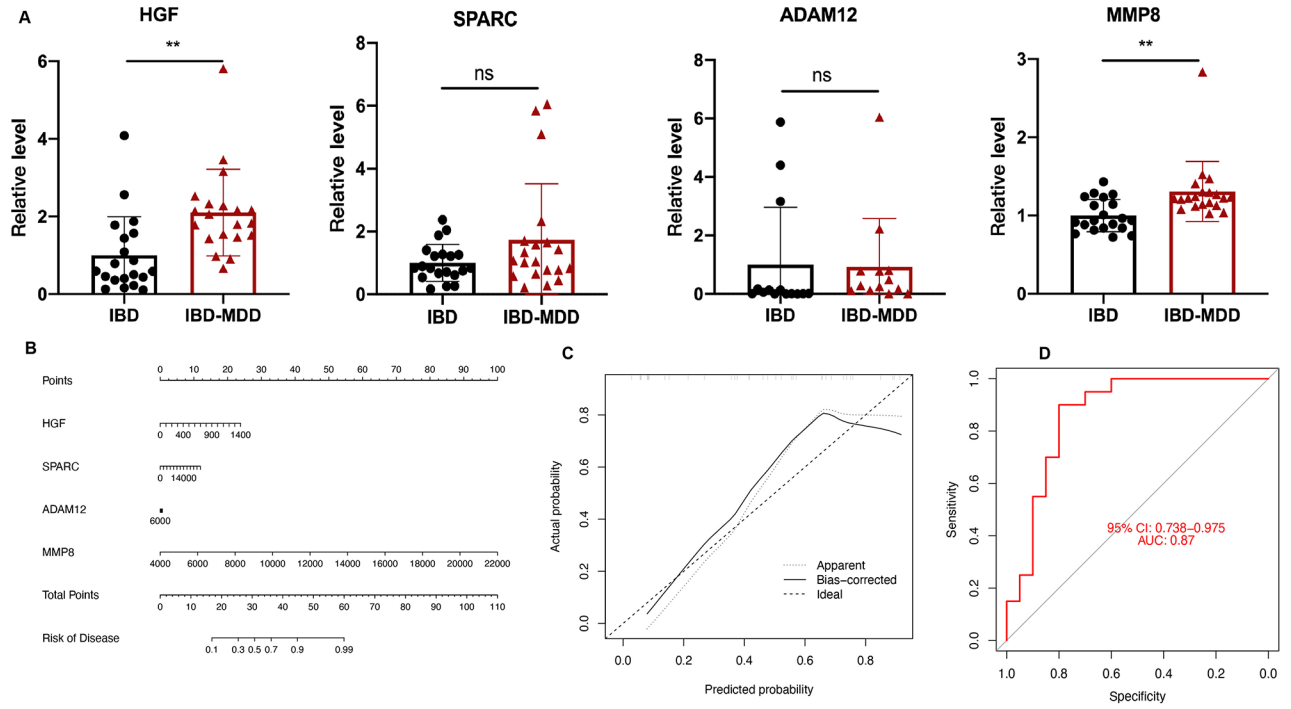

Fig. 8 Validation of the expression patterns of the four candidate diagnostic biomarkers in IBD and IBD-MDD serum samples, and evaluation of the diagnostic performance of the nomogram model for differentiating MDD. **A**. ELISA results showing significant elevation of HGF and MMP8 levels in the serum of IBD-MDD patients, with an increasing trend in SPARC and ADAM12. **B**. Nomogram model developed based on the four diagnostic biomarkers for predicting the risk of MDD. **C**. Calibration curve of the nomogram model predicting MDD in the IBD patients. D. ROC curve of the predictive performance of the four candidate biomarkers and the nomogram model. ** P value<0.01, ns: no significance

stress responses. Th17 cells increased microglial activation in the hippocampus, amygdala, and prefrontal cortex due to mild prenatal stress, and anti-IL-17 treatment rescued depressive and anxious behaviors following perinatal stress [26].Furthermore, using machine learning, we ultimately identified four diagnostic genes, HGF, MMP8, SPARC, and ADAM12. We constructed a nomogram prediction model using the training set (GSE98793) and assessed the accuracy of the prediction model using the ROC curve with an external validation dataset (GSE39653). Finally, the diagnostic value of the diagnostic genes and model was confirmed in our external cohort.

The brain-gut axis is a bidirectional communication system between the brain, intestines, autonomic nervous system, and hypothalamic-pituitary-adrenal (HPA) axis [27]. The vagus nerve, immune and neuroendocrine systems, neurotransmitters and metabolites, and gut microbiota are key nodes that play their respective roles in the brain-gut axis pathway [28]. In our study, KEGG pathway analysis showed that the pathogenic genes of IBD-related MDD are mainly associated with inflammation and immune pathways. Similar findings have been reported in other literature, where some patients with stress-related neurobehavioral disorders such as MDD exhibit a chronic low-grade inflammatory state characterized by increased circulating pro-inflammatory cytokines and leukocytes [29]. Therefore, the occurrence of IBD-related MDD is closely associated with immune dysregulation in the context of intestinal inflammation.

HGF is a heterodimer secreted by mesenchymal cells and is found in the lungs, liver, intestines, and central nervous system. Literature reports indicated that HGF can inhibit inflammatory pathways and improve intestinal symptoms in an IBD animal model [30]. In IBD patients, serum levels of HGF are significantly higher than in the control group and are correlated with IBD activity [31, 32]. HGF is also a key factor in protecting neurons, preventing neuronal death, and promoting neural regeneration. Studies have found that HGF is highly expressed at sites of nerve injury and is used for neural repair [33]. Furthermore, previous studies have found elevated levels of HGF in postpartum depression women and MDD patients [34]. Recent literature also reports a positive correlation between HGF levels in the cerebrospinal fluid of MDD patients and Hamilton Depression Rating Scale scores [35].

MMP8, also known as neutrophil collagenase or collagenase-2, is a predominantly inactive enzyme secreted by neutrophils and plays an important role in mediating inflammation under the influence of biologically active mediators such as interleukins and complement

system components [36]. A recent study reported that MMP8 is increased in the serum of both humans with MDD and stressed mice, and this increase is associated with neurophysiological changes, behavioral alterations, and structural changes in the extracellular space [37]. Additionally, stress promotes a new model of immune cell-brain interaction to regulate social behavior, where MMP8 produced by peripheral immune cells affects neuronal function by altering the extracellular space. Moreover, elevated serum levels of MMP8 are associated with malignant tumors, decreased survival rates, and increased systemic inflammation in conditions such as colorectal cancer, hepatocellular carcinoma, pancreatic cancer, and septic shock [38, 39].

ADAM12, also known as protein 12 with integrin and metalloproteinase domains, is widely expressed in various tissues and is involved in multiple cellular processes, including cell adhesion, cell fusion, protein hydrolysis, and signal transduction [40]. The expression level of ADAM12 is associated with the progression of various diseases, such as cancer, liver fibrosis, and cardiac hypertrophy, and has diagnostic and prognostic value [41]. ADAM12 is the most extensively studied biomarker for diagnosis and prognosis in breast cancer. Studies have also reported that the expression level of ADAM12 is associated with the diagnosis and differentiation of brain tumors from different cell sources [42]. Furthermore, recent research has indicated that ADAM12 can serve as a prognostic marker for rectal cancer patients after radiation therapy [43]. Our study found a trend of increased ADAM12 expression levels in the serum of IBD-MDD patients compared to the IBD group, but without significant statistical differences. It can serve as a combined predictive indicator for the comorbidity of IBD-related MDD.

Secreted protein acidic and rich in cysteine (SPARC), also known as basement membrane protein 40 (BM-40) or osteonectin, is a widely expressed matricellular glycoprotein [44]. It is involved in various biological processes, including regulation of extracellular matrix assembly and deposition, anti-adhesion, modulation of extracellular proteinase activity, and regulation of growth factor/cytokine signaling pathways [45]. Recent literature has also reported the predictive value of the basement membrane gene SPARC in inflammatory bowel disease [46].

Based on the immune infiltration results of four hub genes, it can be inferred that intestinal diseases may involve the migration of immune cells such as CD4+T cells, CD8+T cells, regulatory T cells, NK cells, and mast cells to the central nervous system, leading to neuroimmune activation. The occurrence of IBD-related MDD is closely related to the interaction of the inflammatory immune pathway. Previous studies have also found that intestinal immune cells can directly regulate neuroimmune homeostasis and the brain's response to inflammation. Autoimmune encephalomyelitis (EAE) induces intestinal plasma cells to produce immunoglobulin A (IgA), which migrates extensively to the central nervous system [47]. Antibiotic-induced changes in the gut microbiota can regulate IL-17+γδT cells and regulatory T cells through dendritic cells, limiting neuroinflammatory responses and improving brain injuries [48].

This study has certain limitations. (1) The data used in our study came from the GEO database, and the sample size obtained from the database is limited, which may lead to false-positive results. Therefore, further research with larger sample sizes is necessary to elucidate the exact pathogenesis of IBD-related MDD. (2) The difficulty in obtaining detailed information related to depressive symptoms from the dataset samples, such as the first event, exposure, and specific scores from depression symptom scales. This limitation may be one reason why the AUC value of our diagnostic nomogram's ROC curve is lower in the dataset compared to the results from our own cohort validation. (3) The substantial difference in gender ratios between the two groups, which may explain the lack of significant differences in serum biomarkers between IBD and IBD-MDD.

Conclusion

In conclusion, this study identified four candidate biomarkers for IBD-related MDD, including HGF, SPARC, ADAM12, and MMP8. These candidate biomarkers not only distinguish MDD but may also contribute to the occurrence of MDD through interactions with the inflammatory immune pathway. Our research reveals the shared pathogenesis of IBD and MDD, providing new insights into the diagnosis and therapeutic interventions of serum-based IBD-related MDD.

Abbreviations

IBD	Inflammatory bowel disease
IBD-MDD	IBD comorbid with MDD
PPI	protein-protein interaction
MDD	major depressive disorder
HAMD	Hamilton Depression Rating Scale
BDI	Inventory
RF	random forest
NB	naive Bayes
LR	logistic regression
GMM	Gaussian mixed model
SVM	support vector machine
DT	decision tree
WGCNA	weighted gene co-expression network analysis
NEUT	neutrophil absolute count
HGB	hemoglobin
ESR	erythrocyte sedimentation rate
CRP	reactive protein
WBC	white blood cell count
EO	eosinophil count
PLT	platelets
ALB	albumin
25-OH-VD	hydroxyvitamin D
FCP	Fecal calprotectin

DEGs	Differentially Expressed Genes
GO	Gene Ontology
BP	Biological process
MF	molecular function
DCA	decision curve analysis
AUC	area under the curve

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-024-05758-8.

Supplementary Material 1

Acknowledgements
We thank the patients who participated in this study.

Author contributions
CH, MG, XS, and HG were involved in the conception and design of the study; YL, WT, YZ, and LX were involved in the acquisition of the data; MZ, and XS analysed and interpreted the data; CH and MG drafted the article; HG critically revised for intellectual content; and all authors approved the final version for submission.

Funding
This study was supported by National Key R&D Program of China (2023YFC2507300), Science and Health Joint Medical Research Program of Chongqing Municipality (2024QNXM036, 2024ZDXM009), and National Natural Science Foundation of China (#82400649).

Data availability
The data underlying this article will be shared on reasonable request to the corresponding author.

Declarations

Ethics approval and consent to participate
This study was approved by the Clinical Research Ethics Committee of Chongqing General Hospital (Ethics Approval Number: KY S2023-084-01), and all methods were performed in accordance with the relevant guidelines and regulations, with all participants having signed informed consent.

Consent for publication
Not applicable.

Competing interests
The authors declared no competing interests.

Received: 14 July 2024 / Accepted: 8 October 2024
Published online: 05 November 2024

References

1. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature. 2011;474:307–17.
2. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet. 2017;390:2769–78.
3. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. Lancet. 2017;389:1756–70.
4. Torres J, Mehandru S, Colombel JF, Peyrin-Biroulet L. Crohn's disease. Lancet. 2017;389:1741–55.
5. Neuendorf R, Harding A, Stello N, Hanes D, Wahbeh H. Depression and anxiety in patients with inflammatory bowel disease: a systematic review. J Psychosom Res. 2016;87:70–80.
6. Macer BJ, Prady SL, Mikocka-Walus A. Antidepressants in Inflammatory Bowel Disease: a systematic review. Inflamm Bowel Dis. 2017;23:534–50.
7. Fairbrass KM, Lovatt J, Barberio B, Yuan Y, Gracie DJ, Ford AC. Bidirectional brain-gut axis effects influence mood and prognosis in IBD: a systematic review and meta-analysis. Gut. 2022;71:1773–80.
8. Alexakis C, Kumar S, Saxena S, Pollok R. Systematic review with meta-analysis: the impact of a depressive state on disease course in adult inflammatory bowel disease. Aliment Pharmacol Ther. 2017;46:225–35.
9. Gong W, Guo P, Li Y, Liu L, Yan R, Liu S, et al. Role of the gut-brain Axis in the Shared Genetic Etiology between Gastrointestinal Tract diseases and Psychiatric disorders: a genome-wide pleiotropic analysis. JAMA Psychiatry. 2023;80:360–70.
10. Uher R, Payne JL, Pavlova B, Perlis RH. Major depressive disorder in DSM-5: implications for clinical practice and research of changes from DSM-IV. Depress Anxiety. 2014;31:459–71.
11. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol. 1967;6:278–96.
12. Zhang E, Huang Z, Zang Z, Qiao X, Yan J, Shao X. Identifying circulating biomarkers for major depressive disorder. Front Psychiatry. 2023;14:1230246.
13. Squarcina L, Villa FM, Nobile M, Grisan E, Brambilla P. Deep learning for the prediction of treatment response in depression. J Affect Disord. 2021;281:618–22.
14. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28:882–3.
15. 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.
16. Xu M, Zhou H, Hu P, Pan Y, Wang S, Liu L, et al. Identification and validation of immune and oxidative stress-related diagnostic markers for diabetic nephropathy by WGCNA and machine learning. Front Immunol. 2023;14:1084531.
17. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023;51:D638–46.
18. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51:D587–92.
19. Engebretsen S, Bohlin J. Statistical predictions with glmnet. Clin Epigenetics. 2019;11:123.
20. 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.
21. Zhu E, Shu X, Xu Z, Peng Y, Xiang Y, Liu Y, et al. Screening of immune-related secretory proteins linking chronic kidney disease with calcific aortic valve disease based on comprehensive bioinformatics analysis and machine learning. J Transl Med. 2023;21:359.
22. Xu J, Zhou H, Cheng Y, Xiang G. Identifying potential signatures for atherosclerosis in the context of predictive, preventive, and personalized medicine using integrative bioinformatics approaches and machine-learning strategies. EPMA J. 2022;13:433–49.
23. Bisgaard TH, Allin KH, Keefer L, Ananthakrishnan AN, Jess T. Depression and anxiety in inflammatory bowel disease: epidemiology, mechanisms and treatment. Nat Rev Gastroenterol Hepatol. 2022;19:717–26.
24. Lores T, Goess C, Mikocka-Walus A, Collins KL, Burke ALJ, Chur-Hansen A, et al. Integrated Psychological Care Reduces Health Care Costs at a hospital-based inflammatory bowel Disease Service. Clin Gastroenterol Hepatol. 2021;19:96–e103103.
25. Chao B, Zhang L, Pan J, Zhang Y, Chen Y, Xu M, et al. Stanniocalcin-1 overexpression prevents Depression-Like behaviors through inhibition of the ROS/NF-κB signaling pathway. Front Psychiatry. 2021;12:644383.
26. Beurel E, Medina-Rodriguez EM, Jope RS. Targeting the adaptive Immune System in Depression: Focus on T Helper 17 cells. Pharmacol Rev. 2022;74(2):373–86.
27. Cryan JF, O'Riordan KJ, Cowan CSM, Sandhu KV, Bastiaanssen TFS, Boehme M, et al. The Microbiota-Gut-Brain Axis. Physiol Rev. 2019;99:1877–2013.
28. Morais LH, Schreiber HLt, Mazmanian SK. The gut microbiota-brain axis in behaviour and brain disorders. Nat Rev Microbiol. 2021;19:241–55.
29. Dowlati Y, Herrmann N, Swardfager W, Liu H, Sham L, Reim EK, et al. A meta-analysis of cytokines in major depression. Biol Psychiatry. 2010;67:446–57.
30. Oh K, Iimuro Y, Takeuchi M, Kaneda Y, Iwasaki T, Terada N, et al. Ameliorating effect of hepatocyte growth factor on inflammatory bowel disease in a murine model. Am J Physiol Gastrointest Liver Physiol. 2005;288:G729–735.

31. Naguib R, El-Shikh WM. Clinical Significance of Hepatocyte Growth Factor and Transforming Growth Factor-Beta-1 Levels in Assessing Disease Activity in Inflammatory Bowel Disease. Can J Gastroenterol Hepatol. 2020; 2020:2104314.
32. Sturm A, Schulte C, Schatton R, Becker A, Cario E, Goebell H, et al. Transforming growth factor-beta and hepatocyte growth factor plasma levels in patients with inflammatory bowel disease. Eur J Gastroenterol Hepatol. 2000;12:445–50.
33. Ko KR, Lee J, Lee D, Nho B, Kim S. Hepatocyte growth factor (HGF) promotes peripheral nerve regeneration by activating repair Schwann cells. Sci Rep. 2018;8:8316.
34. Brann E, Fransson E, White RA, Papadopoulos FC, Edvinsson A, Kamali-Moghaddam M, et al. Inflammatory markers in women with postpartum depressive symptoms. J Neurosci Res. 2020;98:1309–21.
35. Hidese S, Hattori K, Sasayama D, Tsumagari T, Miyakawa T, Matsumura R, et al. Cerebrospinal fluid neuroplasticity-associated protein levels in patients with psychiatric disorders: a multiplex immunoassay study. Transl Psychiatry. 2020;10:161.
36. Zalewska EA, Lawicka R, Grygorczuk P, Nowosielska M, Kicman A, Lawicki S. Importance of metalloproteinase 8 (MMP-8) in the diagnosis of Periodontitis. Int J Mol Sci. 2024; 25.
37. Cathomas F, Lin HY, Chan KL, Li L, Parise LF, Alvarez J, et al. Circulating myeloid-derived MMP8 in stress susceptibility and depression. Nature. 2024;626:1108–15.
38. Juurikka K, Butler GS, Salo T, Nyberg P, Astrom P. The role of MMP8 in Cancer: a systematic review. Int J Mol Sci 2019; 20.
39. Kong C, Zhu Y, Xie X, Wu J, Qian M. Six potential biomarkers in septic shock: a deep bioinformatics and prospective observational study. Front Immunol. 2023;14:1184700.
40. Giebeler N, Zigrino P. A disintegrin and metalloprotease (ADAM): historical overview of their functions. Toxins (Basel). 2016;8:122.
41. Lukaszewicz-Zajac M, Dulewicz M, Mroczko B. A disintegrin and metalloproteinase (ADAM) family: their significance in malignant tumors of the Central Nervous System (CNS). Int J Mol Sci. 2021; 22.
42. Kanakis D, Lendeckel U, Theodosiou P, Dobrowolny H, Mawrin C, Keilhoff G, et al. ADAM 12: a putative marker of oligodendrogliomas? Dis Markers. 2013;34:81–91.
43. Piotrowski KB, Blasco LP, Samsoe-Petersen J, Eefsen RL, Illemann M, Oria VO, et al. ADAM12 expression is upregulated in cancer cells upon radiation and constitutes a prognostic factor in rectal cancer patients following radiotherapy. Cancer Gene Ther. 2023;30:1369–81.
44. Nagaraju GP, El-Rayes BF. SPARC and DNA methylation: possible diagnostic and therapeutic implications in gastrointestinal cancers. Cancer Lett. 2013;328:10–7.
45. Pascual-Pasto G, Castillo-Ecija H, Unceta N, Aschero R, Resa-Pares C, Gomez-Caballero A, et al. SPARC-mediated long-term retention of nab-paclitaxel in pediatric sarcomas. J Control Release. 2022;342:81–92.
46. Lin P, Hua J, Teng Z, Lin C, Liu S, He R, et al. Screening of hub inflammatory bowel disease biomarkers and identification of immune-related functions based on basement membrane genes. Eur J Med Res. 2023;28:247.
47. Rojas OL, Probstel AK, Porfilio EA, Wang AA, Charabati M, Sun T, et al. Recirculating intestinal IgA-Producing cells regulate Neuroinflammation via IL-10. Cell. 2019;176:610–e624618.
48. Benakis C, Brea D, Caballero S, Faraco G, Moore J, Murphy M, et al. Commensal microbiota affects ischemic stroke outcome by regulating intestinal gammadelta T cells. Nat Med. 2016;22:516–23.

Publisher's note