

Identification of Key Biomarkers and Immune Microenvironment Features in Ulcerative Colitis: An Integrated Analysis Using WGCNA and Multiple Machine Learning Algorithms

Yingchao Qi^{1,*}, Yichen Wang^{2,*}, Siyao Zhang^{1,3,*}, Xinxin Pan⁴, Wenkai Li², Meijia Cheng², Wenjing Ma², Jiajia Li², Yue Pei², Yunen Liu^{1,2}, Yongduo Yu⁵

¹First Clinical School, Liaoning University of Traditional Chinese Medicine, Shenyang, People's Republic of China; ²School of Shuren International, Shenyang Medical College, Shenyang, People's Republic of China; ³Department of Anorectal Surgery, The Third Affiliated Hospital of Liaoning University of Traditional Chinese Medicine, Shenyang, People's Republic of China; ⁴Department of Anorectal Surgery, Wuxi Huishan District Hospital of Traditional Chinese Medicine, Wuxi, People's Republic of China; ⁵Department of Anorectal Surgery, The Second Affiliated Hospital of Liaoning University of Traditional Chinese Medicine, Shenyang, People's Republic of China

*These authors contributed equally to this work

Correspondence: Yongduo Yu, Department of Anorectal Surgery, The Second Affiliated Hospital of Liaoning University of Traditional Chinese Medicine, Shenyang, People's Republic of China, Email yuyongduo@163.com

Background: Ulcerative colitis (UC) is an inflammatory bowel disease (IBD) characterized by a dysregulated mucosal immune response in the intestine. The disease poses significant challenges for both diagnosis and treatment. This study aims to identify reliable biomarkers for UC and investigate its immunological characteristics, with the goal of improving diagnostic accuracy and informing treatment strategies.

Methods: This study integrated multiple GEO datasets and identified UC-associated hub genes through differential expression and Weighted Gene Co-expression Network Analysis (WGCNA), with subsequent refinement using least absolute shrinkage and selection operator (LASSO), randomForest (RF), and support vector machine-recursive feature elimination (SVM-RFE) algorithms. These genes were used to construct a feedforward neural network (FNN) diagnostic model, whose performance was assessed using receiver operating characteristic (ROC) curve analysis. Immune profiling based on CIBERSORT, ssGSEA, Gene set variation analysis (GSVA), and ESTIMATE revealed associations between hub gene expression, immune cell infiltration, and inflammatory activity. Immunohistochemistry was performed to validate the protein expression of hub genes.

Results: Ten hub genes (ACOX2, MMP3, CPT2, CTSK, CHP2, VCAM1, SLC25A34, BASP1, NCF2, and GLB1L2) were identified, and the FNN model showed strong diagnostic accuracy. Notably, NCF2 expression correlated with immune cell infiltration and immune/inflammation scores, and was confirmed to be elevated in UC tissues, suggesting a key role in disease pathogenesis.

Conclusion: This study identified ten hub genes with potential as biomarkers for the diagnosis and treatment of UC. NCF2 may contribute to UC pathogenesis by modulating inflammatory and immune responses, serving as a potential immune-related biomarker for improved UC diagnosis and treatment.

Keywords: ulcerative colitis, biomarkers, immune microenvironment, machine learning

Introduction

Ulcerative colitis (UC) is a chronic inflammatory bowel disease (IBD) characterized by recurrent episodes of relapse and remission. As of 2023, it is estimated that approximately 5 million individuals are affected by UC globally, and its prevalence continues to increase.¹⁻³ The diagnosis of UC depends on a comprehensive assessment of clinical symptoms, endoscopic observations, and histopathological features, which poses challenges in distinguishing it from other intestinal

disorders. Therefore, there is an urgent need to develop novel biomarkers to enhance the diagnostic accuracy and improve the precision of UC treatment.

The pathogenesis of UC is complex and remains incompletely understood. Current treatments primarily aim to induce and maintain remission; however, the success rate of induction therapy in clinical trials typically does not exceed 20%–30%, and real-world studies report remission rates of only 30%–60%.¹ Evidence suggests that dysregulation of the intestinal mucosal immune response plays a critical role in the initiation and progression of UC.^{4,5} The colonic mucosal immune system is finely tuned to balance immune tolerance and immune activation. Disruption of this equilibrium can result in acute or chronic inflammation, characterized by mucosal ulceration, bleeding, and diarrhea. Therefore, a deeper understanding of the immunological features of UC is essential for elucidating its pathological mechanisms and developing targeted therapeutic strategies. Currently, UC immunotherapy mainly relies on immunosuppressants and biologics (such as anti-TNF- α antibodies, anti-IL-12/23 antibodies, etc). However, these treatments often face challenges such as unstable efficacy, resistance, and potential side effects.^{6,7} Although immunotherapy has improved clinical symptoms in some UC patients, it still cannot address the treatment needs of all patients. As a result, exploring precise immunotherapy strategies—especially through the study of the immune microenvironment and immune biomarkers—has become crucial for enhancing treatment outcomes and achieving personalized treatment.⁸

With the rapid development of microarray technology and bioinformatics, machine learning has been widely employed for high-throughput data analysis and biomarker identification, owing to its superior classification performance,^{9–11} and has yielded significant advancements. Although numerous studies have investigated biomarkers in UC, several limitations remain. First, most existing studies are based on relatively small sample sizes (typically fewer than 100 cases), making them susceptible to statistical bias and increasing the likelihood of false-negative results.¹² Second, imbalanced sample class distributions hinder the training efficiency and generalizability of machine learning models.¹³ Third, the functional roles of identified biomarkers and their regulatory mechanisms within the immune system remain poorly understood, limiting further insights into the pathogenesis of UC. Moreover, many studies lack validation using clinical samples, thereby reducing the translational value of their findings.

In light of the above context, this study integrated multiple GEO datasets to enlarge the sample size and applied a range of machine learning algorithms to identify reliable diagnostic biomarkers for UC. To address class imbalance during model development, a class reweighting strategy was implemented by configuring the `class_weight` parameter in the feedforward neural network (FNN) model. Unlike previous studies that primarily compared immune features between UC patients and healthy controls, our study further assessed hub genes in relation to the immune microenvironment, using CIBERSORT, ssGSEA, ESTIMATE, and inflammatory response scores. Immunohistochemical (IHC) assays were then performed on clinical samples, providing experimental validation for the computationally identified hub genes.

Collectively, these findings provide a strong theoretical basis for the precise diagnosis and personalized treatment of UC. The overall study design and analytical workflow are illustrated in [Figure 1](#).

Materials and Methods

Data Source

The gene expression data used in this study were obtained from the Gene Expression Omnibus (GEO) database of the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/geo/>). Six UC-related gene expression datasets were included: GSE48959 (13 UC samples and 8 normal controls), GSE66407 (161 UC samples and 99 normal controls), GSE87466 (87 UC samples and 21 normal controls), and GSE107499 (75 inflamed UC tissue samples and 44 non-inflamed UC tissue samples), with GSE75214 (97 UC samples and 22 normal controls) and GSE47908 (39 UC samples and 15 normal controls) used as external validation datasets. All datasets were derived from colon tissue samples and encompass gene expression profiles from UC patients, healthy controls, and tissues representing different inflammatory conditions.

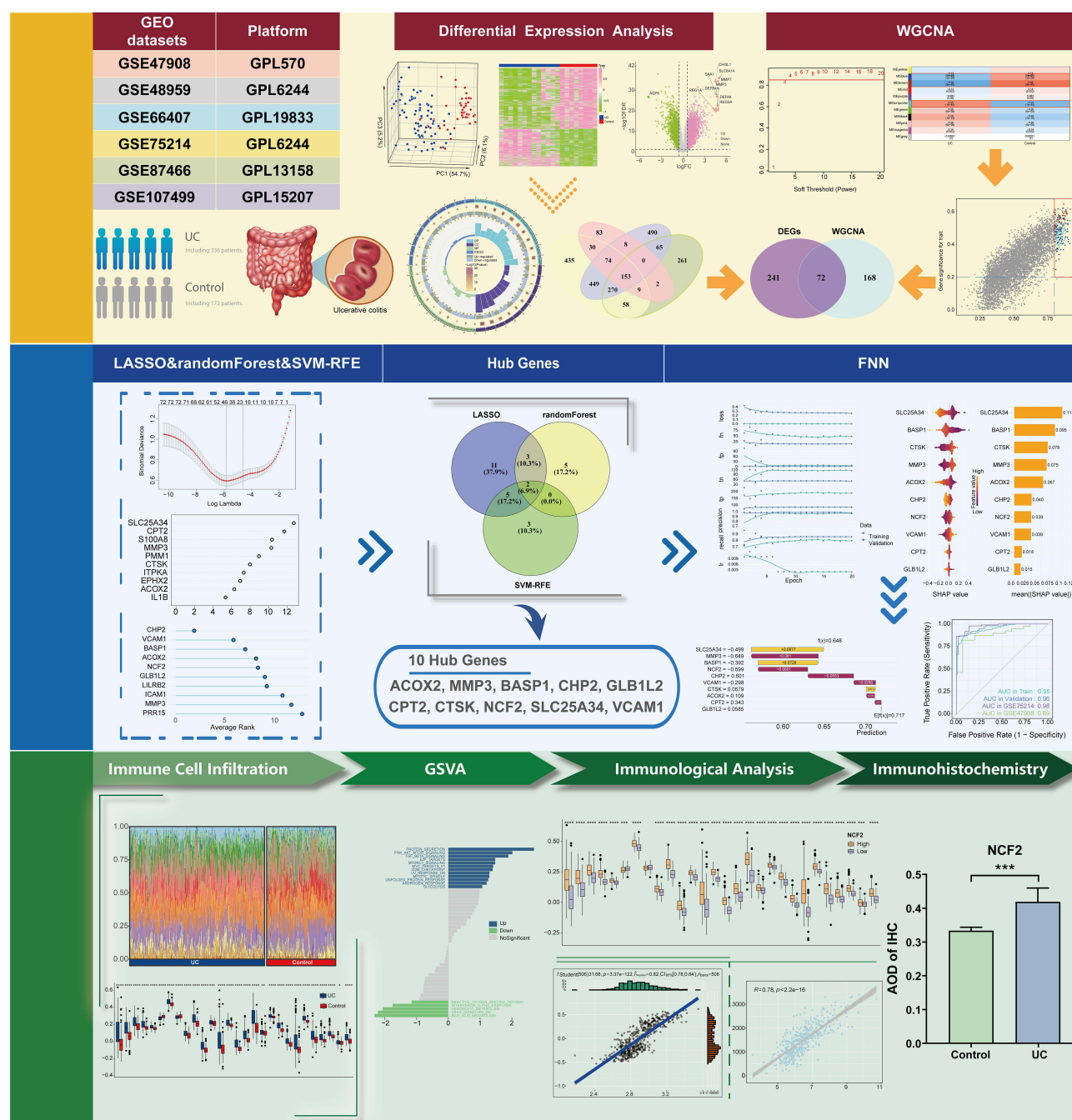


Figure 1 The workflow of this study. In the Immunological Analysis panel, adjusted P values (Benjamini-Hochberg method) are indicated as $P < 0.05$ (*), < 0.01 (**), < 0.001 (***), and < 0.0001 (****). In the immunohistochemistry panel, data are presented as Mean \pm SD; *** indicates $P < 0.001$.

Differential Expression Analysis and Visualization

Differential expression analysis was performed for each dataset using the GEO2R online tool, with a threshold of FDR < 0.05 and $|\log_2\text{FC}| > 0.585$. Expression matrices were retrieved using the R package “GEOquery”,¹⁴ and gene annotation and ID conversion were conducted with “org.Hs.eg.db”. To eliminate analytical bias, data duplications were removed using the *duplicated()* function from the base R package. To assess sample distribution patterns and potential batch effects, three-dimensional principal component analysis (PCA) plots were generated using the “plot3D” package.¹⁵ Volcano plots and heatmaps were employed to visualize the distribution of differentially expressed genes (DEGs)

between the UC group and the control group. Venn diagrams were used to integrate DEGs from four datasets, enabling the identification of commonly upregulated and downregulated genes.

Functional Enrichment Analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted using the R package “clusterProfiler”,¹⁶ with an adjusted *p*-value threshold of < 0.05 . GO analysis was performed to elucidate gene functions across three categories: biological processes (BP), cellular components (CC), and molecular functions (MF). KEGG analysis was employed to identify the signaling pathways and molecular interaction networks involving the associated genes.

Expression Matrix Integration and Batch Effect Correction

Four GEO datasets (GSE48959, GSE66407, GSE87466, and GSE107499) were integrated by identifying common genes across all datasets. The resulting expression matrix was normalized using the `normalizeBetweenArrays` function from the R package “limma”,¹⁷ yielding an integrated UC expression dataset. To reduce the impact of technical variability on downstream analyses, batch effects were corrected using the ComBat algorithm implemented in the R package “sva”.¹⁸

Weighted Gene Co-Expression Network Analysis (WGCNA)

The R package “WGCNA”¹⁹ was used to construct a weighted gene co-expression network. Initially, Pearson correlation coefficients were computed between gene pairs to generate a correlation matrix, which was then transformed into an adjacency matrix using a power function to ensure the resulting network adhered to a scale-free topology. The optimal soft-thresholding power (β) was selected based on the scale-free topology fit index (R^2) and average connectivity. The adjacency matrix was subsequently converted into a topological overlap matrix (TOM), and co-expression modules were identified through hierarchical clustering. Genes within modules significantly associated with UC were selected using thresholds of Module Membership ($MM > 0.8$) and Gene Significance ($GS > 0.2$). These genes were intersected with DEGs to derive the final set of candidate genes.

Screening of Hub Genes Using Multiple Machine Learning Methods

To identify hub genes, feature selection was performed on the candidate gene set using three machine learning approaches: least absolute shrinkage and selection operator (LASSO), support vector machine-recursive feature elimination (SVM-RFE), and randomForest (RF). For all analyses, random seeds were fixed to ensure reproducibility.

LASSO: The R package “glmnet”²⁰ was used to perform 10-fold cross-validation, selecting feature genes corresponding to the largest regularization parameter (λ_{1se}) within one standard error of the minimum cross-validated error. The alpha parameter was set to 1 for LASSO regression (L1 regularization).

RF: A randomForest model was trained using the R package “randomForest”²¹ with 100 trees, and feature importance was assessed based on the Mean Decrease in Gini index.

SVM-RFE: Implemented with the R packages “e1071”²² and “caret”,²³ using 10-fold cross-validation and a stepwise elimination approach for feature selection. Hub genes were defined as those identified by at least two of the three algorithms, a strategy that balances robustness of selection with the dimensionality required for downstream FNN modeling. Their overlaps were visualized using a Venn diagram.

FNN Model Construction and Performance Evaluation

The FNN model was constructed using the expression matrix of the identified hub genes and built with the R package “Keras”.²⁴ The architecture consisted of an input layer, three hidden layers, and a sigmoid-activated output layer, with dropout applied to reduce overfitting. The model was trained with the Adam optimizer and binary cross-entropy loss on a dataset randomly divided into 70% training and 30% validation. Optimization strategies included early stopping, adaptive learning rate reduction, and model checkpointing. To address class imbalance, class weights were defined inversely proportional to class frequency, with each class *c* assigned a weight proportional to $1/n_c$. In addition, random seeds were fixed to ensure reproducibility. Model performance was evaluated using accuracy, sensitivity, specificity,

precision, recall, area under the curve (AUC), and counts of true/false positives and negatives. Model interpretability was assessed with SHapley Additive exPlanations (SHAP). To examine generalization and cross-platform robustness, the model was further validated on two independent external cohorts (GSE47908 and GSE75214). Detailed architecture and hyperparameter settings are provided in [Supplementary Table S1](#).

Immune Microenvironment Analysis

To characterize the immune microenvironment in UC and its association with hub genes, we applied several computational tools, including the R packages “CIBERSORT”,²⁵ “ssGSEA”, “GSVA”,²⁶ and “ESTIMATE”.²⁷ CIBERSORT, using the LM22 signature, quantified immune cell infiltration across groups. ssGSEA calculated enrichment scores for 28 immune cell types to assess immune status. *P* values obtained from CIBERSORT and ssGSEA comparisons were adjusted using the Benjamini–Hochberg false discovery rate (BH-FDR) method. Gene set variation analysis (GSVA) was performed to evaluate the activity of Hallmark pathways from the Molecular Signatures Database (MSigDB), thereby identifying immune-related pathways associated with UC. Samples were stratified by hub gene expression levels, and inflammation response scores were calculated using the “Hallmark inflammatory response” gene set. Immune scores were computed via the ESTIMATE algorithm. Correlation analyses were then performed to explore associations among hub gene expression, inflammation response, and immune scores.

Immunohistochemical (IHC) Validation

To assess the protein expression levels of hub genes, IHC analysis was retrospectively performed on clinical tissue samples. The study included 12 patients diagnosed with UC by pathological examination, who underwent surgery or colonoscopy at the Third Affiliated Hospital of Liaoning University of Traditional Chinese Medicine between March and May 2025, as well as 5 healthy control individuals. Formalin-fixed, paraffin-embedded tissue sections were collected from all subjects. The study was approved by the Ethics Committee of the Third Affiliated Hospital of Liaoning University of Traditional Chinese Medicine (Approval No.: LLSL-ZY-GC-2025-005-01), and written informed consent was obtained from all participants; all procedures were conducted in accordance with the Declaration of Helsinki.

IHC staining was carried out according to standard protocols, including deparaffinization, rehydration, antigen retrieval, and blocking. Sections were incubated with primary antibodies (15551-1-AP, Proteintech, Wuhan Sanying, China) overnight at 4°C, followed by visualization using 3,3'-diaminobenzidine, hematoxylin counterstaining, and mounting. Tissue images were digitized using the ImageScope scanning system. For each slide, three regions were randomly selected at 10× and 40× magnification. The average optical density (AOD) was measured using ImageJ software to assess staining intensity, categorized as negative, weak, moderate, or strong. The combined score was calculated by multiplying the intensity score by the percentage of stained area.

Statistical Analysis

All statistical analyses were conducted in R software (version 4.3.3). The Wilcoxon rank-sum test was employed for the assessment of continuous variables, and Spearman's rank correlation coefficient was used for correlation analysis. *p* < 0.05 was considered statistically significant.

Results

Screening for DEGs

Principal component analysis (PCA) results ([Figure 2A](#)) demonstrated a clear separation between the UC and control groups at the transcriptomic level across all four GEO datasets. Samples within each group exhibited tight clustering without evident outliers, indicating that the data were consistent and reliable. Differential expression analysis identified 2013, 1771, 5451, and 6496 DEGs in the four datasets, respectively ([Figure 2B and C](#)). Heatmaps revealed distinct expression differences between the UC and control groups in each dataset. By intersecting the DEGs from all four datasets, a total of 313 common differentially expressed genes (co-DEGs) were identified, comprising 160 upregulated and 153 downregulated genes ([Figure 2D](#)).

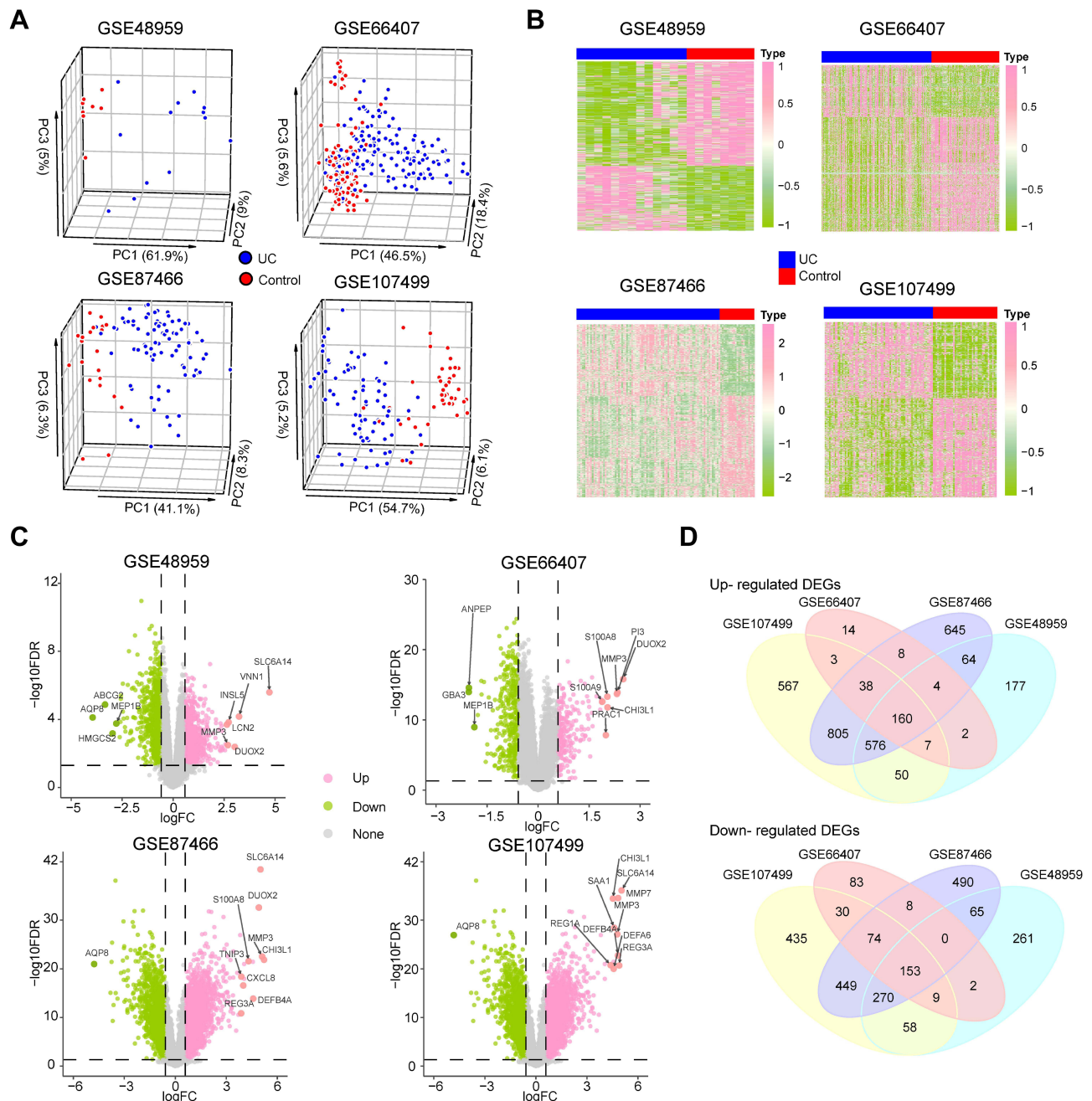


Figure 2 Differentially Expressed Gene Analysis Results. **(A)** PCA plot shows the separation of UC and control samples across datasets. **(B)** Heatmap reveals clustering patterns of DEGs between UC and control groups. **(C)** Volcano plot identifies significantly up- and downregulated genes. **(D)** Venn diagram indicates 313 co-DEGs shared across datasets, with 160 upregulated and 153 downregulated.

Functional Enrichment Analysis of Co-DEGs

To elucidate the pathogenic mechanisms of UC, GO and KEGG enrichment analyses were performed on 313 co-DEGs. GO analysis showed significant enrichment in biological processes related to immune responses and external stimuli, as well as components and functions associated with extracellular signaling (Figure 3A). KEGG analysis revealed involvement in key pathways, including cytokine–cytokine receptor interaction, IL-17, TNF, NF- κ B signaling, and fatty acid metabolism. Circos plots (Figure 3B) highlighted differential enrichment trends between up- and downregulated genes. Network analyses further demonstrated strong connections among immune-related GO terms (Figure 3C) and

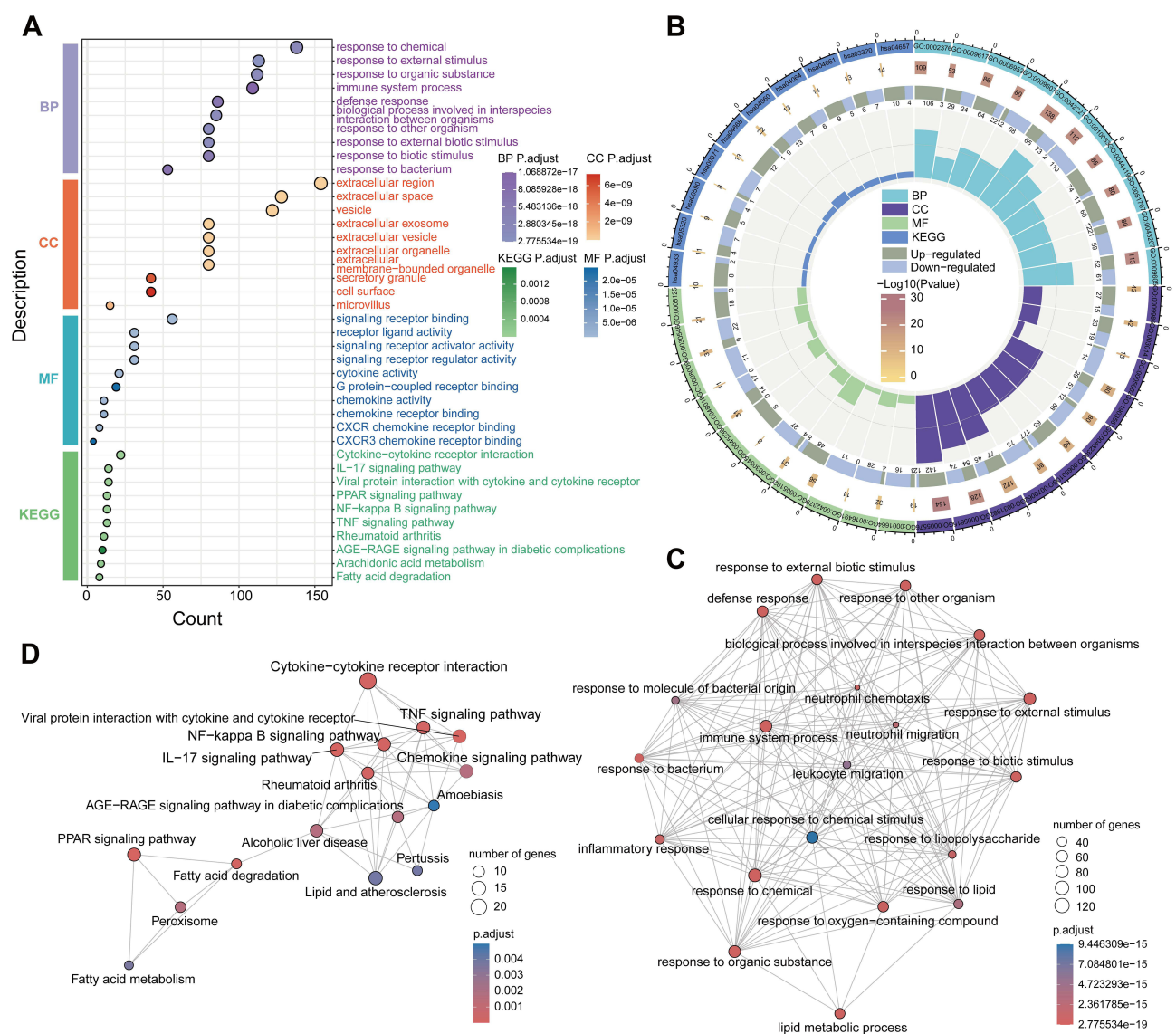


Figure 3 Functional Enrichment Analysis of co-DEGs. **(A)** Bubble plot shows GO (BP, CC, MF) and KEGG pathways significantly enriched in co-DEGs. **(B)** Circos plot presents the distribution of up- and downregulated co-DEGs across functional terms. **(C)** GO network diagram illustrates term significance, gene count per term, and gene overlap. **(D)** KEGG network diagram reveals inter-pathway associations based on shared genes.

interactions across immune and metabolic pathways in KEGG (Figure 3D). These findings suggest that co-DEGs may contribute to UC pathogenesis through coordinated regulation of immune-inflammatory and metabolic pathways.

WGCNA Analysis Identifies UC-Associated Co-Expression Modules and Modules Genes

WGCNA was performed using merged expression data from four GEO datasets. Sample clustering confirmed high gene expression consistency (Figure 4A), and a soft-thresholding power of 5 yielded a scale-free topology ($R^2 > 0.85$) with strong network connectivity (Figure 4B). Several co-expression modules were identified (Figure 4C), among which the blue ($r = -0.47$, $p = 3 \times 10^{-29}$), brown ($r = -0.64$, $p = 9 \times 10^{-60}$), and turquoise ($r = 0.55$, $p = 1 \times 10^{-40}$) modules exhibited distinct module characteristics (Figure 4D and E). Correlation analysis showed significant negative associations between the blue and brown modules and UC, whereas the turquoise module was positively correlated (Figure 4F). Based on MM > 0.8 and GS > 0.2 , 240 key genes were identified, mainly within these three modules (Figure 4G). KEGG enrichment analysis (Supplementary Figure S1A and C) showed that these genes were primarily involved in metabolic and

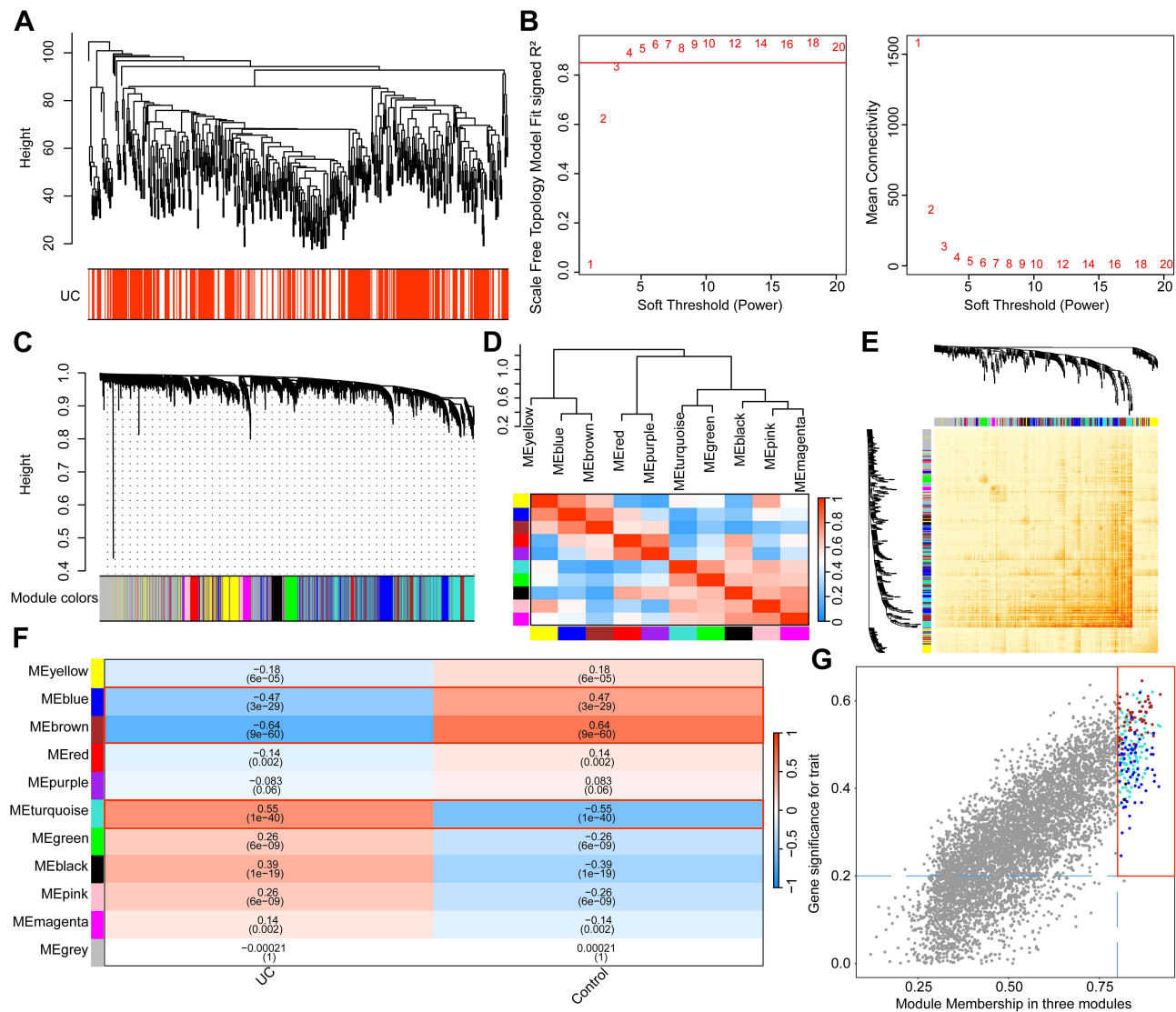


Figure 4 Identification of UC-Associated Gene Co-expression Modules Using WGCNA. (A) Sample clustering and UC status heatmap show sample groupings and disease classification. (B) Network topology analysis defines optimal soft-thresholding for scale-free network construction. (C) Gene clustering dendrogram identifies modules via dynamic tree cutting. (D) Module eigengene heatmap reveals correlations among modules. (E) Gene co-expression heatmap illustrates global expression relationships. (F) Correlation heatmap links modules to UC status based on Pearson coefficients. (G) MM vs GS scatter plot highlights 240 key genes associated with UC.

inflammatory pathways, including fatty acid degradation, lipid metabolism, and NF- κ B signaling. GO analysis (Supplementary Figure S1B and D) indicated associations with immune-related biological processes, peroxisomal localization, and molecular functions such as RAGE and fatty acid binding.

Overall, WGCNA effectively identified UC-associated co-expression modules and modules genes, highlighting their involvement in metabolic and immune regulation and providing new insights into UC pathogenesis and biomarker discovery.

Identification of Key Diagnostic Genes by Machine Learning

To identify genes with diagnostic potential, we integrated results from differential expression analysis and WGCNA, then applied LASSO, randomForest, and SVM-RFE for feature selection. As shown in Figure 5A, 313 co-DEGs and 240 WGCNA-derived genes yielded 72 overlapping candidates for model construction. Functional enrichment analysis (Supplementary Figure S2A) indicated that the candidate genes were mainly involved in lipid and organic acid metabolism, extracellular localization, and functions such as oxidoreductase activity and fatty acid binding. These

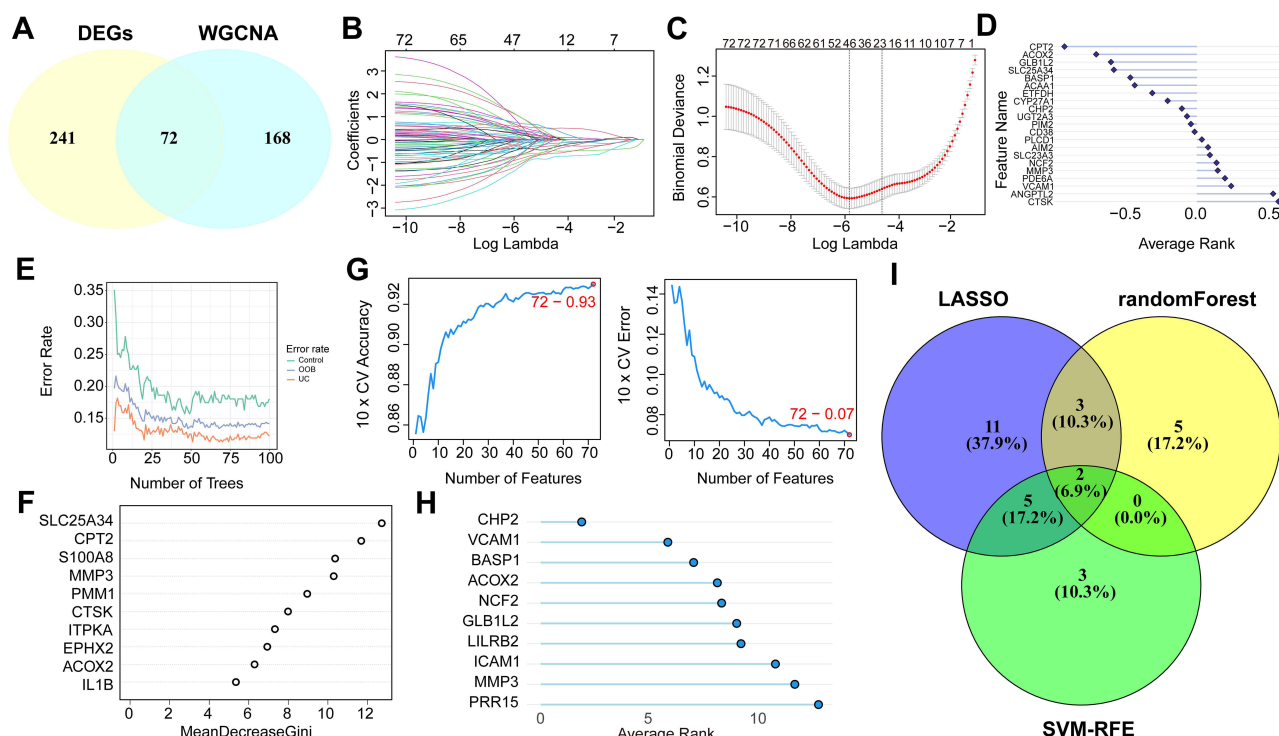


Figure 5 Machine Learning-Based Screening of Key Biomarkers for UC. (A) Venn diagram shows overlap between DEGs and WGCNA modules. (B) LASSO trajectory plot tracks gene coefficient changes across λ values. (C) Cross-validation curve evaluates LASSO model error at different λ . (D) Importance ranking of 21 genes selected by LASSO. (E) randomForest error curve reveals performance across decision tree counts. (F) Top 10 feature genes ranked by MeanDecreaseGini in randomForest. (G) SVM-RFE performance plot assesses accuracy and error by feature count. (H) Top 10 genes selected by SVM-RFE ranked by average importance. (I) Venn diagram compares gene selection results across all three algorithms.

genes were notably enriched in inflammation-related pathways, including TNF, IL-17, and NF- κ B signaling. GO and KEGG network diagrams (Supplementary Figure S2B-C) illustrated the functional-pathway associations, consistent with the inflammatory features of UC. LASSO regression showed that most gene coefficients decreased toward zero as the regularization parameter λ increased (Figure 5B). Ten-fold cross-validation identified the optimal λ ($\lambda_{1se} = 0.0097$) corresponding to the simplest model within one standard error of the minimum error (Figure 5C), resulting in 21 diagnostic genes (Figure 5D). In the randomForest model, the out-of-bag (OOB) error stabilized after ~50 trees (Figure 5E). Based on MeanDecreaseGini scores, the top 10 important genes were selected, including SLC25A34, CPT2, S100A8, MMP3, PMM1, CTSK, ITPKA, EPHX2, ACOX2, and IL1B (Figure 5F). SVM-RFE analysis showed optimal performance when 72 features were selected, achieving 93% accuracy and a 7% error rate in ten-fold cross-validation (Figure 5G). Based on average rank, the top 10 genes were identified, including CHP2, VCAM1, BASP1, ACOX2, NCF2, GLB1L2, LILRB2, ICAM1, MMP3, and PRR15 (Figure 5H). To improve biomarker selection reliability, overlapping genes from all three algorithms were analyzed using a Venn diagram (Figure 5I). Ten genes were shared by at least two methods, with ACOX2 and MMP3 identified by all three, highlighting their strong diagnostic potential.

In summary, integrating three machine learning approaches and cross-validation identified ten hub genes with potential diagnostic value—ACOX2, MMP3, CPT2, CTSK, CHP2, VCAM1, SLC25A34, BASP1, NCF2, and GLB1L2—offering promising targets for UC biomarker development.

Construction and Evaluation of the FNN-Based Prediction Model for UC

To evaluate the diagnostic potential of the 10 hub genes identified by multiple machine learning algorithms, the FNN model was constructed using their expression matrix (Figure 6A). During training, loss steadily declined while accuracy and recall improved (Figure 6B), indicating strong learning and generalizability. SHAP analysis was applied to enhance model interpretability by quantifying each gene's contribution. As shown in Figure 6C, SLC25A34, BASP1, and CTSK

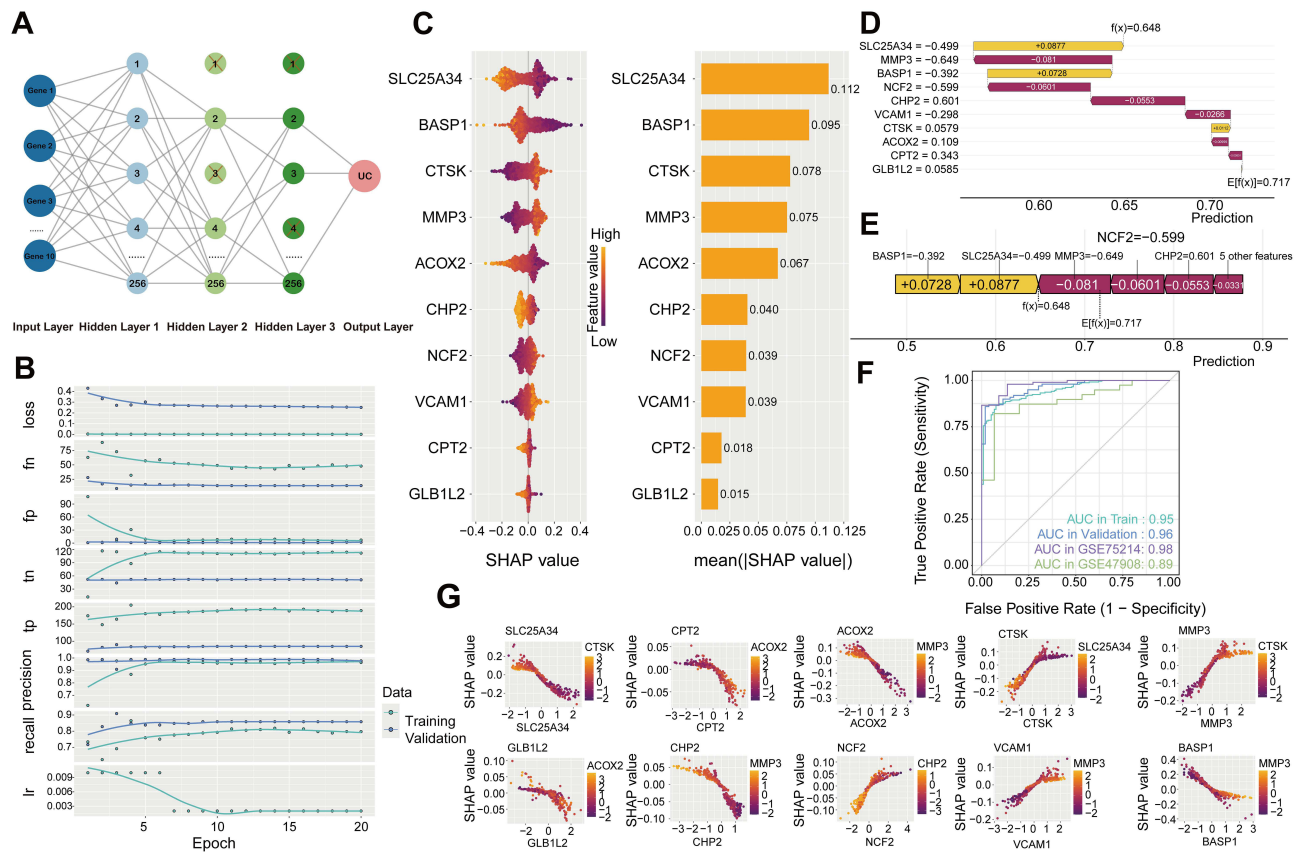


Figure 6 Development and Evaluation of a FNN-Based Prediction Model for UC. **(A)** Diagram of the FNN model architecture with 10 input genes and three hidden layers, where red X marks indicate neurons randomly dropped during training by the dropout regularization process. **(B)** Training curves showing changes in loss, accuracy, and learning rate over epochs. **(C)** SHAP analysis identifies key genes driving model predictions. **(D and E)** SHAP force plots interpret gene-level contributions in a representative sample. **(F)** ROC curves and AUC values evaluate model performance across datasets. **(G)** SHAP dependence plots reveal how gene expression levels influence predictions.

had the highest impact based on mean absolute SHAP values. Force plots (Figure 6D and E) further illustrated how gene-level contributions adjusted prediction probabilities, clarifying the model's decision logic. The FNN model achieved high diagnostic accuracy across the training, internal validation, and external datasets, with AUCs of 0.95, 0.96, 0.98 (GSE75214), and 0.89 (GSE47908), respectively (Figure 6F), demonstrating robust performance. Correlation analysis of SHAP values with gene expression (Figure 6G) highlighted the predictive relevance of genes such as SLC25A34, CPT2, and ACOX2. In addition, bar charts of top SHAP contributors across the integrated GEO datasets and the two external validation cohorts (GSE75214 and GSE47908) confirmed the consistency of gene-level drivers in UC prediction (Supplementary Figure S3A–C). Finally, confusion matrices for the training, internal validation, and external validation cohorts (GSE75214 and GSE47908) further demonstrated reliable predictive accuracy across datasets (Supplementary Figure S4A–D).

Overall, the hub gene-based FNN model enables accurate UC prediction, while SHAP analysis provides interpretability by revealing gene-specific contributions to model output.

Immune Microenvironment Analysis of UC

To characterize the immune microenvironment in UC, we analyzed integrated gene expression data from four GEO datasets. Compared to controls, UC patients exhibited significant changes in immune cell composition based on CIBERSORT analysis, with elevated levels of inflammation-related cells. Specifically, activated CD4+ memory T cells, plasma cells, and M1 macrophages were significantly increased in UC, while memory and naïve B cells were more abundant in controls (Figure 7A–C). Immune status was further assessed using ssGSEA, which revealed

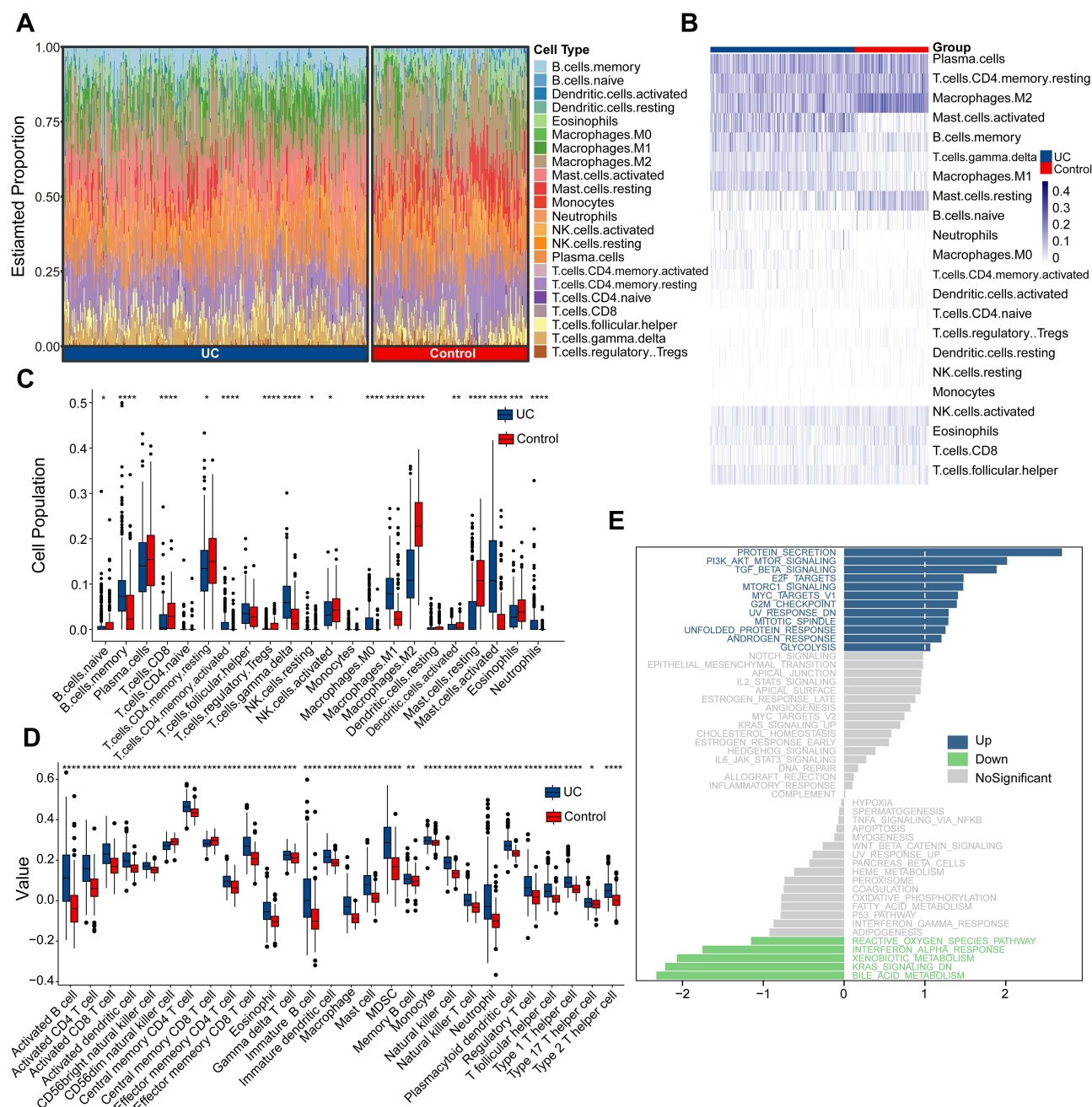


Figure 7 Immune Microenvironment in UC. (A) Stacked bar plot showing the relative proportions of 22 immune cell types across UC and control samples. (B) Heatmap illustrating immune cell infiltration patterns between UC and control groups. (C) Boxplots comparing immune cell proportions between UC and control groups. (D) Boxplots comparing immune status scores between UC and control groups. (E) GSEA of Hallmark pathways showing differential immune-related pathway activities between groups. Adjusted *P* values (Benjamini-Hochberg method) are indicated as *P* < 0.05 (*), < 0.01 (**), < 0.001 (***), and < 0.0001 (****).

upregulation of immunostimulatory cells and reduced immunosuppressive populations in UC (Figure 7D). GSEA revealed pathway-level alterations, with marked upregulation of inflammation-related pathways such as protein secretion, PI3K–AKT–mTOR, TGF- β , and MYC targets ($T > 1$), while pathways related to oxidative stress, fatty acid oxidation, and tryptophan metabolism were downregulated ($T < -1$) (Figure 7E).

These results suggest that the immune microenvironment in UC characterized by inflammatory cell enrichment, diminished immunosuppression, heightened pro-inflammatory signaling, and altered metabolism, which may contribute to disease pathophysiology.

Potential Immunoregulatory Roles of Hub Genes in UC

This study explored the roles of ten hub genes in UC's immune microenvironment by comparing immune profiles between high- and low-expression groups ([Supplementary Figure S5–S13A–D](#)). High expression of SLC25A34, GLB1L2, CHP2, CPT2, and ACOX2 was linked to elevated levels of immunosuppressive cells (eg, resting memory CD4+ T cells, M2 macrophages), reduced pro-inflammatory cell infiltration, and lower inflammation and immune scores. In contrast, elevated BASP1, VCAM1, CTSK, MMP3, and NCF2 expression correlated with increased pro-inflammatory cells (eg, M1 macrophages, neutrophils) and higher immune scores. The expression of all ten hub genes was further validated across four integrated GEO datasets and two independent external cohorts ([Supplementary Figure S14A](#)). Specifically, MMP3, CTSK, VCAM1, BASP1, and NCF2 were consistently upregulated in UC, while ACOX2, CPT2, CHP2, SLC25A34, and GLB1L2 were downregulated, supporting the robustness of our findings. Among these genes, several (eg, MMP3) have already been validated in UC patient tissues, whereas NCF2, despite its strong association with ROS and immune-inflammatory diseases,^{28,29} remains unverified in clinical UC specimens; thus, we selected it for in-depth analysis. Based on NCF2 expression levels, CIBERSORT revealed enrichment of neutrophils, M1 macrophages, and activated mast cells in the high-expression group, whereas the low-expression group showed higher proportions of resting memory CD4+ T cells, M2 macrophages, and resting mast cells ([Figure 8A](#)). ssGSEA further indicated enhanced

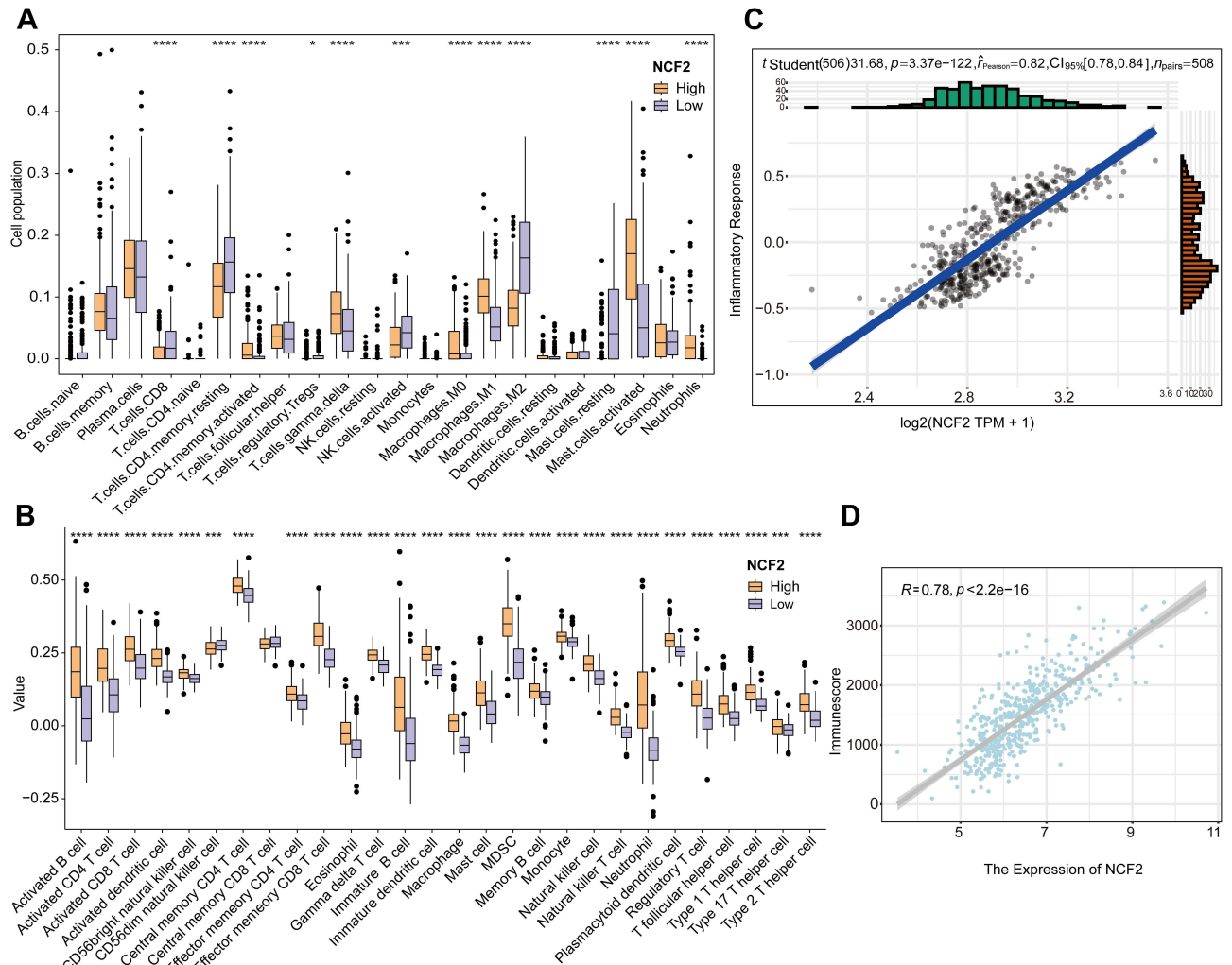


Figure 8 Association Between NCF2 Expression and Immune Microenvironment. **(A)** Immune cell infiltration compared between high and low NCF2 expression groups. **(B)** Immune status scores compared between high and low NCF2 expression groups. **(C)** Correlation between NCF2 expression and inflammatory response score. **(D)** Correlation between NCF2 expression and immune score. Adjusted *P* values (Benjamini-Hochberg method) are indicated as $P < 0.05$ (*), < 0.001 (**), and < 0.0001 (***).

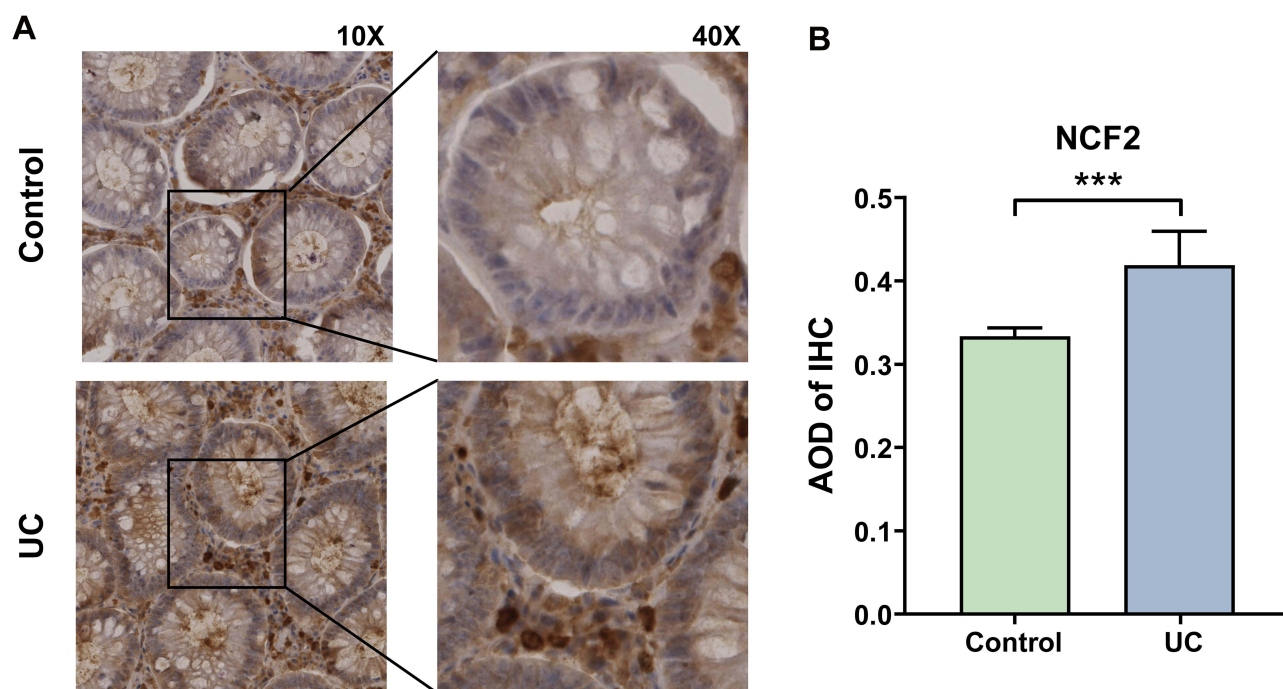


Figure 9 IHC Validation of NCF2 Expression in Colonic Tissues from UC and HC Groups. **(A)** Representative IHC images of NCF2 staining in UC and HC tissues ($\times 10$, $\times 40$). **(B)** Quantitative analysis of NCF2 expression based on AOD values (Mean \pm SD, *** $P < 0.001$).

immune activity in activated B cells and memory T cells with elevated NCF2 expression (Figure 8B). Correlation analysis using the Hallmark inflammatory response gene set revealed a strong positive association between NCF2 expression and inflammation response scores ($r = 0.82$, $p = 3.37 \times 10^{-122}$) (Figure 8C). ESTIMATE-derived immune scores were also strongly correlated with NCF2 expression ($r = 0.78$, $p < 2.2 \times 10^{-16}$) (Figure 8D).

Overall, the ten hub genes are pivotal in immune regulation in UC, with NCF2 potentially contributing to disease pathophysiology through its association with increased pro-inflammatory cell infiltration and enhanced immune activation.

IHC Validation of NCF2 Expression

IHC analysis demonstrated that NCF2 was positively expressed in the intestinal tissues of patients with UC, with predominant localization in the cytoplasm of epithelial cells and in infiltrating immune cells within the inflamed mucosa. The staining intensity was generally moderate, although certain regions exhibited strong staining, while tissues from the healthy control (HC) group showed only mild staining (Figure 9A). Quantitative analysis further indicated that the AOD in the UC group was significantly higher than that in the HC group, with a statistically significant difference ($p < 0.001$) (Figure 9B). These findings support the observation of elevated NCF2 expression in UC tissues, suggesting a potential role for NCF2 in the pathogenesis of UC.

Discussion

Ulcerative colitis (UC), a major subtype of inflammatory bowel disease (IBD), is characterized by chronic, relapsing inflammation resulting from dysregulated immune responses in the intestinal mucosa.³⁰ A hallmark of early-stage IBD is the excessive infiltration of polymorphonuclear leukocytes, particularly neutrophils, which contribute to disease onset and progression by disrupting the epithelial barrier, inducing oxidative stress, mediating proteolytic tissue damage, and releasing pro-inflammatory mediators.³¹ Increasing evidence suggests that abnormal immune cell infiltration plays a central role in UC pathogenesis.^{32,33} However, the lack of specific biomarkers that reliably reflect immune dysregulation remains a major limitation in clinical practice, impeding early diagnosis and the development of targeted therapies.³⁴

Therefore, identifying molecular markers closely associated with the immune microenvironment in UC holds significant clinical value.

This study integrated gene expression profile data from multiple GEO datasets and employed a systematic bioinformatics framework, including differential expression analysis, WGCNA, and multiple machine learning algorithms. Using this approach, ten genes strongly associated with UC were identified: ACOX2, MMP3, CPT2, CTSK, CHP2, VCAM1, SLC25A34, BASP1, NCF2, and GLB1L2. The FNN model constructed with these genes demonstrated high classification accuracy across training, internal validation, and external validation cohorts, highlighting their potential utility as robust diagnostic biomarkers for UC.

Further immunological analyses revealed distinct patterns of immune cell infiltration and altered immune states in patients with UC. Specifically, activated CD4⁺ memory T cells, M1 macrophages, and plasma cells were significantly enriched in UC samples, whereas memory B cells and naïve B cells were more prevalent in the control group, indicative of sustained immune activation in UC. Moreover, GSVA-based pathway enrichment analysis showed that inflammation-related pathways—including protein secretion, PI3K-AKT-mTOR signaling, and TGF- β signaling—were markedly upregulated in UC tissues, while metabolic pathways such as reactive oxygen species metabolism, fatty acid oxidation, and tryptophan metabolism were significantly downregulated. These findings suggest an association between immune dysregulation, metabolic imbalance, and UC. The analysis provides an overview of the immune microenvironment in UC, but the causal relationship and underlying mechanisms still require further validation.

This study explored the roles of ten hub genes in the immune microenvironment of ulcerative colitis (UC) and found their expression levels closely linked to immune cell infiltration and inflammation. Elevated expression of SLC25A34, GLB1L2, CHP2, CPT2, and ACOX2 correlated with more immunosuppressive cells and reduced inflammation, suggesting anti-inflammatory or regulatory functions. In contrast, higher levels of BASP1, VCAM1, CTSK, MMP3, and NCF2 were associated with pro-inflammatory cell enrichment and increased inflammation, indicating potential roles in promoting immune activation. These findings suggest that these genes are context-dependent in modulating the UC immune microenvironment. Consistent validation of the ten hub genes across multiple cohorts underscores the robustness of our findings and their potential value as immune-related biomarkers in UC.

Among the hub genes analyzed, NCF2 (neutrophil cytosolic factor 2) is of particular biological significance due to its critical role in inflammatory responses and immune regulation. The p67^{phox} protein encoded by NCF2 is a core component of the NADPH oxidase complex and plays a central role in regulating the production of reactive oxygen species (ROS).^{35,36} While ROS are essential for pathogen elimination,³⁷ excessive ROS production can lead to tissue damage and contribute to the progression of inflammatory diseases such as UC.^{38–40} In our study, high NCF2 expression in UC was significantly associated with immune infiltration and immune status, particularly with pro-inflammatory cells, such as neutrophils and M1 macrophages. This elevated NCF2 expression was also positively correlated with higher inflammatory gene sets and immune scores. Given the limited research on NCF2 in UC, we further confirmed its elevated expression in UC tissues through IHC, supporting the validity of our findings.

MMP3 (matrix metalloproteinase 3), a matrix metalloproteinase that degrades extracellular matrix components, plays a critical role in tissue remodeling and inflammatory processes through NF- κ B signaling and immune cell activation. Previous studies have shown that MMP3 is closely linked to UC pathogenesis, with serum levels positively correlating with endoscopic activity and being significantly elevated in inflamed mucosal tissues.^{41,42} These findings are consistent with our results, supporting MMP3's pivotal role in UC-related inflammation.

CTSK (cathepsin K), a lysosomal cysteine protease involved in bone matrix resorption, has been implicated in regulating tumor growth and metastasis via the IL-17/CTSK/EMT axis.⁴³ To our knowledge, this is the first study to demonstrate that CTSK is highly expressed in UC, where its expression correlates with increased pro-inflammatory immune infiltration, especially neutrophils and M1 macrophages.

VCAM1 (vascular cell adhesion molecule 1), an adhesion molecule essential for immune surveillance and inflammatory responses, has been shown to correlate with UC disease activity and endoscopic severity, being markedly elevated in active mucosal tissues.⁴⁴ However, inconsistencies remain regarding its serum levels before and after treatment.⁴⁵ Our findings align with this, showing that VCAM1 is upregulated in UC and strongly associated with pro-inflammatory immune infiltration, as well as higher inflammation and immune-related scores.

BASP1 (brain acid-soluble protein 1), a membrane-associated protein involved in immune processes in various cancers,^{46,47} has not previously been studied in UC. Here, we report for the first time that BASP1 is upregulated in UC, with its expression correlating strongly with immune cell infiltration, inflammation scores, and immune activity. This suggests BASP1 plays a key role in UC immunopathogenesis.

ACOX2 (acyl-CoA oxidase 2), a peroxisomal enzyme involved in fatty acid degradation, has been linked to oxidative-stress-related diseases and shown to be significantly decreased in UC mouse models.^{48,49} In our study, reduced ACOX2 expression in UC was associated with increased infiltration of pro-inflammatory immune cells, such as neutrophils and CD8⁺ T cells, consistent with a heightened inflammatory state.

CPT2 (carnitine palmitoyltransferase 2), a key enzyme in mitochondrial long-chain fatty acid oxidation, has been shown to suppress tumor progression by blocking Wnt/ β -catenin signaling and inhibiting ROS/NF- κ B signaling.^{50,51} Our findings demonstrate that CPT2 is downregulated in UC, aligning with prior observations of decreased fatty acid oxidation and carnitine metabolism in UC patients.⁵² Lower CPT2 expression was also associated with greater pro-inflammatory immune infiltration.

CHP2 (solute carrier family 25 member 34), which regulates Na⁺/H⁺ exchange to maintain intracellular pH, has been suggested as a biomarker for Crohn's disease and a predictor of infliximab response in UC.^{53,54} In our study, CHP2 was downregulated in UC, with its lower expression linked to increased neutrophil and M1 macrophage infiltration, further exacerbating the inflammatory environment.

SLC25A34 (solute carrier family 25 member 34), a member of the mitochondrial solute carrier family, is downregulated in colorectal cancer, where its overexpression suppresses malignant phenotypes.⁵⁵ Given that mitochondrial function regulates macrophage inflammatory polarization,^{56,57} our findings of SLC25A34 downregulation in UC and its association with pro-inflammatory immune infiltration suggest a role in immune-metabolic regulation.

GLB1L2 (galactosidase beta 1-like 2), involved in sphingolipid and glycosaminoglycan metabolism, has not been previously studied in UC. In our study, reduced expression of GLB1L2 was associated with increased pro-inflammatory immune infiltration, highlighting its potential role in UC-related inflammation.

Our study identified NCF2 as markedly upregulated in UC and strongly associated with the infiltration of pro-inflammatory immune cells within the immune microenvironment. Given its established role in regulating ROS production, and considering that excessive ROS can activate NF- κ B signaling and amplify downstream inflammatory cascades,^{38,58} these findings indicate that NCF2 warrants further clinical investigation. Recent ROS-targeted therapeutic strategies—including ROS scavengers, nanoparticle-based delivery systems, and intestine-targeted formulations—have also shown promise in mitigating UC-related inflammation.^{7,59} In this context, clarifying the mechanistic role of NCF2 in UC is essential, as it could emerge as a potential stratification biomarker. Moreover, the FNN model built on ten hub genes provides additional diagnostic and risk-assessment value and could be further developed into a clinical decision-support tool to advance precision medicine in UC.

In conclusion, this study employed a multidimensional bioinformatics strategy to systematically identify ten hub genes associated with UC and to construct a robust predictive model, while also highlighting the potential regulatory role of NCF2 within the UC immune microenvironment. These findings not only deepen our understanding of UC pathogenesis but also provide a theoretical basis for the development of personalized diagnostic and therapeutic strategies centered on immune modulation.

This study has several limitations. First, most hub genes lack experimental validation, as only NCF2 was confirmed in clinical samples, although all ten hub genes were consistently validated across GEO and external cohorts. Therefore, further clinical studies are required to establish their diagnostic and mechanistic relevance. Second, considering the challenges of translational application, a more concise biomarker panel may be needed, and refinement using prospective datasets together with additional screening strategies will be important. Third, immune infiltration analysis was based on CIBERSORT bulk RNA-seq deconvolution, which has limited resolution and may not fully capture cellular heterogeneity. Future studies employing higher-resolution approaches, such as single-cell or spatial transcriptomics, combined with functional validation, are expected to provide deeper mechanistic insights and facilitate precision diagnosis and treatment of UC.

Conclusion

This study identified ten genes associated with UC and highlighted their immunological relevance. Elevated NCF2 expression was strongly correlated with immune cell infiltration and inflammatory activity. These findings enhance our understanding of UC pathogenesis and support the development of targeted diagnostic and therapeutic strategies.

Abbreviations

AOD, average optical density; AUC, area under the curve; BP, biological processes; CC, cellular components; co-DEGs, common differentially expressed genes; DEGs, differentially expressed genes; FNN, feedforward neural network; GEO, gene expression omnibus; GO, gene ontology; GSVA, gene set variation analysis; HC, healthy control; IBD, inflammatory bowel disease; IHC, Immunohistochemical; KEGG, kyoto encyclopedia of genes and genomes; LASSO, least absolute shrinkage and selection operator; lambda.lse, largest regularization parameter; MF, molecular functions; MSigDB, the Molecular Signatures Database; NCBI, national center for biotechnology information; NCF2, neutrophil cytosolic factor 2; OOB, out-of-bag; PCA, principal component analysis; RF, randomforest; ROC, receiver operating characteristic; ROS, reactive oxygen species; SHAP, SHapley Additive exPlanations; SVM-RFE, support vector machine-recursive feature elimination; TOM, topological overlap matrix; UC, ulcerative colitis; WGCNA, weighted gene co-expression network analysis.

Data Sharing Statement

The raw and processed data used and analyzed in this study are available from the corresponding author (Yongduo Yu) upon reasonable request. Publicly available datasets analyzed in this study can also be accessed from the GEO database (<https://www.ncbi.nlm.nih.gov/>).

Ethics Approval and Consent to Participate

All human samples and associated data used in this study were collected in compliance with national and institutional ethical guidelines. The study protocol was approved by the Ethics Committee of the Third Affiliated Hospital of Liaoning University of Traditional Chinese Medicine (LLSL-ZY-GC-2025-005-01) and conducted in accordance with the Declaration of Helsinki.

Acknowledgments

We express our gratitude to all the contributors to the Gene Expression Omnibus and MSigDB databases.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This research was supported by the Liaoning Provincial Natural Science Foundation Joint Fund (Doctoral Research Start-up Project, 2023-BSBA-291).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Le Berre C, Honap S, Peyrin-Biroulet L. Ulcerative colitis. *Lancet*. 2023;402(10401):571–584. doi:10.1016/s0140-6736(23)00966-2
2. Kontola K, Oksanen P, Huhtala H, Jussila A. Increasing incidence of inflammatory bowel disease, with greatest change among the elderly: a nationwide study in Finland, 2000–2020. *J Crohn's Colitis*. 2023;17(5):706–711. doi:10.1093/ecco-jcc/jjac177

3. Lophaven SN, Lyng E, Burisch J. The incidence of inflammatory bowel disease in Denmark 1980-2013: a nationwide cohort study. *Aliment Pharmacol Ther.* 2017;45(7):961–972. doi:10.1111/apt.13971
4. Kobayashi T, Siegmund B, Le Berre C, et al. Ulcerative colitis. *Nat Rev Dis Primers.* 2020;6(1):74. doi:10.1038/s41572-020-0205-x
5. Kudo T, Shimizu T. Mucosal immune systems of pediatric inflammatory bowel disease: a review. *Pediatrics Int.* 2023;65(1):e15511. doi:10.1111/ped.15511
6. Gros B, Kaplan GG. Ulcerative colitis in adults: a review. *JAMA.* 2023;330(10):951–965. doi:10.1001/jama.2023.15389
7. Xiao Q, Li X, Li Y, et al. Biological drug and drug delivery-mediated immunotherapy. *Acta pharmaceutica Sinica B.* 2021;11(4):941–960. doi:10.1016/j.apsb.2020.12.018
8. Alsoud D, Verstockt B, Fiocchi C, Vermeire S. Breaking the therapeutic ceiling in drug development in ulcerative colitis. *Lancet Gastroenterol Hepatol.* 2021;6(7):589–595. doi:10.1016/s2468-1253(21)00065-0
9. Wang Y, Xu Y, Yang Z, Liu X, Dai Q. Using recursive feature selection with random forest to improve protein structural class prediction for low-similarity sequences. *Comput Math Methods Med.* 2021;2021:5529389. doi:10.1155/2021/5529389
10. Kong R, Xu X, Liu X, He P, Zhang MQ, Dai Q. 2SigFinder: the combined use of small-scale and large-scale statistical testing for genomic island detection from a single genome. *BMC Bioinf.* 2020;21(1):159. doi:10.1186/s12859-020-3501-2
11. Dai Q, Bao C, Hai Y, et al. MTGPick allows robust identification of genomic islands from a single genome. *Briefings Bioinf.* 2018;19(3):361–373. doi:10.1093/bib/bbw118
12. Rajput D, Wang WJ, Chen CC. Evaluation of a decided sample size in machine learning applications. *BMC Bioinf.* 2023;24(1):48. doi:10.1186/s12859-023-05156-9
13. Kaur H, Pannu HS, Malhi AK. A systematic review on imbalanced data challenges in machine learning. *ACM Computing Surveys.* 2019. doi:10.1145/3343440
14. Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and bioconductor. *Bioinformatics.* 2007;23(14):1846–1847. doi:10.1093/bioinformatics/btm254
15. Soetaert K. plot3D: plotting multi-dimensional data. 2016.
16. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation.* 2021;2(3):100141. doi:10.1016/j.xinn.2021.100141
17. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47. doi:10.1093/nar/gkv007
18. 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(6):882–883. doi:10.1093/bioinformatics/bts034
19. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 2008;9:559. doi:10.1186/1471-2105-9-559
20. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Software.* 2010;33(1):1–22. doi:10.1016/j.jspi.2009.07.020
21. Liaw A, Wiener M. Classification and Regression by randomForest. *R News.* 2002;23(23):18–22.
22. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien [R package e1071 version 1.7-4]. 2020.
23. Kuhn M. Building predictive models in R using the caret package. *J Stat Software.* 2008;28(5):1–26. doi:10.18637/jss.v028.i05
24. Beygelzimer A, Kakade S, Langford J, Arya S, Li S. FNN: fast nearest neighbor search algorithms and applications. 2013.
25. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol.* 2018;1711:243–259. doi:10.1007/978-1-4939-7493-1_12
26. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf.* 2013;14:7. doi:10.1186/1471-2105-14-7
27. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612. doi:10.1038/ncomms3612
28. Bakutenko IY, Haurlych ID, Nikitchenko NV, et al. Neutrophil cytosolic factor 2 (NCF2) gene polymorphism is associated with juvenile-onset systemic lupus erythematosus, but probably not with other autoimmune rheumatic diseases in children. *Mol Genet Genom Med.* 2022;10(1):e1859. doi:10.1002/mgg3.1859
29. Jia Y, Nie J, Wang H, Han Z, Zhang Z, Li L. Correlation study of NCF2 in chronic rhinosinusitis with nasal polyps. *Lin chuang er bi yan hou tou jing wai ke za zhi =J Clin Otorhinolaryngol Head Neck Surg.* 2024;38(4):303–309. doi:10.13201/j.issn.2096-7993.2024.04.008
30. de Souza HS, Fiocchi C, de Souza HSP. Immunopathogenesis of IBD: current state of the art. *Nat Rev Gastroenterol Hepatol.* 2016;13(1):13–27. doi:10.1038/nrgastro.2015.186
31. Brazil JC, Louis NA, Parkos CA. The role of polymorphonuclear leukocyte trafficking in the perpetuation of inflammation during inflammatory bowel disease. *Inflamm Bowel Dis.* 2013;19(7):1556–1565. doi:10.1097/MIB.0b013e318281f54e
32. Yang C, Wang W, Li S, et al. Identification of cuproptosis hub genes contributing to the immune microenvironment in ulcerative colitis using bioinformatic analysis and experimental verification. *Front Immunol.* 2023;14:1113385. doi:10.3389/fimmu.2023.1113385
33. Tang D, Pu B, Liu S, Li H. Identification of cuproptosis-associated subtypes and signature genes for diagnosis and risk prediction of Ulcerative colitis based on machine learning. *Front Immunol.* 2023;14:1142215. doi:10.3389/fimmu.2023.1142215
34. Jones R, Charlton J, Latinovic R, Gulliford MC. Alarm symptoms and identification of non-cancer diagnoses in primary care: cohort study. *BMJ.* 2009;339:b3094. doi:10.1136/bmj.b3094
35. Xu W, Li Y, Wan S, et al. S100A8 induces cyclophosphamide-induced alopecia via NCF2/NOX2-mediated ferroptosis. *Free Radic Biol Med.* 2025;230:112–126. doi:10.1016/j.freeradbiomed.2025.02.014
36. Kong M, Chen X, Lv F, et al. Serum response factor (SRF) promotes ROS generation and hepatic stellate cell activation by epigenetically stimulating NCF1/2 transcription. *Redox Biol.* 2019;26:101302. doi:10.1016/j.redox.2019.101302
37. Bedard K, Krause KH. The NOX family of ROS-generating NADPH oxidases: physiology and pathophysiology. *Physiol Rev.* 2007;87(1):245–313. doi:10.1152/physrev.00044.2005

38. Muro P, Zhang L, Li S, et al. The emerging role of oxidative stress in inflammatory bowel disease. *Front Endocrinol.* **2024**;15:1390351. doi:10.3389/fendo.2024.1390351
39. Wang Z, Wu H, Chang X, et al. CKMT1 deficiency contributes to mitochondrial dysfunction and promotes intestinal epithelial cell apoptosis via reverse electron transfer-derived ROS in colitis. *Cell Death Dis.* **2025**;16(1):177. doi:10.1038/s41419-025-07504-4
40. Segal AW. NADPH oxidases as electrochemical generators to produce ion fluxes and turgor in fungi, plants and humans. *Open Biol.* **2016**;6(5):160028. doi:10.1098/rsob.160028
41. Louis E, Ribbens C, Godon A, et al. Increased production of matrix metalloproteinase-3 and tissue inhibitor of metalloproteinase-1 by inflamed mucosa in inflammatory bowel disease. *Clin Exp Immunol.* **2000**;120(2):241–246. doi:10.1046/j.1365-2249.2000.01227.x
42. Kourkoulis P, Michalopoulos G, Katifelis H, et al. Leucine-rich alpha-2 glycoprotein 1, high mobility group box 1, matrix metalloproteinase 3 and annexin A1 as biomarkers of ulcerative colitis endoscopic and histological activity. *Eur J Gastroenterol Hepatol.* **2020**;32(9):1106–1115. doi:10.1097/meg.0000000000001783
43. Wu N, Wang Y, Wang K, et al. Cathepsin K regulates the tumor growth and metastasis by IL-17/CTSK/EMT axis and mediates M2 macrophage polarization in castration-resistant prostate cancer. *Cell Death Dis.* **2022**;13(9):813. doi:10.1038/s41419-022-05215-8
44. Takagi T, Uchiyama K, Asaeda K, et al. Association of vascular cell adhesion molecule-1 expression in colonic mucosa with mucosal inflammation and subsequent relapse in patients with ulcerative colitis. *J Gastroenterol Hepatol.* **2025**;40(7):1719–1727. doi:10.1111/jgh.17000
45. Umehara Y, Kudo M, Nakaoka R, Kawasaki T, Shiomi M. Serum proinflammatory cytokines and adhesion molecules in ulcerative colitis. *Hepato-Gastroenterol.* **2006**;53(72):879–882.
46. Wang T, Liu X, Wang T, Zhan L, Zhang M. BASP1 expression is associated with poor prognosis and is correlated with immune infiltration in gastric cancer. *FEBS Open Bio.* **2023**;13(8):1507–1521. doi:10.1002/2211-5463.13654
47. Pan X, Xu X, Wang L, et al. BASP1 is a prognostic biomarker associated with immunotherapeutic response in head and neck squamous cell carcinoma. *Front Oncol.* **2023**;13:1021262. doi:10.3389/fonc.2023.1021262
48. Deng B, Zhen J, Xiang Z, et al. Unveiling and validating the role of fatty acid metabolism in ulcerative colitis. *J Inflamm Res.* **2024**;17:6345–6362. doi:10.2147/jir.S479011
49. Du X, Ma Z, Xing Y, et al. Identification and validation of potential biomarkers related to oxidative stress in idiopathic pulmonary fibrosis. *Immunobiology.* **2024**;229(5):152791. doi:10.1016/j.imbio.2024.152791
50. Zhang X, Zhang Z, Liu S, et al. CPT2 down-regulation promotes tumor growth and metastasis through inducing ROS/NFκB pathway in ovarian cancer. *Transl Oncol.* **2021**;14(4):101023. doi:10.1016/j.tranon.2021.101023
51. Li H, Chen J, Liu J, et al. CPT2 downregulation triggers stemness and oxaliplatin resistance in colorectal cancer via activating the ROS/Wnt/β-catenin-induced glycolytic metabolism. *Exp Cell Res.* **2021**;409(1):112892. doi:10.1016/j.yexcr.2021.112892
52. Jiang LJ, Guo MY, Yang H. [Changes in the expression of genes related to intestinal fatty acid oxidation and carnitine metabolism in patients with ulcerative colitis]. *Zhonghua Yi Xue Za Zhi.* **2024**;104(36):3422–3429. Chinese
53. Chen Y, Zhang L, Huang WY, et al. Multiple machine learning models, molecular subtyping and singlecell analysis identify PANoptosis-related core genes and their association with subtypes in crohn's disease. *Curr Med Chem.* **2024**. doi:10.2174/0109298673330894241008060309
54. Chen X, Jiang L, Han W, et al. Artificial neural network analysis-based immune-related signatures of primary non-response to infliximab in patients with ulcerative colitis. *Front Immunol.* **2021**;12:742080. doi:10.3389/fimmu.2021.742080
55. Li C, Yu G, Chen W, Ouyang J, Wang X, Wang Z. E3 ubiquitination ligase MYLIP mediates the NKRF/SLC25A34 axis to suppress malignant progression in colorectal cancer. *Dig Dis Sci.* **2025**;70(2):581–597. doi:10.1007/s10620-024-08735-9
56. Liu W, Tong B, Xiong J, et al. Identification of macrophage polarisation and mitochondria-related biomarkers in diabetic retinopathy. *J Transl Med.* **2025**;23(1):23. doi:10.1186/s12967-024-06038-1
57. Roy N, Alencastro F, Roseman BA, et al. Dysregulation of lipid and glucose homeostasis in hepatocyte-specific SLC25A34 knockout mice. *Am J Pathol.* **2022**;192(9):1259–1281. doi:10.1016/j.ajpath.2022.06.002
58. Morgan MJ, Liu ZG. Crosstalk of reactive oxygen species and NF-κB signaling. *Cell Res.* **2011**;21(1):103–115. doi:10.1038/cr.2010.178
59. Li D, Li J, Chen T, et al. Injectable bioadhesive hydrogels scavenging ROS and restoring mucosal barrier for enhanced ulcerative colitis therapy. *ACS Appl Mater Interfaces.* **2023**;15(32):38273–38284. doi:10.1021/acsami.3c06693