ORIGINAL ARTICLE
Iran J Allergy Asthma Immunol
In press.

Identification of Immune-Related Diagnostic Genes and Subtypes of Active Ulcerative Colitis

Tianlong Tang[1], Xinquan Zan[2], Fei Liu[3], Xiao Shi[1], Qiaoming Zhi[1], and Ye Han[1]

[1] Department of General Surgery, The First Affiliated Hospital of Soochow University, Suzhou, China
[2] Department of General Surgery, The Affiliated Wuxi People's Hospital of Nangjing Medical University, Wuxi, China
[3] Department of Gastroenterology, The First Affiliated Hospital of Soochow University, Suzhou, China

Received: 16 June 2025; Received in revised form: 8 September 2025; Accepted: 18 September 2025

ABSTRACT

Ulcerative colitis (UC) is a chronic immune-mediated disease with a growing global burden. In this study, bioinformatics analysis and machine learning methods were employed to screen the potential immune microenvironment-related feature genes.

We identified 4 immune-related genes that are consistently dysregulated in UC and correlate with immune infiltration, including *APOBEC3B* (Apolipoprotein B mRNA Editing Enzyme Catalytic Subunit 3B), *CXCL11*, *PLA2G2A*, and *TMEM173*. Their diagnostic performance was verified in an external cohort and in our clinical samples.

Then, the proportion of ambiguous clustering (PAC) successfully classified UC patients into 2 molecular subtypes, including subtype 1 (metabolism-related subtype) and subtype 2 (immune-related subtype). The single sample gene set enrichment analysis (ssGSEA) algorithm revealed that subtype 2, with a higher score, of the majority of immune cells presented a worse inflammatory response.

In addition, we assessed scores of partial novel drugs querying the cMAP database and found that the efficacy of clinical small-molecule compounds presented different results across UC subtypes. These findings identify biomarkers, establish a concise immune-based classification of UC, and support subtype-guided therapy.

Keywords: Immunology; Machine learning; Ulcerative colitis

INTRODUCTION

Ulcerative colitis (UC), the main subtype of inflammatory bowel disease (IBD), is characterized as a chronic inflammatory disease that primarily affects the colon and rectum. The clinical manifestations of UC include abdominal pain, diarrhea, bloody stools, and tenesmus. During follow-up, part of some UC patients can develop perforation, sepsis, bowel obstruction, and colitis-associated cancer (CAC).[1,2] The etiology of UC

Corresponding Author: Ye Han, MD;
Department of General Surgery, The First Affiliated Hospital of Soochow University, Suzhou, China Tel: (+86 180) 1311 6265, Fax: (+86 051) 2677 81108, Email: hanyeor@163.com

Qiaoming Zhi, PhD;
Department of General Surgery, The First Affiliated Hospital of Soochow University, Suzhou, China. Tel: (+86 152) 5027 8285, Fax: (+86 051) 2677 81108, Email: strexboy@163.com

The first, second and third authors contributed equally to this study

Published by Tehran University of Medical Sciences (http://ijaai.tums.ac.ir)

is still indeterminate. Epidemiological and experimental studies have shown that multiple risk factors, such as environmental factors, genomic alterations, intestinal epithelial barrier defect, and dysbacteriosis, are closely associated with the initiation and progression of UC.[3,4] But it cannot be ignored that UC is a chronic immune-mediated inflammatory disease, and the complex interactions between the intestinal microbiota (or metabolites) and the host's immune system can mainly promote the disease state.[5,6]

The intestinal surface is constitutively exposed to an enormous diversity of food and microbial antigens. An intact intestinal barrier, including the mucosal barrier and different specialized epithelial cells, is able to separate the intestinal lumen from the lamina propria and protect the human intestine against the potentially hostile antigens. Once the intestinal barrier is injured, its mucosal permeability increases, which can lead to the migration of pathogenic toxins and microorganisms.[7,8] Subsequently, various immune cells accumulate, and multiple inflammatory factors are released. The vertebrate immune response is one of the physiological functions, which can be divided into innate and adaptive immunity, and protects the host from infections and injuries.[9] Innate immunity is the first line of defense against pathogen invasion, and is enacted by different immune sentinel cells (e.g., monocytes, neutrophils, macrophages, mast cells, dendritic cells, innate lymphoid cells, and natural killer cells). This reaction is neither specific nor memorable, but rapid within a short time.[10] Unlike innate immunity, adaptive immune responses involved in IBD development are more critical, time-consuming, precise, and complex.[11] Distinct types of regulatory immune cells, such as $Foxp3^+$ regulatory T cells, suppressor T helper type 17 cells, and regulatory B cells, can respond to the subsequent pathogenic infections after initial antigen-specific stimulation, and aberration of adaptive immunity contributes to the potential inflammation and UC disease.[6,12]

Currently, the therapeutic management for patients suffering from UC consists of the stepwise use of 5-aminosalicylates (5-ASA), corticosteroids, immunomodulators, biologics, and some emerging pharmacotherapies.[13] Despite the reduced morbidity and mortality over the past decades, the patient care in UC remains challenging. For instance, when patients are treated for daily clinical practice, there are still many patients affected by UC who do not response well to therapeutic drugs, and many others loss of efficacy. This means that a considerable proportion of UC patients still receive treatment failure. Due to the complex immune responses presented in UC, we conjectured that deregulation of immunity in the UC progression may influence the pharmacological properties, efficacy in specific population as well as safety.[14-16] In this study, we selected genes based on their significant discriminative value in immune cell infiltration abundance, including *APOBEC3B*, *CXCL11*, *PLA2G2A*, and *TMEM173*, using machine learning methods. Meanwhile, the potential diagnostic role of these feature genes was validated in published cohorts and our collected samples. According to the expressing levels of these feature genes, we subsequently divided UC patients into two immune microenvironment-related subtypes (subtype 1 and subtype2). Finally, the cMAP database was employed to estimate the efficacy of different clinical small-molecule compounds when targeting these two subtypes of UC patients. These data may help investigators search for newly UC molecular classifications, and select more applicable drugs appropriate for personalized treatment modality.

MATERIALS AND METHODS

Data selection

We conducted a comprehensive search for UC datasets in the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) using the search term "ulcerative colitis". Then we applied the following inclusion criteria to screen the datasets: (1) Homo sapiens as the study subject; (2) Expression profiling by array as the experimental method; (3) A minimum of 100 total samples for the discovery cohort; (4) Datasets obtained from different platforms. In addition, we excluded samples from pediatric ulcerative colitis (pediatric UC) to ensure the accuracy and consistency of our analysis. Finally, we obtained three datasets: GSE87466 (comprising 87 UC samples and 21 healthy samples), GSE13367 (comprising 16 UC samples and 20 healthy samples), and GSE16879 (comprising 24 UC samples and 6 healthy samples), respectively (Table 1).

Table 1. The information on microarrays included in this study.

GEO ID	Platform	Samples	Tissues	Attribute
GSE87466	GPL13158	87 active UC and 21 controls	Mucosal	Discovery set
GSE16879	GPL570	24 active UC and 6 controls	Mucosal	Validation set
GSE13367	GPL570	16 active UC and 20 controls	Mucosal	Validation set

UC: ulcerative colitis.

Identification of differentially Expressed Genes (DEGs)

We utilized R software (version 4.2.2) and the R package "limma" (version 3.54.1) to identify differentially expressed genes (DEGs).[17] A statistical significance threshold of adjusted $p<0.05$ and log FC>1.0 was applied. The results were graphically represented using volcano plots generated by the R packages "ggplot2" (version 3.4.1) and "ggrepel" (version 0.9.3). Besides, we employed TBTools (version 1.108, available at https://github.com/CJ-Chen/TBtools/releases/) to partially visualize the DEGs in the form of heatmaps.

Evaluation of Immune Infiltrating Cells and Identification of Characteristic Immune Cells

To explore the different distribution of immune cells between UC samples and healthy samples, we employed the CIBERSORT deconvolution algorithm.[18] The "LM22" dataset file (leukocyte signature matrix) was utilized as a reference to estimate the proportion of 22 immune cell types. Thus, we established the least absolute shrinkage and selection operator (LASSO) regression model, which was implemented using the R package "glmnet" (version 4.1.6) to identify the optimal variables.[19,20] We set an iteration of 1000 and used the minimum criteria to determine the optimal penalty parameter. Finally, the non-zero coefficients were utilized for further analysis.

Weighted Gene Co-expression Network Analysis (WGCNA)

We utilized the R package "WGCNA" (version 1.72.1) to establish a weighted gene co-expression network analysis (WGCNA) to identify the most relevant modules of characteristic immune cells.[21] The top 15 000 genes ranked by standard deviation were selected as input genes. Then, we selected an appropriate soft threshold to ensure that the scale-free topology fit index exceeds 0.85. After that, we made a gene hierarchical clustering dendrogram to identify co-expression modules and calculate the module eigengene, in which modules with a height of less than 0.25 were merged. Finally, we calculated the correlation between module eigengenes and immune cells as well as UC to identify the most relevant modules.

Identification of Immune Microenvironment-related DEGs and Functional Enrichment

To identify immune microenvironment-associated DEGs, we searched for immune-related genes (IRGs) from the InnateDB database (https://www.innatedb.com/), and 1225 IRGs were obtained. The overlapped genes of DEGs, IRGs, and genes from the most immune microenvironment-related modules were considered as immune microenvironment-related DEGs. Subsequently, we performed Gene Ontology (GO) functions enrichment analysis, Kyoto Encyclopedia of Genes Genomes (KEGG) pathways enrichment analysis, and disease enrichment analysis on the overlapped genes of DEGs and genes from immune microenvironment-related modules, using the online tool "Enrichr" (https://maayanlab.cloud/Enrichr/), to develop the potential roles that these genes played in.

PPI Network Construction and KEGG Enrichment Analysis

In order to investigate the potential connections among the immune microenvironment-related DEGs, we searched the STRING database (https://cn.string-db.org/), which was considered a functional protein association network containing almost all known proteins, and enlarged the number of genes to get a general network covering all microenvironment-associated DEGs. After that, KEGG pathways

enrichment analysis was performed via the online tool "Enrichr" to investigate the pathways in which the microenvironment-related DEGs were involved.

Construction of Multiple Machine Learning Models and Identification of Feature Genes

We established 4 machine learning models to identify significant genes. We constructed a LASSO regression model using the same method as described above. The support vector machine recursive feature elimination (SVM-RFE) algorithm, which is a generalized linear classifier for bivariate data based on supervised learning, was employed and performed by the R package "e1071".[22] We constructed an XGBoost model using the R package "caret", in which genes with non-zero importance were selected for further analysis.[23] We also established a RandomForest model using the R package "randomForest", in which we set the number of decision trees to 20 000 to obtain a stable result, and genes with an importance score greater than 1 were considered significant and were selected for further analysis.[24] Finally, genes obtained from the intersection of 4 machine learning models were considered as potential feature genes.

Validation of Feature Genes

To ensure the robustness of our results, we utilized two additional datasets obtained from different platforms: GSE13367 and GSE16879. We compared the gene expression levels between UC samples and healthy samples in both discovery and validation cohorts, which were visualized in boxplots. Additionally, we established receiver operating characteristic (ROC) curves and calculated the area under the ROC curve (AUC) to evaluate the diagnostic values for these feature genes.

Quantitative Real-time Polymerase Chain Reaction (qRT-PCR)

The 57 samples for the UC and 20 human controls were from The First Affiliated Hospital of Soochow University in 2023. The biopsy of 57 samples for the UC group was obtained from the inflamed areas of ulcerative colitis patients who were in active ulcerative colitis, while the biopsy of 20 samples for the HC (Human Control) group was obtained from normal healthy individuals. RNA was extracted from the fresh frozen colonic tissues using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA). The complementary DNA (cDNA) synthesis was conducted using the Prime Script RT Master Mix (TaKaRa, Tokyo, Japan). The primers, which were designed on Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi), were listed as follows: *APOBEC3B*: forward, 5'-CGCCAGACCTACTTGTGCTAT-3', reverse, 5'-CATTTGCAGCGCCTCCTTAT-3'; *CXCL11*: forward, 5'-GAGTGTGAAGGGCATGGCTA-3', reverse, 5'-ACATGGGGAAGCCTTGAACA-3'; *TMEM173* (*STING1*): forward, 5'-GAGTGTGTGGAGTCCTGCTC-3', reverse, 5'-CTGGAGTGGGGCATCTTCTG-3'; *PLA2G2A*: forward, 5'-CTGTCTCCAAACAGCCTTGTG-3', reverse, 5'- CTGCTGGGTGGTCTCAACTT-3'. The FastStart Universal SYBR Green Master (Roche, Basel, Switzerland) was used for the qRT-PCR analysis, and the data were analyzed with the LightCycler 96 System (Roche, Basel, Switzerland). The expression data of genes were normalized to *GAPDH*, and the $2^{-\Delta\Delta Ct}$ method was employed to analyze the relative expressions of target genes.

Identification of Immune Microenvironment-related Subtypes

We identified immune microenvironment-related subtypes using the R package "ConsensusClusterPlus". K value was set from 2 to 9, and the optimal K value was determined by the measure of proportion of ambiguous clustering (PAC). Our classification criteria are based on unsupervised clustering of the expression levels of identified feature genes.

Single Sample Gene set Enrichment Analysis (ssGSEA) and Gene Set Enrichment Analysis (GSVA)

To investigate the differences between immune microenvironment-related subtypes, we evaluated the enrichment scores of 28 immune cell subtypes using the single-sample gene set enrichment analysis (ssGSEA) method as previously reported[25], which was realized by the R packages "GSEABase" and "GSVA", and visualized by the R packages "pheatmap" and "ggpubr". Additionally, we conducted gene set enrichment analysis (GSVA) to uncover the characteristics of subtypes.[26]

Prediction of Small-molecule Compounds

Statistical Analyses

All statistical analyses were performed using R software

(version 4.2.2) and Statistic Package for Social Science (SPSS) software (version 19.0) (IBM, Armonk, NY, USA). The Spearman method was used for correlation analysis, while the Wilcoxon rank-sum test was utilized for comparison analysis. The Pearson χ^2 test was employed to compare the clinical characteristics, and $p<0.05$ was considered statistically significant. P values from group comparisons were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) correction implemented in the base R function p.adjust.

RESULTS

Identification of DEGs between UC and healthy individuals in the discovery dataset

The workflow for our study is presented in Figure 1. We normalized the expression data of GSE87466, GSE13367, and GSE16879 using the "limma" package. The PCA plot revealed a significant biological difference on the PC1 axis between UC and healthy samples (Figure 2A, Supplementary Figure 1A). The DEGs of GSE87466 were visualized in a volcano plot, and a corresponding heatmap containing the top 50 up- and down-regulated genes was also presented in Figure 2B–C. Similar information of GSE13367 and GSE16879 was also shown in supplementary files (Supplementary Figure S1B–C). In these 3 included data sets, we identified a total of 1004 (630 up- and 374 down-regulated genes in GSE87466), 505 (403 up- and 102 down-regulated genes in GSE13367), and 3061 DEGs (1700 up- and 1361 down-regulated genes in GSE16879) in UC samples, respectively, compared to healthy controls.

Identification of Significant Immune Cells between UC and Healthy Individuals

The proportions of 22 included types of immune cells in UC and healthy samples were displayed in Figure 2D. Among them, a total of 15 types of immune-related cells, including dendritic cells activated, mast cells resting, neutrophils, T cells CD4 memory activated, macrophages M1, macrophages M0, B cells naive, T cells follicular helper, T cells CD4 memory resting, NK cells activated, macropahges M2, mast cells resting, plasma cells, T cells CD8 and T cells regulator (Tregs), presented a significant distributing difference between UC samples and healthy samples (Figure 2E–F). 10 variables had a non-zero coefficient in LASSO regression analysis, including macrophages M1, macrophages M0, mast cells activated, dendritic cells activated, T cells CD8, T cells CD4 memory resting, plasma cells, mast cells resting, NK cells activated, and T cells regulatory (Tregs) (Figure 3A–B).

WGCNA and Identification of Feature Genes

We established a sample cluster tree of 108 samples based on the top 15 000 genes ranked by standard deviation. 100 samples were reserved for further analysis, with 8 outliers being rejected. Subsequently, the cluster tree (Figure 3C) of 100 samples was established, and the scale-free topology network and connectivity grew optimally when the soft threshold β was set at 13 (Figure 3D). 30 differently colored gene modules were calculated, in which modules with a height lower than 0.25 were merged. As a result, we achieved 15 different colored gene modules and figured out the r-value between modules and types of immune cells. The dark turquoise module, along with the magenta module, was closely correlated with immune cells because of a statistically significant correlation ($p<0.05$) with all 10 types of immune cells (Figure 3E). Fortunately, our data also showed that the dark turquoise and magenta modules also had close relationships with clinical traits, with a separate r value of 0.6 and 0.78, respectively (Figure 3F). Afterwards, we searched and achieved 1227 immune-related genes from the InnateDB database, and the Venn analysis (DEGs, genes from WGCNA modules, and InnateDB database) successfully screened 19 key genes as potential immune microenvironment-related DEGs (Figure 3G).

PPI Network and Functional Enrichment

We detected the expression levels of 19 immune microenvironment-related DEGs between UC and healthy controls, and all these 19 genes presented a statistically significant difference (Figure 4A). The PPI network displayed the potential interactions among the expanded genes of 19 immune microenvironment-related DEGs, which might partly reveal the possible mechanisms of the immune response to UC (Figure 4B). The 19 immune microenvironment-related DEGs collectively engaged in the pathways of the NF-kappa B signaling pathway, NOD-like receptor signaling pathway, and TNF signaling pathway, which were closely linked to the development of UC (Supplementary Figure 2A–B). After that, we depicted the correlations between the 19 immune

Published by Tehran University of Medical Sciences (http://ijaai.tums.ac.ir)

microenvironment-related DEGs and 22 immune cell subsets. The heatmap implied that most of these 19 microenvironment-related DEGs were closely correlated with immune cells, except Dendritic cells resting and T cells CD4 naive (Figure 4C).

Meanwhile, we also performed the enrichment analysis on the genes obtained from the intersections between DEGs and genes in darkturquoise (and magenta) modules. In the GO enrichment analysis, these genes were mainly involved in neutrophil chemotaxis, granulocyte chemotaxis, neutrophil migration in biological process, secretory granule lumen, specific granule lumen in cellular component, CXCR chemokine receptor binding, chemokine activity, CXCR3 chemokine receptor binding, and chemokine receptor binding in molecular function, which revealed a strong chemotactic movement embedded in individual defense response (Figure 4D). In the KEGG pathway enrichment, these genes were also mainly enriched for some immune-related signals, such as the IL-17 signaling pathway, cytokine-cytokine receptor interaction, chemokine signaling pathway, TNF signaling pathway, and NOD-like receptor signaling pathway, along with inflammation, UC, IBD, CD, and colitis in disease enrichment via the DisGeNET database (Figure 4E).

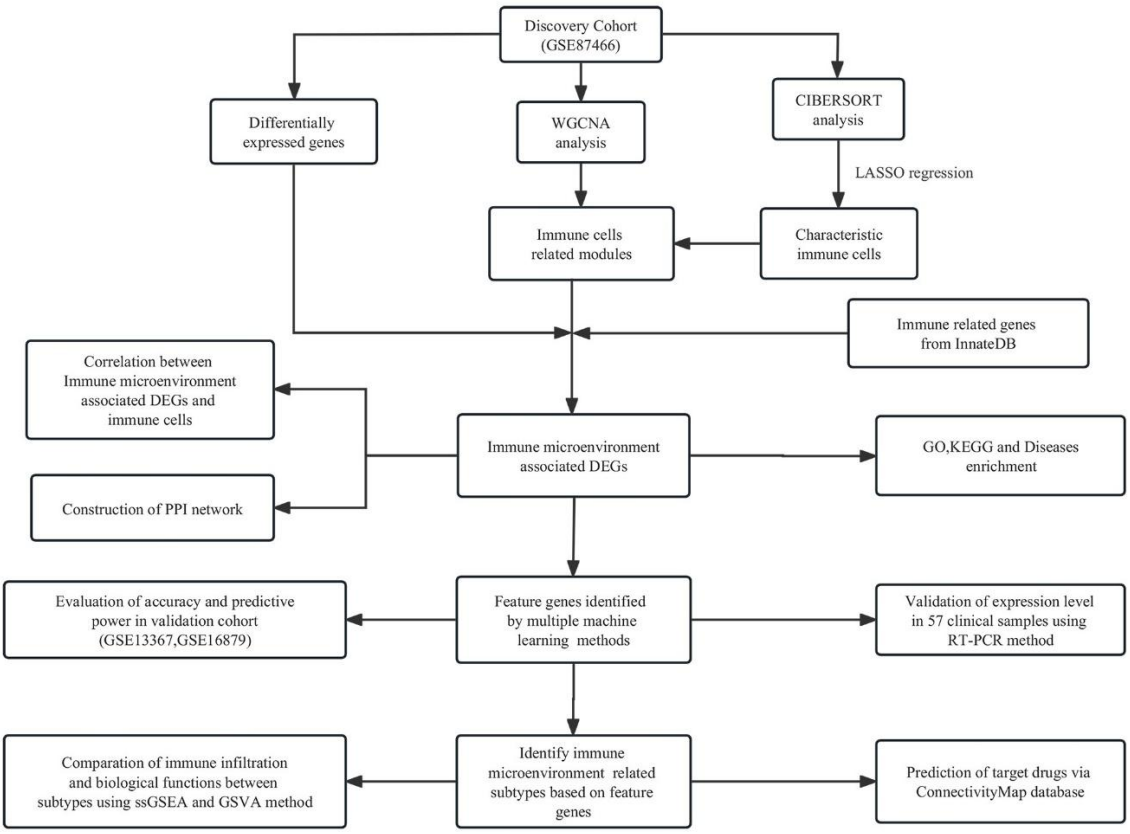


Figure 1. Flowchart of this study. WGCNA: weighted gene co-expression network analysis; CIBERSORT: cell-type identification by estimating relative subsets of RNA transcripts; LASSO: least absolute shrinkage and selection operator; DEGs: differentially expressed genes; PPI: protein–protein interaction; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; ssGSEA: single-sample gene set enrichment analysis; GSVA: gene set variation analysis; RT-PCR: reverse transcription polymerase chain reaction; cMAP: Connectivity Map

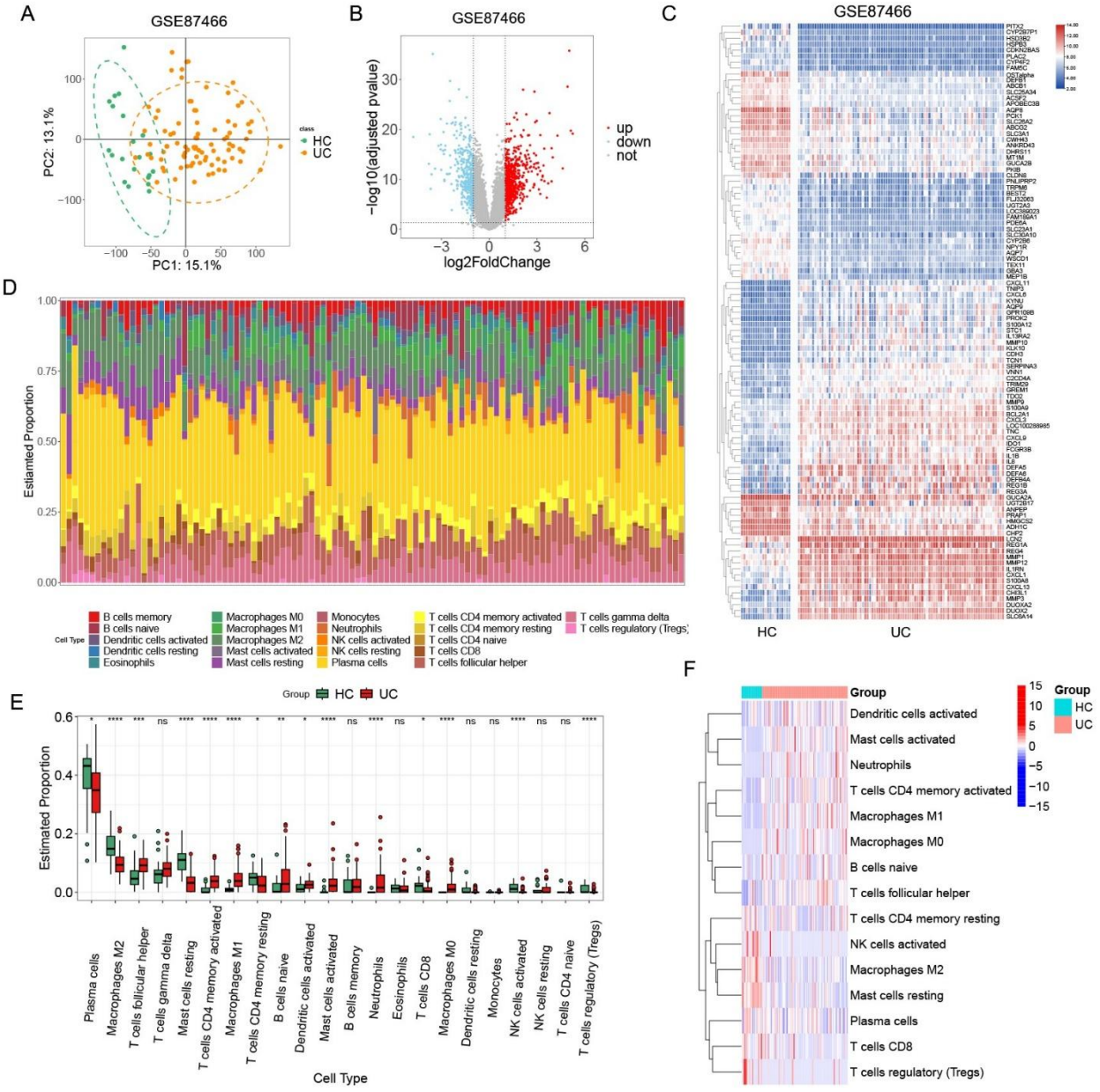


Figure 2. Evaluation of biological differences and immune cell infiltration between ulcerative colitis (UC) and healthy individuals in the discovery dataset. A and B. The principal component analysis (PCA) and volcano plot of the discovery cohort revealed a significant biological difference and differentially expressed genes (DEGs) between UC and healthy samples. C. The top 50 up- and down-regulated genes were presented as a heatmap. D. Barplot showed the proportions of 22 immune cells in the discovery dataset between UC and healthy individuals. E and F. A total of 15 types of infiltrated immune cells showed significant differences (The FDR *q<0.05, **q<0.01, *q<0.001, ****q<0.0001, and ns, no significance). HC: healthy control; FDR: false discovery rate; q: q-value; ns: no significance; NK cells: natural killer cells; M0 macrophages: classically unpolarized macrophages; M1 macrophages: classically activated macrophages; M2 macrophages: alternatively activated macrophages; T cells CD4 memory activated: activated $CD4^+$ memory T cells; T cells CD4 memory resting: resting $CD4^+$ memory T cells; T cells CD8: $CD8^+$ T cells; T cells follicular helper: follicular helper T cells; T cells gamma delta: γδ T cells; T cells regulatory (Tregs): regulatory T cells; B cells memory: memory B cells; B cells naïve: naïve B cells; Dendritic cells resting: resting dendritic cells; Dendritic cells activated: activated dendritic cells.**

Published by Tehran University of Medical Sciences (http://ijaai.tums.ac.ir)

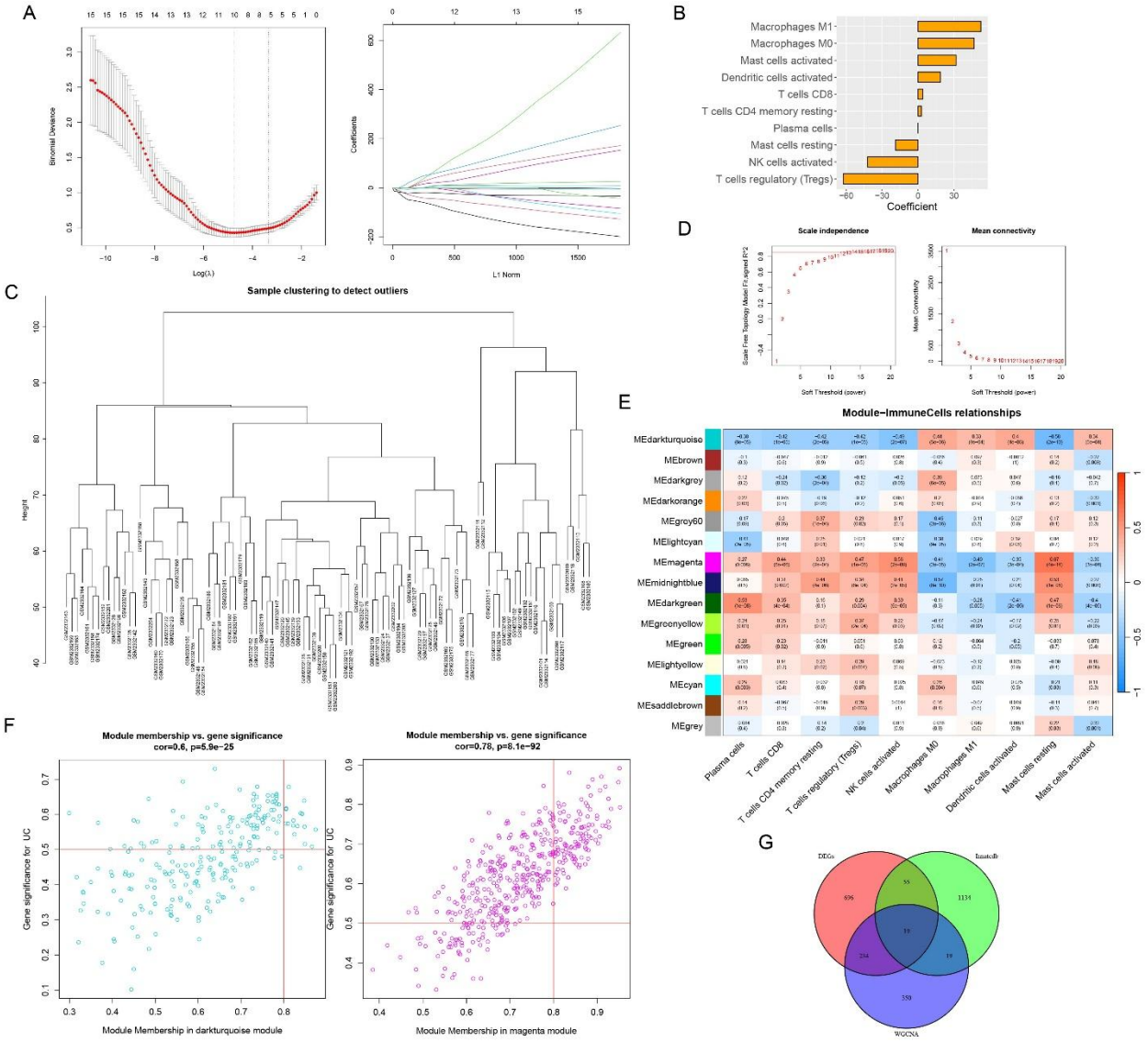


Figure 3. Identification of ulcerative colitis (UC)-related immune cells and 19 associated genes. A. 10-fold cross-validation (10-fold CV) of least absolute shrinkage and selection operator (LASSO) regression analysis and coefficient profiles of 10 differentially expressed immune cells. B. Barplots showed the coefficient of 10 immune cells. C. Cluster trees after rejecting outliers. D. Soft threshold power screening and scale-free network construction. E. Heatmap showed the relationships between modules and immune cells. F. Scatterplot of gene significance in darkturquoise and magenta module, along with the value of correlation between modules and UC. G. Venn diagram finally identified 19 immune microenvironment-associated differentially expressed genes (DEGs) (1004 DEGs, 622 genes from weighted gene co-expression network analysis (WGCNA) modules, and 1227 genes from InnateDB database) for our further experiment. B cells memory: memory B cells; B cells naïve: naïve B cells; Dendritic cells activated: activated dendritic cells; Dendritic cells resting: resting dendritic cells; Macrophages M0: unpolarized macrophages; Macrophages M1: classically activated macrophages; Macrophages M2: alternatively activated macrophages; NK cells: natural killer cells; T cells CD4 memory activated: activated $CD4^+$ memory T cells; T cells CD4 memory resting: resting $CD4^+$ memory T cells; T cells CD8: $CD8^+$ T cells; T cells follicular helper: follicular helper T cells; T cells gamma delta: γδ T cells; T cells regulatory (Tregs): regulatory T cells.

Published by Tehran University of Medical Sciences (http://ijaai.tums.ac.ir)

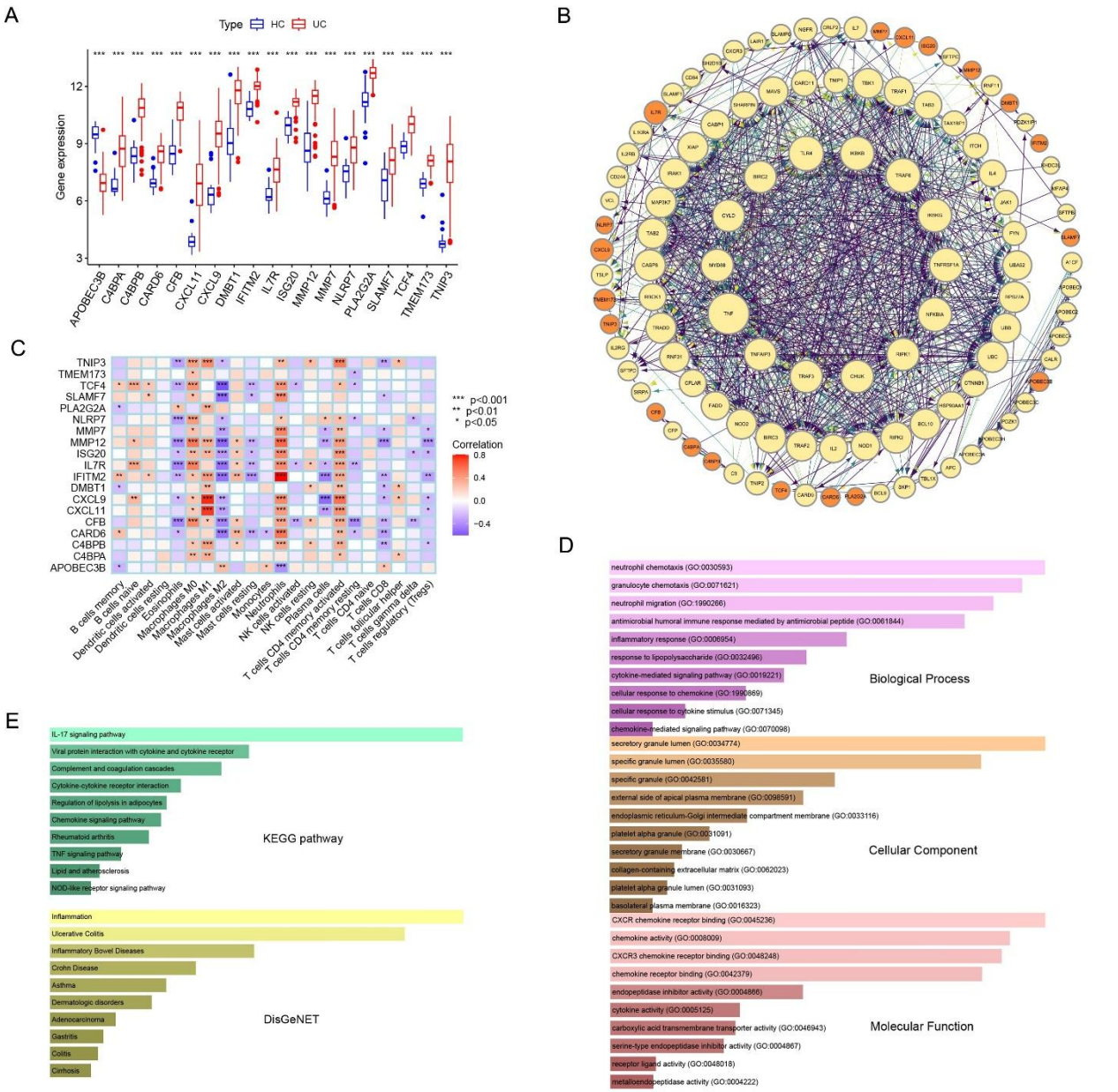


Figure 4. Correlation and functional enrichment. A. The expression levels of 19 immune microenvironment-related genes were compared between ulcerative colitis (UC) and healthy individuals. B. Protein–protein interaction (PPI) network of expanded genes based on the 19 immune microenvironment-related genes. C. Correlations between immune cells and these 19 genes. D and E. Barplots showed the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment ranked by adjusted *p* value. (*p*<0.05, **p*<0.01, ***p*<0.001, ****p*<0.0001, and ns, no significance). Abbreviations: HC, healthy controls; UC, ulcerative colitis; B cells memory, memory B cells; B cells naïve, naïve B cells; Dendritic cells activated, activated dendritic cells; Dendritic cells resting, resting dendritic cells; Macrophages M0, unpolarized macrophages; Macrophages M1, classically activated macrophages; Macrophages M2, alternatively activated macrophages; NK cells, natural killer cells; T cells CD4 memory activated, activated CD4⁺ memory T cells; T cells CD4 memory resting, resting CD4⁺ memory T cells; T cells CD8, CD8⁺ T cells; T cells follicular helper, follicular helper T cells; T cells gamma delta, γδ T cells; T cells regulatory (Tregs), regulatory. T cells.

Published by Tehran University of Medical Sciences (http://ijaai.tums.ac.ir)

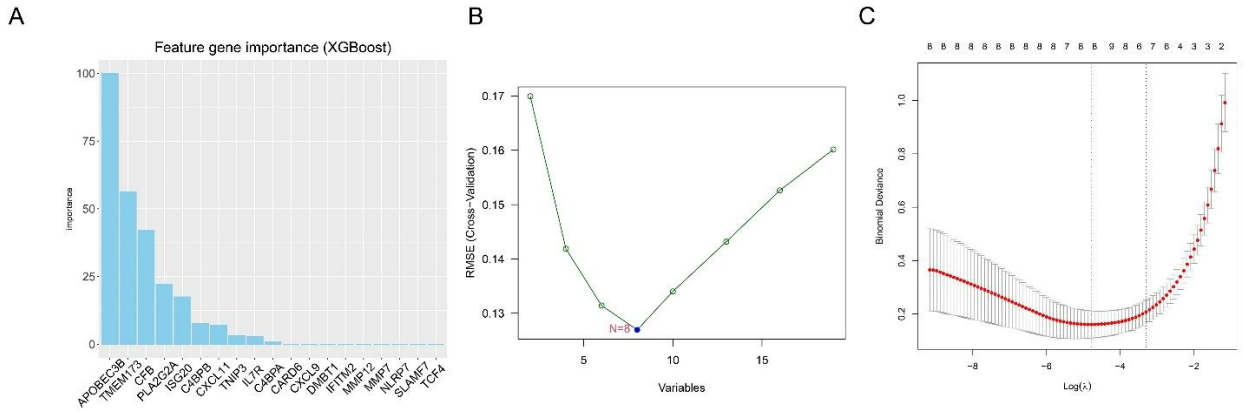
Feature Genes Screened by Machine Learning Models

We established four machine learning methods to screen the optimal variables from the 19 immune microenvironment-related DEGs. The importance of the 19 genes in the XGBoost model was displayed in Figure 5A, of which 10 genes were considered important with a non-zero value. Similarly, 8 optimal variables were identified by the SVM-RFE algorithm (Figure 5B). As to the LASSO regression model, the optimal penalty parameter was determined via the minimum criteria, and 8 non-zero coefficients were identified (Figure 5C–D). In addition, 9 variables were obtained with the value of importance more than 1 in the Random Forest model (Figure 5E–F). Eventually, we identified 4 overlapping feature genes: *APOBEC3B*, *CXCL11*, *PLA2G2A,* and *TMEM173* (Figure 5G).

Validation of Feature Genes

In order to acquire a credible result, we examined the expression levels of feature genes between UC and healthy controls in both the discovery cohort and validation cohort. ROC curves were also calculated, and the area under the ROC curve (AUC) was used to evaluate the diagnostic values of these 4 genes in UC patients in these 3 cohorts. The data showed that the tissue mRNA expressing levels of *APOBEC3B* were significantly lower in UC patients than those in healthy samples, while the other 3 genes, including *CXCL11*, *PLA2G2A,* and *TMEM173*, were all up-regulated both in the discovery cohort (Figure 5H) and validation cohorts (Figure 6A–B). The AUC of these 4 feature genes was all more than 0.9 in the discovery cohort (Figure 5I), and the minimum of the AUC in the other two validation datasets was 0.772 (Figure 6C–D).

Furthermore, we used our clinical results to validate the reliability of the above data. The method of qRT-PCR was employed to determine the mRNA expressions of these 4 feature genes in our collected colonic tissues between UC patients and healthy individuals. Fortunately, our PCR data also sd the similar results that the tissue *APOBEC3B* levels in UC patients were significantly down-regulated, while *PLA2G2A*, *CXCL11,* and *TMEM173* expressions were up-regulated, compared to healthy individuals (Figure 6E). More importantly, the results of these 4 feature genes' ROC curves were also satisfactory for diagnosis, and the AUCs were 0.859, 0.903, 0.883, and 0.803, respectively (Figure 6F). In addition, according to the median of expressing levels, we divided these UC clinical samples into high- and low-expression groups, and the relationships between gene expressions and clinical features were calculated. The data showed that these 4 feature genes were significantly associated with some UC-related clinicopathological factors, such as the gender (*APOBEC3B*, p=0.046), smoking (*PLA2G2A*, p=0.028), stress (*APOBEC3B*, p=0.027), and Mayo score (*APOBEC3B*, p=0.008; *CXCL11*, p=0.038; *TMEM173*, p=0.030) (Table 2).



Published by Tehran University of Medical Sciences (http://ijaai.tums.ac.ir)

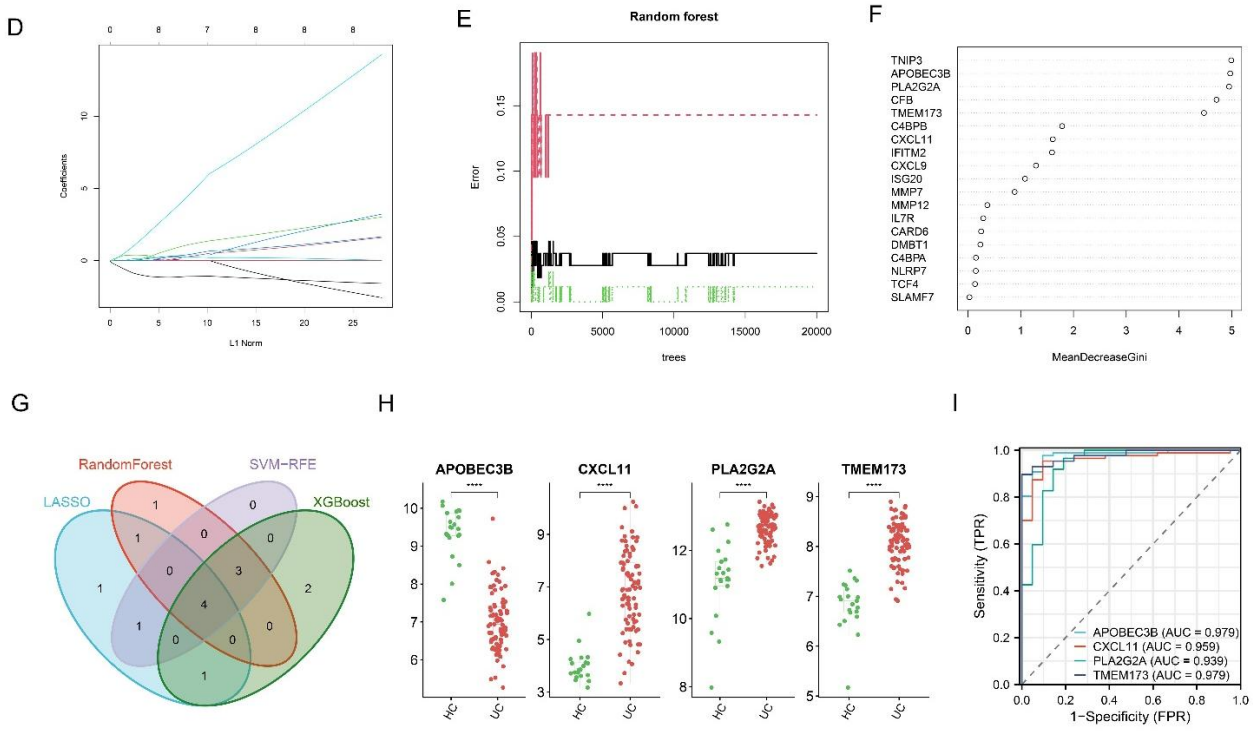
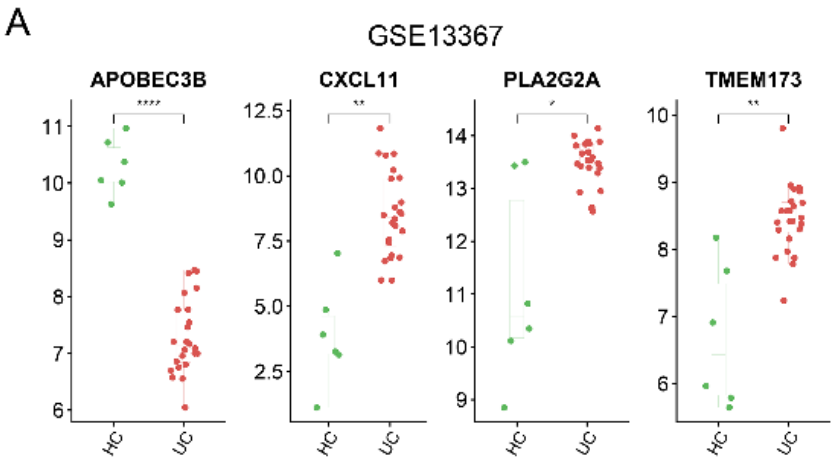


Figure 5. Feature genes screened by multiple machine learning models and their expression validation in the discovery dataset. A–F. Important coefficients obtained from extreme gradient boosting (XGBoost) algorithms, support vector machine–recursive feature elimination (SVM-RFE) algorithms, least absolute shrinkage and selection operator (LASSO) regression, and Random Forest (RF) algorithms. G. Venn plot showed 4 feature genes among XGBoost, SVM-RFE, LASSO, and Random Forest algorithm. H. The tissue messenger RNA (mRNA) expressing levels of these 4 feature genes in the discovery cohort. I. The area under the receiver operating characteristic curve (AUC) of these 4 feature genes was calculated in the discovery cohort. (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001, and ns, no significance). Abbreviations: AUC: area under the receiver operating characteristic curve; HC: healthy controls; RF: Random Forest; SVM-RFE: support vector machine–recursive feature elimination; UC: ulcerative colitis; XGBoost: extreme gradient boosting



Published by Tehran University of Medical Sciences (http://ijaai.tums.ac.ir)

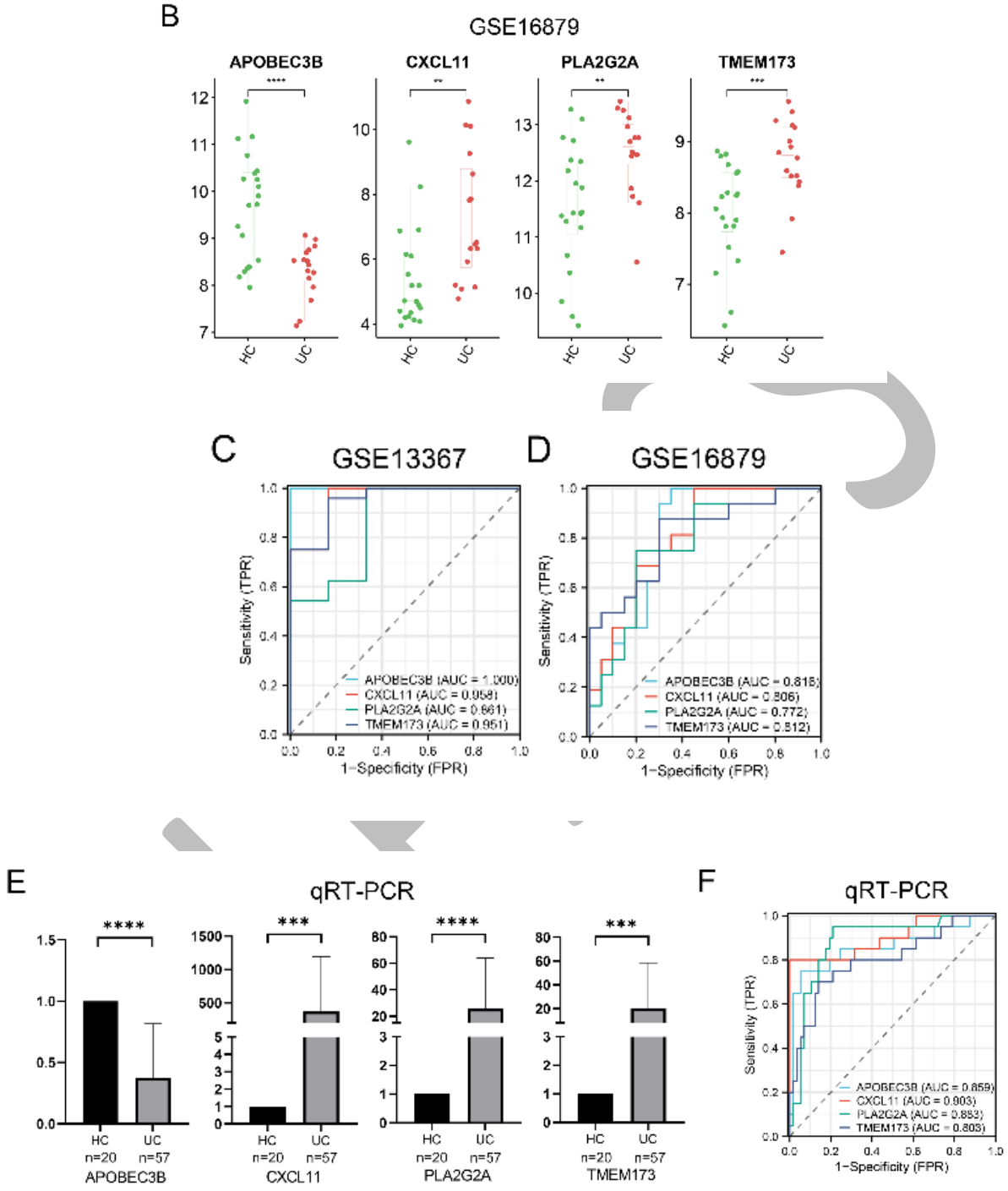


Figure 6. Validation of 4 feature genes in the two validation cohorts (GSE13367 and GSE16879) and our collected samples. A and B. The tissue messenger RNA (mRNA) expressing levels and C and D. Area under the receiver operating characteristic curve (AUC) of these 4 feature genes in the discovery cohorts. E. These 4 feature genes were also determined in our collected colonic tissues between ulcerative colitis (UC) patients and healthy individuals. F. The AUC of these 4 feature genes was also calculated in our collected samples. (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001, and ns, no significance). Abbreviations: AUC: area under the receiver operating characteristic curve; HC: healthy controls; mRNA: messenger RNA; UC: ulcerative colitis.

Published by Tehran University of Medical Sciences (http://ijaai.tums.ac.ir)

Table 2. Correlations between feature gene expressions and clinicopathological factors.

Feature genes	Groups	Gender	Median age (years)	High-fat diet	High-sugar diet	Drinking milk	Smoking (male)	Stress	Family history	Mayo score
APOBEC3B	High expression(n=29)	11	48	17	12	10	9	13	1	6.48
	Low expression(n=28)	19	38	20	11	2	6	18	4	7.78
	p	0.046*	0.609	0.462	1	0.027	0.465	0.027*	0.328	0.008*
CXCL11	High expression(n=29)	17	39	19	8	8	5	19	4	7.62
	Low expression(n=28)	13	49	18	15	4	10	12	1	6.61
	p	0.512	0.411	1	0.084	0.365	0.144	0.147	0.371	0.038*
PLA2G2A	High expression(n=29)	19	39	20	9	9	4	17	4	7.41
	Low expression(n=28)	11	48.5	17	14	3	11	14	1	6.82
	p	0.086	0.380	0.708	0.234	0.120	0.028*	0.699	0.371	0.146
TMEM173	High expression(n=29)	15	39	19	11	8	8	17	5	7.65
	Low expression(n=28)	15	47.5	18	12	4	7	14	0	6.57
	p	1	0.638	1	0.913	0.120	1	0.699	0.067	0.030*

*p<0.05. All these data strongly implied that these 4 genes could be considered as UC-associated feature molecules for our subsequent study.

AUC: area under the receiver operating characteristic curve; *APOBEC3B*: apolipoprotein B mRNA editing enzyme catalytic subunit 3B; CDF: cumulative distribution function; cMAP: connectivity map; *CXCL11*: C-X-C motif chemokine ligand 11; FDR: false discovery rate; GO: Gene Ontology; GSVA: gene set variation analysis; HC: healthy controls; KEGG: Kyoto Encyclopedia of Genes and Genomes; LASSO: least absolute shrinkage and selection operator; MEK: mitogen-activated protein kinase kinase; mRNA: messenger RNA; PAC: proportion of ambiguous clustering; PPI: protein–protein interaction; qRT-PCR: quantitative reverse transcription polymerase chain reaction; RF: Random Forest; ssGSEA: single-sample gene set enrichment analysis; SVM-RFE: support vector machine–recursive feature elimination; *PLA2G2A*: phospholipase A2 group IIA; *TMEM173*: transmembrane protein 173; TNF: tumor necrosis factor; UC: ulcerative colitis; XGBoost: extreme gradient boosting.

Recognition of Immune Microenvironment-related Subtypes in UC Patients

We grouped the UC samples in the discovery cohort using the proportion of ambiguous clustering (PAC) measure based on the expression data of 4 feature genes. K=2 was selected as the optimal value based on the consensus matrix, CDF plot, relative change of the CDF curve area, the cluster-consensus score, and the PAC value, for getting a minimum of 0.071 when we attempted to figure out the PAC value based on k values ranging from 2 to 9 (Figure 7A–C). Finally, we identified 2 subtypes (subtype1 and subtype2, respectively) in UC patients according to these 4 feature genes, and the significant expression differences of 4 feature genes between 2 subtypes were also confirmed, except for *APOBEC3B* (Figure 7D).

Immune Microenvironment Analysis and Functional Enrichment between two subtypes

We estimated the scores of 28 kinds of immune cells between the two subtypes using the ssGSEA method, in which a majority of immune cells represented a significant difference between the two subtypes (Figure 7E). After that, we performed GSVA analysis aiming to explore the different characteristics between these two subtypes. The evaluated scores of the categories for UC samples were displayed as heatmaps (Supplementary Figure 3A–B). As displayed in Figure 8A, the biological functions of subtype 2 were mainly involved in immune response: Defense response to symbiont, Negative regulation of innate immune response, Cytolytic granule, AIM2 inflammasome complex, CXCR3 chemokine receptor binding, and MHC class IB protein binding. Instead, Branching involved in prostate gland morphogenesis, Cellular response to acidic PH, Cellular water homeostasis, Corpus callosum development, Ciliary membrane, Egg coat, Polymerase II CTD heptapeptide repeat phosphatase activity, and potassium channel inhibitor activity were highly enriched in subtype 1. The subsequent KEGG pathway enrichment showed a similar trend that subtype2 closely correlated with Human Diseases (Circulatory system) and Organismal Systems (immune system), while subtype1 was connected with Metabolism (Figure 8B). These data demonstrated that subtype 2 might present an immune-related subtype, while subtype 1 was only proposed as a metabolism-related subtype.

Prediction of Small-molecule Compounds Targeting these two Subtypes of UC Patients

We uploaded the DEGs of two subtypes compared with healthy controls. Different small-molecule compounds were predicted and evaluated when targeting both subtypes. As shown in Figure 8C, the MEK inhibitors (PD-184352 and selumetinib) achieved the top scores on both subtype 1 and subtype 2 among thousands of small-molecule compounds, which might imply a promising therapeutic modality for UC patients (Figure 8C). We also compared other compounds that proved efficient or novel efficient for UC patients between subtype 1 and subtype 2, including the JAK inhibitors, TNF production inhibitors, glucocorticoid receptor (GR) agonists, NF-κB pathway inhibitors, p38 MAPK inhibitors, AKT inhibitors, and PI3K inhibitors. These predicted results provided us with novel and useful clues that different subtypes of UC patients showed different or contradictory responses to these included small-molecule compounds (Figure 8D). This data may help us make a more reasonable selection when these drugs are inevitably needed for UC treatments in the clinic.

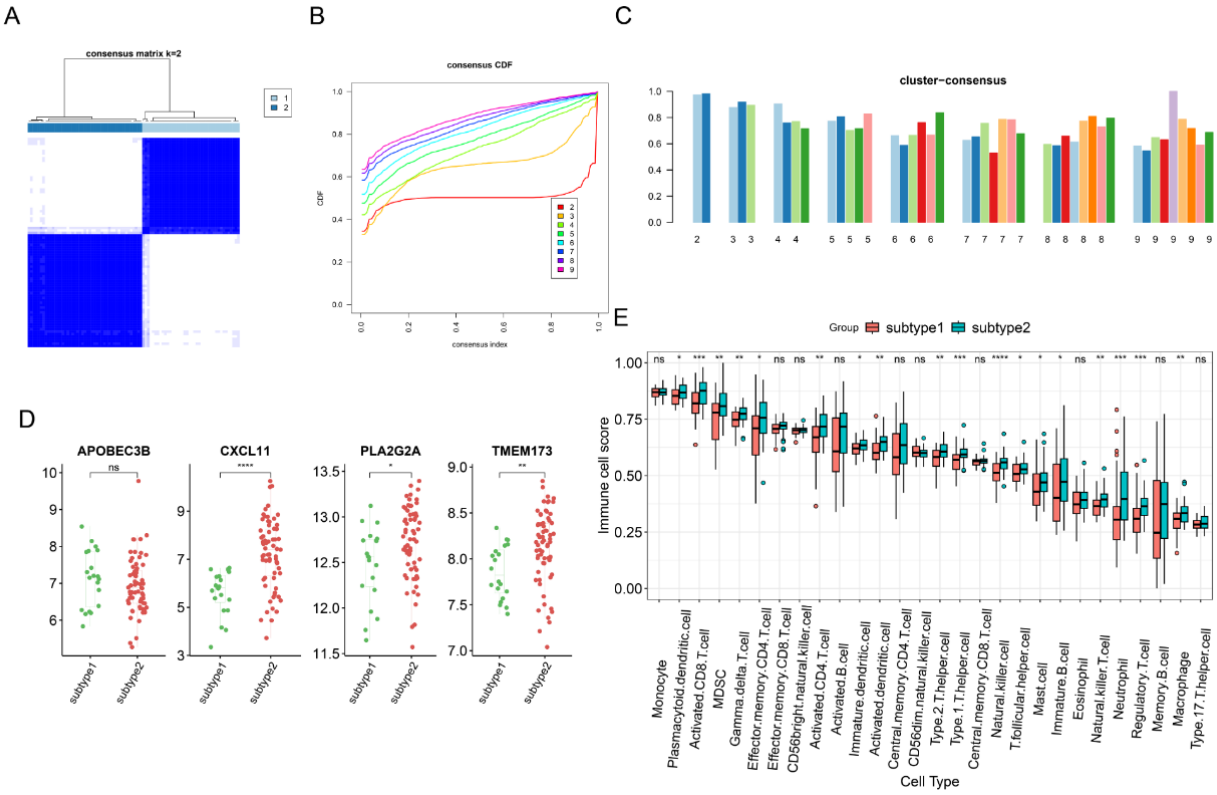


Figure 7. Recognition of immune microenvironment-related subtypes in ulcerative colitis (UC) patients and differences of immune cell infiltration between two subtypes by proportion of ambiguous clustering (PAC) measure and single-sample gene set enrichment analysis (ssGSEA). A. Consensus clustering matrix when k = 2. B. Consensus cumulative distribution function (CDF) curves when k = 2 to 9. C. Consensus score of each subtype when k = 2 to 6. D. The expression levels of 4 feature genes between subtype 1 and subtype 2. E. Box plot showed the differences of infiltrated immune cells in the discovery dataset between UC and normal individuals (The false discovery rate (FDR) q<0.05, *q<0.01, **q<0.001, *q<0.0001, and ns, no significance).**

Published by Tehran University of Medical Sciences (http://ijaai.tums.ac.ir)

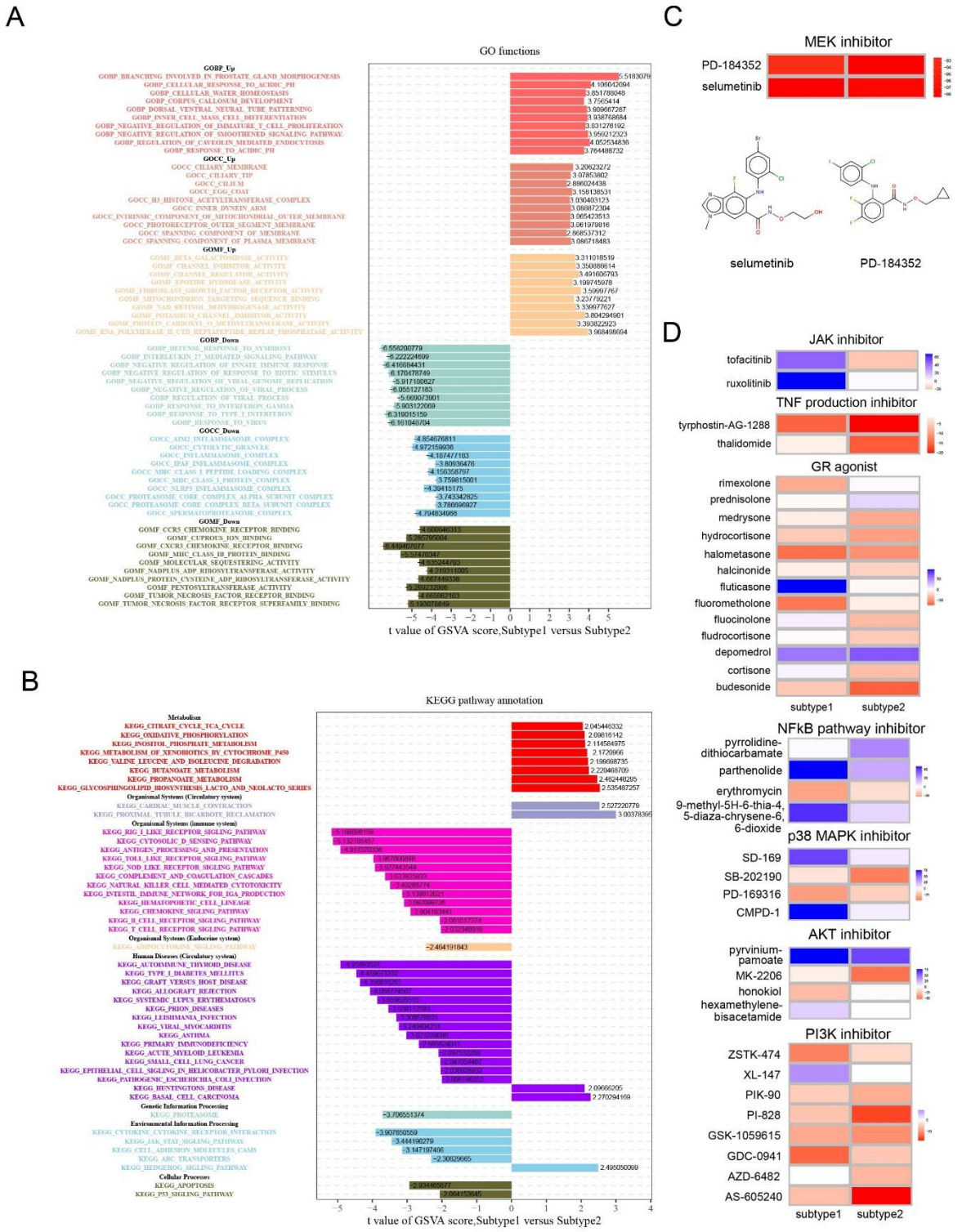


Figure 8. Enrichment analysis and connectivity map (cMAP) analysis between two subtypes. A and B. Differences in enriched biological functions and leading Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways between subtypes ranked by t values of gene set variation analysis (GSVA) scores. C. Potential compounds (mitogen-activated protein kinase kinase (MEK) inhibitors) with cMAP scores less than −90 targeting both subtype 1 and subtype 2. D. cMAP scores of small-molecule compounds targeting related pathways proved theoretically effective on ulcerative colitis (UC) therapy.

DISCUSSION

New data indicate that the incidence and prevalence of UC have significantly increased over the past decades, and it correspondingly places an increasing and inescapable global medical burden on health and social systems.[27,28] In order to accurately diagnose UC in the early stage, clinical symptoms, histological analysis, and combined with different modern medical tests, including endoscopy, gastrointestinal radiography, and laboratory examination, are widely used. However, these routine tests still have some inevitable limitations.[29,30] It has become an urgent clinical need to seek more effective biomarkers to improve UC diagnosis and enable individualizing risk stratification regarding monitoring, treatment, and follow-up.[31] For this reason, investigators have uncovered many newly high-confidence molecules in colon tissues, plasma, stools and rectal mucus.[32] For instance, in 2014, Van der Goten et al collected the colonic mucosal biopsies from 17 UC patients and 10 normal individuals, and used microarrays to identify a number of miRNAs and mRNAs that are differentially expressed between active UC and controls.[33] Zhou et al also systematically summarized the recent usefulness of miRNA for diagnosis, monitoring, and prognosis of UC.[34] In 2022, our previous data also firstly implied that human-derived circular RNAs could be novel biomarkers that distinguished patients with UC as well as CAC with normal controls.[35] In this study, we aimed to search for new biomarkers from the perspective of dysregulated immunity in UC. Firstly, we analyzed the DEGs between UC and healthy individuals in the discovery dataset. Meanwhile, we evaluated the 22 types of immune infiltrating cells in UC and healthy samples, and finally identified 10 variables, including macrophages M1, macrophages M0, mast cells activated, dendritic cells activated, T cells CD8, T cells CD4 memory resting, plasma cells, mast cells resting, NK cells activated, and T cells regulatory (Tregs), as characteristic immune cells in the UC progression, which had a non-zero coefficient in LASSO regression analysis. Subsequently, using the WGCNA, the dark turquoise module, along with the magenta module, was found to closely associate with 10 types of characteristic UC-related immune cells. This meant that these two modules possibly included some key immune-related genes. The Venn analysis (DEGs, genes from WGCNA modules, and InnateDB database) initially identified 19 key genes as immune microenvironment-related DEGs. Machine learning is an interdisciplinary and machine learning algorithms enable machines to learn from massive data, recognize patterns, and make decisions without being explicitly programmed. Machine learning algorithms were applied for predicting diagnostic signatures in multiple diseases, including UC.[36] Finally, four machine learning methods, including the XGBoost model, SVM-RFE algorithm, LASSO regression model, and Random Forest model, successfully helped us identify 4 overlapping feature genes: *APOBEC3B*, *CXCL11*, *PLA2G2A,* and *TMEM173*. The expression levels of these 4 feature genes were significantly up- or down-regulated in UC samples, which could also serve as diagnostic biomarkers in UC patients.

Over the last centuries and decades, the classification of UC was primarily on the basis of the clinical presentations, such as the age of patients, location, and lesion extensions, and duration or stage of the disease. However, such classifications have not proven informative or satisfactory for predicting UC progression and response to treatments.[37-39] Therapeutic decisions are mainly based on subjective evaluations of disease severity. Available evidence suggests that abnormal immune responses against the microorganisms of the intestinal flora have classically been considered to play a pivotal role during the pathogenesis of UC.[6,40] Based on our newly identified immune-related feature genes in UC, a bold assumption was that specific molecular UC classifications might exist.[41] Using the PAC measure based on the expression data of 4 feature genes, two stable subtypes (subtype1 and subtype2) were successfully identified. Subsequent ssGSEA analysis revealed that subtype 2 had higher scores for the majority of immune cells than subtype 1. To confirm whether subtype 2 had a more severe inflammatory response than subtype 1, we performed GSVA enrichment analysis for biological functions. The results showed that subtype 2 was closely associated with immune system and circulatory system diseases in KEGG analysis. The proven UC-related signaling pathways, such as the Toll-like receptor signaling pathway, NOD-like receptor signaling pathway, Natural killer cell-mediated cytotoxicity, Intestinal immune network for IgA production, Chemokine signaling pathway, B cell receptor signaling pathway, and T cell receptor signaling pathway, were significantly enriched in subtype 2. These results demonstrated that subtype 2 of UC patients presented a worse inflammatory response and needed to recruit excessive immune cells to participate in the anti-inflammatory process.

Over the past two decades, treatment for UC has dramatically improved. These anti-inflammatory agents included 5-ASA compounds, systemic and topical corticosteroids, and immunomodulators. In addition, a number of biologicals and small molecules have been explored and applied, which provide us with more treatment options for personalized medicine.[42-45] Now that we had immunologically distinguished UC patients as subtype 1 and subtype 2, it was believed that such molecular classifications might affect the sensitivities of different anti-UC medications. Then we assessed scores of partial novel drugs querying the cMAP database, and found that the efficacy of different small-molecule compounds presented different results across subtypes. For instance, the MEK inhibitors (PD-184352 and selumetinib) achieved the top scores both in subtype 1 and subtype 2, compared to thousands of other small-molecule compounds. MEK inhibitors are promising therapeutic agents for UC, but they are merely inferred from the data and require subsequent experimental and clinical validation. More interestingly, even when including the same sub-classification of drugs, the small-molecule compounds presented completely different responses. Previous data have proved that the p38 MAPK signal transduction pathway plays an essential role in the pathogenesis of UC. Blockade of p38 MAPK signals was found to ameliorate inflammation, at least in part, by reducing secretions of some pro-inflammatory cytokines, suggesting that the p38 MAPK signaling pathway might be considered as a new target for UC treatment.[46-48] However, in our data, p38 MAPK inhibitors, including SD-169, SB-202190, PD-169316 and CMPD-1, showed different or contradictory responses between subtype1 and subtype2 of UC patients. Similar phenomena were also shown in other types of medications, such as the JAK inhibitors, TNF production inhibitors, GR agonists, NF-κB pathway inhibitors, AKT inhibitors, and PI3K inhibitors. In our previous study, *SLC26A2* was negatively correlated with the IL-17 signaling pathway and positively associated with the tight junction, which led to abnormal immune cell infiltration and inflammatory injuries. According to the *SLC26A2* expression, UC patients were divided into different subgroups. The potential target drugs for UC treatment, such as progesterone, tetradioxin, and dexamethasone, were initially predicted and exert anti-inflammatory effects via the common molecule SLC26A2.49.

While we had previously emphasized the importance of these small molecules, it was important to note that these findings were limited to database predictions and still require further experimental validation. And there are some limitations in our study. It identified that the ROC of four genes (*APOBEC3B*, *CXCL11*, *PLA2G2A*, and *TMEM173*) ranged from 0.772 to 0.9, while the clinical relevance was not fully developed and lacked protein-level validation. These signals are hypothesis-generating, and we plan to conduct further experimental validation, including patient-derived organoids or peripheral immune cell assays, including patient-derived organoids or peripheral immune cell assays. We focused on the clinical pathological factors of UC, with an emphasis on the etiological factors of UC. In subsequent studies, we will investigate whether these molecular subtypes are consistent with the endoscopic or histological criteria for UC or not. To translate these genes into useful diagnostic assays in real-world practice, several steps are necessary. First, large-scale validation studies should be conducted to confirm the association of these genes with clinical outcomes across diverse patient populations. Second, sensitive and specific detection methods need to be developed to accurately measure the expression levels of these genes in clinical samples. Third, these gene expression data should be integrated with other clinical parameters to build machine learning models that can predict treatment responses in ulcerative colitis patients. However, this process requires a substantial amount of sequencing data on treatment outcomes to ensure the model's accuracy and reliability. Therefore, it is recommended to gather and analyze more comprehensive patient data before proceeding with model development.

In conclusion, this study firstly used bioinformatic predictions and experimental verification to identify 4 immune-related feature genes: *APOBEC3B*, *CXCL11*, *PLA2G2A,* and *TMEM173*. These 4 genes were significantly up- or down-regulated in UC samples and could serve as satisfactory diagnostic predictors for UC patients. Then PAC measure of these 4 feature genes successfully classified UC patients into two molecular subtypes (subtype1 and subtype2), and the ssGSEA algorithm revealed that subtype2, with a higher score of the majority of immune cells, presented a worse inflammatory response. Moreover, we assessed scores of partial novel drugs querying the cMAP database, and found that the efficacy of different small-molecule compounds presented different results across subtypes. These findings will further our molecular understanding

of UC subtypes with a heterogeneous immune pattern and different responses to small-molecule compounds, which may provide us more new ideas and directions to investigate UC patients.

STATEMENT OF ETHICS

This study was approved by the ethics committee of The First Affiliated Hospital of Soochow University (2022-431). Written informed consent was obtained from all participants before recruitment.

FUNDING

This study was supported by grants from the Suzhou Medical Youth Talents Project (Qngg2024004), the National Natural Science Foundation of China (82573771), the Suzhou Gusu Health Talent Research Project (GSWS2023039), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (24KJA320005), and Horizontal Cooperation Project of Soochow University (H230753 and H241466).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ACKNOWLEDGMENTS

We thank Yuting Kuang and Daiwei Wang for providing us with the network data sets.

DATA AVAILABILITY

The datasets (No. GSE87466, GSE16879, and GSE13367) generated and/or analysed during the current study are available in the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/).

AI ASSISTANCE DISCLOSURE

ChatGPT was used in the language improvement of this manuscript.

REFERENCES

1. Yu YR, Rodriguez JR. Clinical presentation of Crohn's, ulcerative colitis, and indeterminate colitis: Symptoms, extraintestinal manifestations, and disease phenotypes. Semin Pediatr Surg. 2017;26(6):349-55.
2. Gu Y, Zhao H, Zheng L, Zhou C, Han Y, Wu A, et al. Non-coding RNAs and colitis-associated cancer: Mechanisms and clinical applications. Clin Transl Med. 2023;13(5):e1253.
3. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. Lancet. 2017;389(10080):1756-70.
4. Porter RJ, Kalla R, Ho GT. Ulcerative colitis: Recent advances in the understanding of disease pathogenesis. F1000Res. 2020;9:
5. Park JH, Peyrin-Biroulet L, Eisenhut M, Shin JI. IBD immunopathogenesis: A comprehensive review of inflammatory molecules. Autoimmun Rev. 2017;16(4):416-26.
6. Tatiya-Aphiradee N, Chatuphonprasert W, Jarukamjorn K. Immune response and inflammatory pathway of ulcerative colitis. J Basic Clin Physiol Pharmacol. 2018;30(1):1-10.
7. Camilleri M, Madsen K, Spiller R, Greenwood-Van MB, Verne GN. Intestinal barrier function in health and gastrointestinal disease. Neurogastroent Motil. 2012;24(6):503-12.
8. Minton K. Intestinal barrier protection. Nat Rev Immunol. 2022;22(3):144-5.
9. Peterson LW, Artis D. Intestinal epithelial cells: regulators of barrier function and immune homeostasis. Nat Rev Immunol. 2014;14(3):141-53.
10. Maldonado-Contreras AL, McCormick BA. Intestinal epithelial cells and their role in innate mucosal immunity. Cell Tissue Res. 2011;343(1):5-12.
11. Geremia A, Biancheri P, Allan P, Corazza GR, Di Sabatino A. Innate and adaptive immunity in inflammatory bowel disease. Autoimmun Rev. 2014;13(1):3-10.
12. Kmiec Z, Cyman M, Slebioda TJ. Cells of the innate and adaptive immunity and their interactions in inflammatory bowel disease. Adv Med Sci-Poland. 2017;62(1):1-16.
13. Hanzel J, Hulshoff MS, Grootjans J, D'Haens G. Emerging therapies for ulcerative colitis. Expert Rev Clin Immu. 2022;18(5):513-24.
14. Ashton JJ, Green Z, Kolimarala V, Beattie RM. Inflammatory bowel disease: long-term therapeutic challenges. Expert Rev Gastroent. 2019;13(11):1049-63.
15. Tran V, Limketkai BN, Sauk JS. IBD in the Elderly: Management Challenges and Therapeutic Considerations. Curr Gastroenterol Rep. 2019;21(11):60.
16. Ferretti F, Cannatelli R, Maconi G, Ardizzone S. Therapeutic Management of Adults with Inflammatory

Bowel Disease and Malignancies: A Clinical Challenge. Cancers. 2023;15(2):
17. 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(7):e47.
18. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453-7.
19. Engebretsen S, Bohlin J. Statistical predictions with glmnet. Clin Epigenetics. 2019;11(1):123.
20. Gui J, Li H. Penalized Cox regression analysis in the high-dimensional and low-sample-size settings, with applications to microarray gene expression data. Bioinformatics. 2005;21(13):3001-8.
21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
22. Kim S. Margin-maximised redundancy-minimised SVM-RFE for diagnostic classification of mammograms. Int J Data Min Bioin. 2014;10(4):374-90.
23. Li W, Yin Y, Quan X, Zhang H. Gene Expression Value Prediction Based on XGBoost Algorithm. Front Genet. 2019;10:1077.
24. Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. Brief Bioinform. 2019;20(2):492-503.
25. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 2017;18(1):248-62.
26. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.
27. Le Berre C, Ananthakrishnan AN, Danese S, Singh S, Peyrin-Biroulet L. Ulcerative Colitis and Crohn's Disease Have Similar Burden and Goals for Treatment. Clin Gastroenterol H. 2020;18(1):14-23.
28. Wei SC, Sollano J, Hui YT, Yu W, Santos EP, Llamado L, et al. Epidemiology, burden of disease, and unmet needs in the treatment of ulcerative colitis in Asia. Expert Rev Gastroent. 2021;15(3):275-89.
29. Seyedian SS, Nokhostin F, Malamir MD. A review of the diagnosis, prevention, and treatment methods of inflammatory bowel disease. J Med Life. 2019;12(2):113-22.
30. Kaenkumchorn T, Wahbeh G. Ulcerative Colitis: Making the Diagnosis. Gastroenterol Clin N. 2020;49(4):655-69.
31. Sakurai T, Saruta M. Positioning and Usefulness of Biomarkers in Inflammatory Bowel Disease. Digestion. 2023;104(1):30-41.
32. Liu D, Saikam V, Skrada KA, Merlin D, Iyer SS. Inflammatory bowel disease biomarkers. Med Res Rev. 2022;42(5):1856-87.
33. Van der Goten J, Vanhove W, Lemaire K, Van Lommel L, Machiels K, Wollants WJ, et al. Integrated miRNA and mRNA expression profiling in inflamed colon of patients with ulcerative colitis. Plos One. 2014;9(12):e116117.
34. Zhou J, Liu J, Gao Y, Shen L, Li S, Chen S. miRNA-Based Potential Biomarkers and New Molecular Insights in Ulcerative Colitis. Front Pharmacol. 2021;12:707776.
35. Wan D, Wang S, Xu Z, Zan X, Liu F, Han Y, et al. PRKAR2A-derived circular RNAs promote the malignant transformation of colitis and distinguish patients with colitis-associated colorectal cancer. Clin Transl Med. 2022;12(2):e683.
36. Lu J, Wang Z, Maimaiti M, Hui W, Abudourexiti A, Gao F. Identification of diagnostic signatures in ulcerative colitis patients via bioinformatic analysis integrated with machine learning. Hum Cell. 2022;35(1):179-88.
37. Mohammed VN, Samaan M, Mosli MH, Parker CE, MacDonald JK, Nelson SA, et al. Endoscopic scoring indices for evaluation of disease activity in ulcerative colitis. Cochrane Db Syst Rev. 2018;1(1):CD011450.
38. Pai RK, Jairath V, Vande CN, Rieder F, Parker CE, Lauwers GY. The emerging role of histologic disease activity assessment in ulcerative colitis. Gastrointest Endosc. 2018;88(6):887-98.
39. Pabla BS, Schwartz DA. Assessing Severity of Disease in Patients with Ulcerative Colitis. Gastroenterol Clin N. 2020;49(4):671-88.
40. Ahluwalia B, Moraes L, Magnusson MK, Ohman L. Immunopathogenesis of inflammatory bowel disease and mechanisms of biological therapies. Scand J Gastroentero. 2018;53(4):379-89.
41. Furey TS, Sethupathy P, Sheikh SZ. Redefining the IBDs using genome-scale molecular phenotyping. Nat Rev Gastro Hepat. 2019;16(5):296-311.
42. Bhattacharya A, Osterman MT. Biologic Therapy for Ulcerative Colitis. Gastroenterol Clin N. 2020;49(4):717-29.
43. Caballol B, Gudino V, Panes J, Salas A. Ulcerative colitis: shedding light on emerging agents and strategies in preclinical and early clinical development. Expert Opin Inv Drug. 2021;30(9):931-46.
44. Lasa JS, Olivera PA, Danese S, Peyrin-Biroulet L. Efficacy and safety of biologics and small molecule drugs for patients with moderate-to-severe ulcerative colitis: a

systematic review and network meta-analysis. Lancet Gastroenterol. 2022;7(2):161-70.
45. Shaaban AA, Abdelhamid AM, Shaker ME, Cavalu S, Maghiar AM, Alsayegh AA, et al. Combining the HSP90 inhibitor TAS-116 with metformin effectively degrades the NLRP3 and attenuates inflammasome activation in rats: A new management paradigm for ulcerative colitis. Biomed Pharmacother. 2022;153:113247.
46. Feng YJ, Li YY. The role of p38 mitogen-activated protein kinase in the pathogenesis of inflammatory bowel disease. J Digest Dis. 2011;12(5):327-32.
47. Elkholy SE, Maher SA, Abd EN, Elsayed HA, Hassan WA, Abdelmaogood A, et al. The immunomodulatory effects of probiotics and azithromycin in dextran sodium sulfate-induced ulcerative colitis in rats via TLR4-NF-kappaB and p38-MAPK pathway. Biomed Pharmacother. 2023;165:115005.
48. Zhou A, Zhang S, Yang C, Liao N, Zhang Y. Dandelion root extracts abolish MAPK pathways to ameliorate experimental mouse ulcerative colitis. Adv Clin Exp Med. 2022;31(5):529-38.
49. Qian L, Hu S, Zhao H, Han Y, Dai C, Zan X, et al. The Diagnostic Significance of SLC26A2 and Its Potential Role in Ulcerative Colitis. Biomedicines. 2025;13(2):461.