



OPEN ACCESS

EDITED BY

Edoardo Rosato,
Sapienza University of Rome, Italy

REVIEWED BY

Antonietta Gigante,
Sapienza University of Rome, Italy
Chiara Pellicano,
Sapienza University of Rome, Italy

*CORRESPONDENCE

Yuting Liang
✉ liangyuting666@126.com
Chao Huang
✉ zjhc2007@126.com
Xianping Zhou
✉ supershan99@163.com

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 02 May 2024

ACCEPTED 03 July 2024

PUBLISHED 07 August 2024

CITATION

Feng B, Zhang Y, Qiao L, Tang Q, Zhang Z, Zhang S, Qiu J, Zhou X, Huang C and Liang Y (2024) Evaluating the significance of ECSCR in the diagnosis of ulcerative colitis and drug efficacy assessment. *Front. Immunol.* 15:1426875. doi: 10.3389/fimmu.2024.1426875

COPYRIGHT

© 2024 Feng, Zhang, Qiao, Tang, Zhang, Zhang, Qiu, Zhou, Huang and Liang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Evaluating the significance of ECSCR in the diagnosis of ulcerative colitis and drug efficacy assessment

Bin Feng^{1†}, Yanqiu Zhang^{2†}, Longwei Qiao^{3,4†}, Qingqin Tang¹, Zheng Zhang¹, Sheng Zhang¹, Jun Qiu¹, Xianping Zhou^{5,6*}, Chao Huang^{3*} and Yuting Liang^{1*}

¹Center for Clinical Laboratory, The First Affiliated Hospital of Soochow University, Suzhou, Jiangsu, China, ²Institute of Clinical Pharmacology, Anhui Medical University, Key Laboratory of Anti-inflammatory and Immune Medicine, Ministry of Education, Anhui Collaborative Innovation Center of Anti-inflammatory and Immune Medicine, Hefei, Anhui, China, ³Center for Reproduction and Genetics, School of Gusu, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital, Nanjing Medical University, Suzhou, Jiangsu, China, ⁴State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China, ⁵Department of Laboratory, Bozhou Hospital Affiliated to Anhui Medical University, Bozhou, Anhui, China, ⁶Department of Laboratory, Anhui Medical University, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China

Background: The main challenge in diagnosing and treating ulcerative colitis (UC) has prompted this study to discover useful biomarkers and understand the underlying molecular mechanisms.

Methods: In this study, transcriptomic data from intestinal mucosal biopsies underwent Robust Rank Aggregation (RRA) analysis to identify differential genes. These genes intersected with UC key genes from Weighted Gene Co-expression Network Analysis (WGCNA). Machine learning identified UC signature genes, aiding predictive model development. Validation involved external data for diagnostic, progression, and drug efficacy assessment, along with ELISA testing of clinical serum samples.

Results: RRA integrative analysis identified 251 up-regulated and 211 down-regulated DEGs intersecting with key UC genes in WGCNA, yielding 212 key DEGs. Subsequently, five UC signature biomarkers were identified by machine learning based on the key DEGs—THY1, SLC6A14, ECSCR, FAP, and GPR109B. A logistic regression model incorporating these five genes was constructed. The AUC values for the model set and internal validation data were 0.995 and 0.959, respectively. Mechanistically, activation of the IL-17 signaling pathway, TNF signaling pathway, PI3K-Akt signaling pathway in UC was indicated by KEGG and GSVA analyses, which were positively correlated with the signature biomarkers. Additionally, the expression of the signature biomarkers was strongly correlated with various UC types and drug efficacy in different datasets. Notably, ECSCR was found to be upregulated in UC serum and exhibited a positive correlation with neutrophil levels in UC patients.

Conclusions: THY1, SLC6A14, ECSCR, FAP, and GPR109B can serve as potential biomarkers of UC and are closely related to signaling pathways associated with UC progression. The discovery of these markers provides valuable information for understanding the molecular mechanisms of UC.

KEYWORDS

ulcerative colitis, ECSCR, machine learning, diagnosis, biomarker

Introduction

Ulcerative colitis (UC) is a chronic inflammatory bowel disease (IBD) that commonly affects individuals in their young and middle-aged years. Clinical symptoms encompass abdominal cramps, pus, mucus, and bloody diarrhea (1–3). The global incidence of UC ranges from 10.5 to 14 cases per 100,000 people annually, with a prevalence of approximately 246.7 cases per 100,000 individuals (4). Nevertheless, a consistent increase in UC incidence is observed in Western countries, coupled with significant surges in South America and East Asia. Furthermore, UC onset is occurring at younger ages, with both developed and developing countries experiencing a rise in UC cases among children (5). Given the challenges associated with curing UC, its propensity for recurrence, and the heightened risk of cancer (6), early initiation of treatment for remission and long-term maintenance to prevent recurrence constitute crucial strategies. A comprehensive understanding of the disease's pathogenesis and the identification of new biomarkers may offer insights for early diagnosis and monitoring disease progression, ultimately contributing significantly to the improvement of overall health outcomes.

Endoscopy and tissue biopsy continue to serve as the exclusive methods for confirming a diagnosis of UC and evaluating disease activity and severity, albeit causing significant discomfort for patients with UC (7). Furthermore, the reluctance of patients to undergo this technique can lead to delayed diagnosis (8). The development of UC is due to a complex interaction between the host's genetic background, microbial changes, and environmental factors, resulting in improper and chronic activation of the mucosal immune system (9–13). The identification of biomarkers could play a crucial role in facilitating precise diagnosis and selecting therapeutic approaches for UC patients. Combining C-reactive protein and fecal calreticulin levels may provide some benefits in the ongoing evaluation and monitoring of UC progression; however, their efficacy is limited (14). Therefore, the urgent need to identify distinctive markers of UC is paramount for facilitating early diagnosis, assessing disease progression, uncovering new pathogenic mechanisms, and enabling the prediction and development of therapeutic strategies for UC. Moreover, Aminosalicylates (5-ASA) and corticosteroids are fundamentally used in treating and controlling mild to moderate UC (15, 16).

Immunomodulators and biologic therapies, such as anti-TNF-alpha agents (infliximab, adalimumab, golimumab), are effective in reducing UC inflammation and improving disease prognosis (11). With the increasing incidence of UC, the number of patients with refractory UC—who exhibit poor or no response to conventional medications, prolonged disease duration, and recurrent exacerbations—is also rising (17). Therefore, a thorough investigation of biomarkers associated with therapy response in UC patients is essential, along with the development of new approaches to enhance response rates.

Gene chip assay technology, combined with second-generation sequencing technology and bioinformatics analysis, is currently widely employed for exploring the pathological characteristics of various diseases and identifying potential novel biomarkers, including UC (9, 10, 18). In this study, using transcriptomic data from a substantial number of intestinal mucosal biopsies, differential genes were identified through Robust Rank Aggregation (RRA) analysis and intersected with UC key genes identified by Weighted Gene Co-expression Network Analysis (WGCNA). Subsequently, machine learning algorithms were employed to identify UC signature genes, which were then utilized to develop predictive models. The diagnostic performance of both the modeled genes and models, as well as their relationship with disease progression and drug efficacy, were validated using external data. The diagnostic efficacy of the modeled genes for UC was further confirmed through ELISA testing of clinical serum samples. The objective of this work is to provide new insights into the early diagnosis of UC. The flow chart of this study is illustrated in Figure 1.

Materials and methods

Data collection

The gene expression datasets for UC were retrieved from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/>). Six specific datasets, namely GSE92415, GSE87473, GSE11223, GSE107499, GSE53306, and GSE206285 (Table 1), were selected for inclusion in our study. Subsequently, both gene expression profiles and corresponding clinical information were downloaded for further analysis.

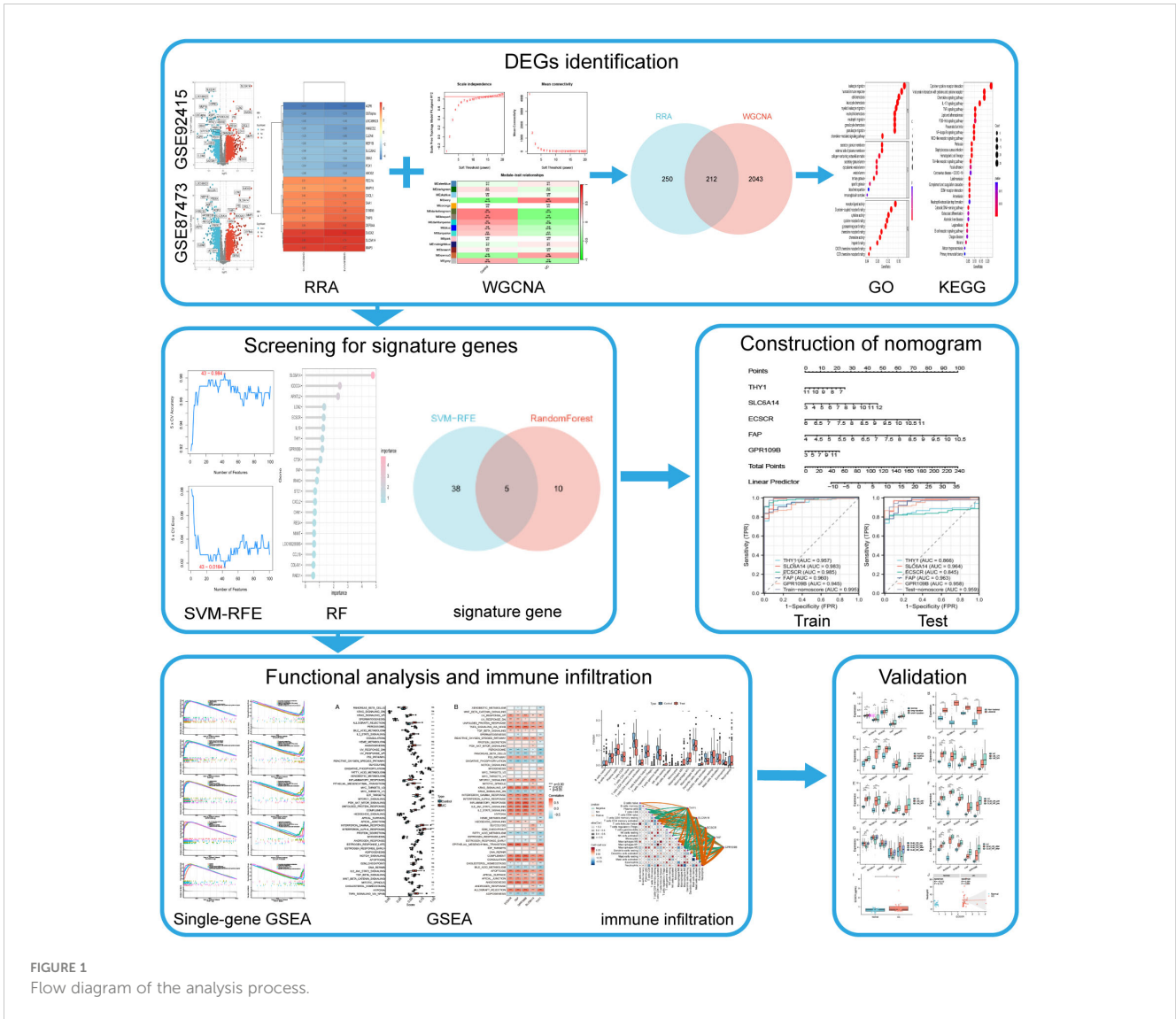


FIGURE 1
Flow diagram of the analysis process.

Identification of differentially expressed genes (DEGs)

The GSE92415 and GSE87473 datasets were selected for differential gene expression analysis. The R package “limma” was used to identify DEGs using $|\log_2\text{fold change (FC)}| > 0.5$ and

adjusted $P < 0.05$ as the cutoff criteria (19–21). The outcomes were visualized using “ggplot2” R packages. Following this, both datasets were amalgamated and subjected to an analysis of common DEGs using the RRA method (22). The lists of up- and down-regulated genes in each dataset were sorted based on logFC. Subsequently, all gene lists were integrated utilizing the “RobustRankAggreg” package.

TABLE 1 Microarray information.

GEO	Platform	Tissue	Samples (number)			Attribute
			Total	UC	Control	
GSE92415	GPL13158	Colon	183	162	21	Training set
GSE87473	GPL13158	Colon	21	106	21	Validation set
GSE11223	GPL1708	Colon	202	129	73	Validation set
GSE107499	GPL15207	Colon	119	75	44	Validation set
GSE53306	GPL14951	Colon	40	28	12	Validation set
GSE206285	GPL13158	Colon	382	364	18	Validation set

Construction of the co-expression network by WGCNA

The R package “WGCNA” was used for gene co-expression network analysis. The WGCNA algorithm is a systems biology approach for characterizing patterns of correlation between genes in different samples (23). Initially, a hierarchical cluster analysis is performed to filter the discrete samples. Subsequently, the optimal soft threshold power (β) was chosen to calculate the adjacency matrix, which was then transformed into a topological overlap matrix (TOM). The identification of different modules was achieved by the dynamic tree-cutting method, specifically filtering modules containing more than 50 genes. Finally, Pearson correlation analysis was used to calculate the correlation between modules and traits. Modules with the most significant correlation coefficients were then selected for further analysis. Intersecting genes identified through both RRA and WGCNA methods were deemed key differential genes for UC.

Functional enrichment analysis of key genes

The R package “org.Hs.eg.db” was used to convert the gene names of the DEGs into gene IDs. The R package “clusterProfiler” was implemented to conduct Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analysis to assess gene-related biological processes (BP), molecular functions (MF), cellular components (CC), and signaling pathways (24, 25). The results were plotted using the R packages “enrichplot” and “ggplot2”.

Screening signature genes by machine learning

To further screen key differential genes, this study employed two machine learning algorithms, support vector machine-recursive feature elimination (SVM-RFE) and Random Forest (RF). SVM-RFE was performed by “caret” package in R at 10-fold cross-validation to determine the variables at the max accuracy (26). RF algorithm was used to rank gene importance using the R package “randomForest” (27). Finally, the intersection of the two machine learning algorithms was identified as signature genes.

Construction and validation of a diagnostic nomogram in UC

The signature genes were used to construct the nomogram in training set (GSE92415) by the R package “rms” (28). The nomoscore based on signature gene expression was calculated by linearly combining coefficients from logistic regression and expression levels. Subsequently, Receiver operating characteristic (ROC) curves were plotted via the R “pROC” package to assess the

diagnostic ability of the nomogram model (29). In addition, the nomogram model was tested in the test set (GSE87473).

$$\text{Nomoscore} = \text{Coef} + \sum_{i=1}^n \text{Expr}_i \times \text{Coef}_i$$

Coef is regression coefficient; Coef_i stands for the coefficient of gene_i; Expr_i is the expression of gene_i.

Gene set enrichment analysis

GSEA was performed to explore the possible function of signature genes using the hallmark gene sets (hall.v7.5.1.symbols), which were obtained from the Molecular Signatures Database (<https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>) (30). R package “GSVA” was applied to compute a hallmark gene set score based on gene expression levels for each sample.

Immune infiltration analysis

To analyze immune infiltration, the CIBERSORT algorithm assessed differential immune cell presence in the colon mucosa of both UC patients and healthy controls (31). Additionally, the Spearman’s correlation analysis was performed to investigate the relationship between immune cells and the identified signature genes (32).

External validation of signature genes and correlation with drug efficacy

GSE11223, GSE107499, GSE53306, and GSE206285 datasets were downloaded for external validation. Differences in expression levels of signature genes in different types of UC and drug efficacy were validated using the Wilcoxon rank sum test, and the results were visualized using the boxplot.

Enzyme-linked immunosorbent assay

Serum samples were collected from 40 UC patients and 25 normal controls from the First Affiliated Hospital of Soochow University. Optical density (OD) was measured at 450 nm using a microplate reader according to the manufacturer’s instructions, and the concentration of ECSCR in the samples was calculated from a standard curve. In this experiment, the detection limit for human ECSCR was 0.1–5 μg/ml. The medical Ethics Committee of the First Affiliated Hospital of Soochow University approved this study.

Statistical analysis

All statistical analysis was performed using R (version 4.3.1) (<https://www.r-project.org>) and associated R packages. A

significance level of $P < 0.05$ was used for all analyses to indicate statistical significance.

Results

Identification of key differential genes and functional enrichment analysis

The GSE92415 and GSE87473 datasets underwent separate differential analyses, leading to the identification of 2761 and 3096 DEGs based on the criteria of $\log_{2}FC > 0.5$ and adjusted $P < 0.05$, respectively (Figures 2A, B). Following this, RRA was employed to integrate the results from the two cohort analyses, revealing 251 up-regulated and 211 down-regulated DEGs. The heatmap illustrates the degree of upregulation or downregulation observed in the top 10 DEGs across both datasets (Figure 2C).

The GSE92415 dataset, in conjunction with the WGCNA software package, was utilized to discern functional clusters linked to UC patients. In the construction of gene co-expression networks, a soft threshold of $\beta=16$ was selected, aligning with a correlation coefficient nearing 0.85 (Figure 3A). Post-merging comparable modules using a MEDissThres of 0.3, a total of 15 modules were generated (Figure 3B). An analysis of module-trait relationships through a heatmap unveiled that the ivory modules exhibited the most robust correlation with UC progression ($r = 0.52$, $P = 4e-14$) (Figure 3C).

A total of 212 genes were identified as differentially expressed in both the RRA and WGCNA analyses, designating them as key differential genes through the application of a Venn diagram (Figure 3D). To thoroughly elucidate the biological processes and pathways associated with these DEGs, we conducted GO and KEGG

enrichment analyses. In the GO analysis, DEGs exhibited significant enrichment in the following processes: leukocyte migration, humoral immune response, cell chemotaxis, leukocyte chemotaxis, secretory granule membrane, external side of the plasma membrane, collagen-containing extracellular matrix, receptor ligand activity, G protein-coupled receptor binding, cytokine activity, and cytokine receptor binding (Figure 3E). For the KEGG analysis, genes displayed notable enrichment in cytokine-cytokine receptor interaction, chemokine signaling pathway, IL-17 signaling pathway, TNF signaling pathway, PI3K-Akt signaling pathway, and rheumatoid arthritis (RA) (Figure 3F).

Using machine learning to identify the signature gene of UC

Two machine learning algorithms, SVM-RFE and RF, were employed to identify potential diagnostic biomarkers from DEGs. Utilizing the SVM-RFE algorithm, we pinpointed 43 genes with distinct features, achieved by minimizing cross-validation error (Figures 4A, B). Additionally, RF highlighted the top 15 genes (Figure 4C) and identified a total of 5 UC signature biomarkers—THY1, SLC6A14, ECSCR, FAP, and GPR109B—by taking the intersection of the genes (Figure 4D).

To enhance our comprehension of the predictive value associated with these signature genes for UC, an analysis was undertaken on the five signature biomarkers across two datasets (Figures 4E, F). The results unveiled a significant upregulation of expression in UC patients for these 5 characterized genes. Following this observation, a logistic regression model was constructed utilizing these 5 characterized genes (refer to Figure 4G). The Area Under the Curve (AUC) values were determined to be 0.995

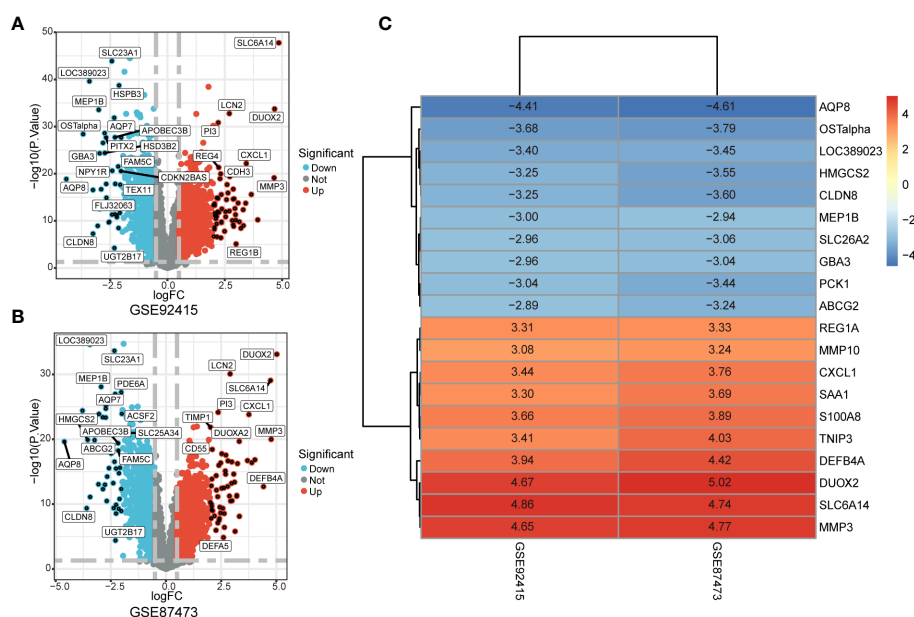
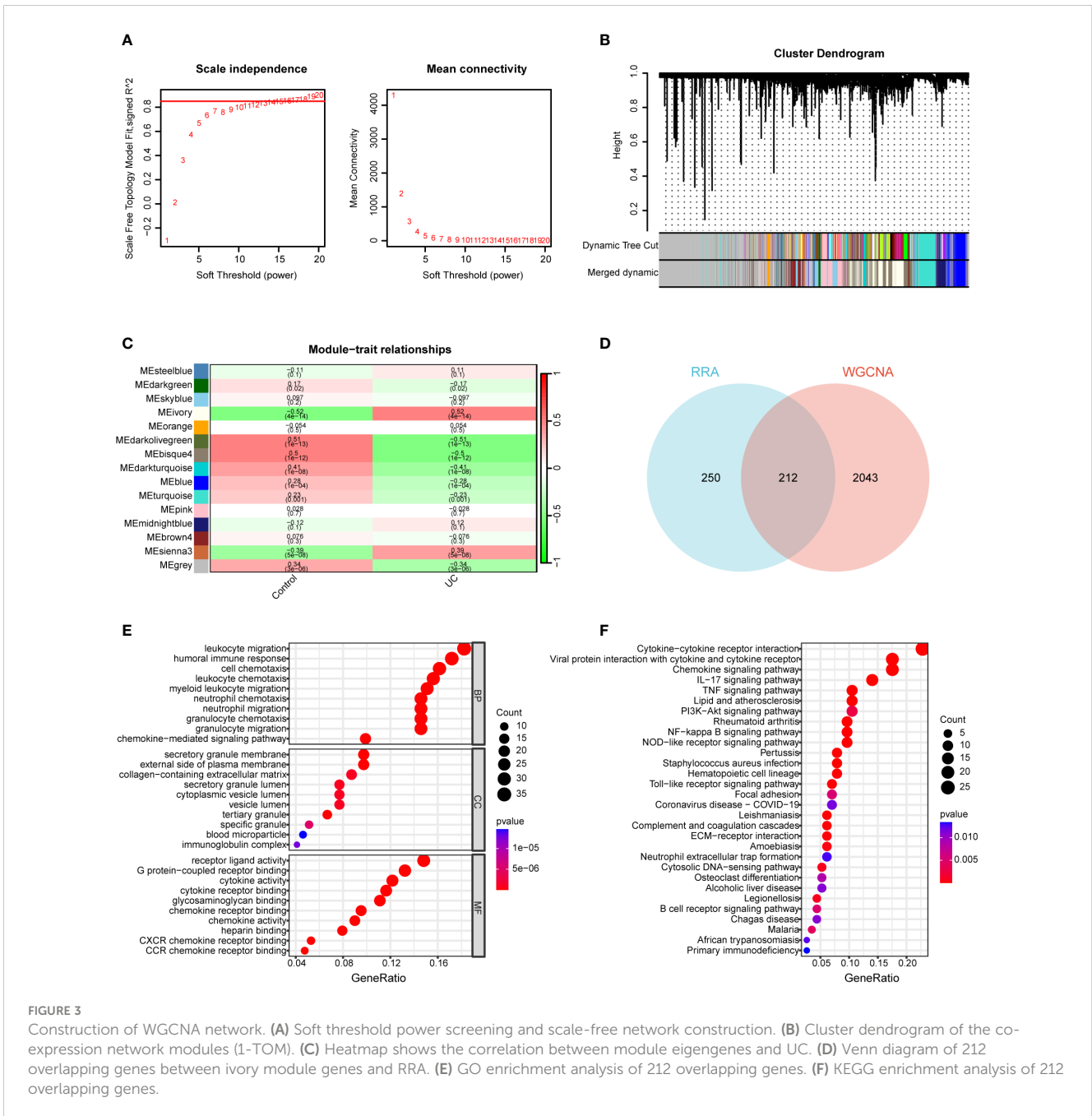


FIGURE 2 Identification of DEGs associated with UC. (A, B) Volcano plot shows DEGs in GSE92415 and GSE87473 datasets. (C) Heatmap of the top 10 up- and down-regulated DEGs identified in the RRA analysis.



and 0.959 in the model set and internal validation data, respectively. This signifies that the model exhibited superior efficacy in distinguishing between UC and healthy control samples compared to the individual prediction of the 5 characterized genes (as depicted in Figure 4H).

Investigation of specific signaling mechanisms associated with the UC signature gene

GSEA analysis was conducted to scrutinize the signaling pathways implicated in the five signature genes and explore their

impact on signaling pathways related to UC progression (Figure 5). The results demonstrated that genes associated with elevated THY1 expression were notably enriched in primary immunodeficiency and viral protein interaction with cytokine and cytokine receptor. In contrast, genes linked to low THY1 expression showed significant enrichment in ascorbate and aldarate metabolism, butanoate metabolism and citrate cycle (TCA cycle). For genes associated with high expression of SLC6A14, there was significant enrichment in the IL-17 signaling pathway and RA. Conversely, genes related to low SLC6A14 expression displayed enrichment in butanoate metabolism and the citrate cycle (TCA cycle). Genes associated with high ECSCR expression were significantly enriched in hematopoietic cell lineage and viral protein interaction with

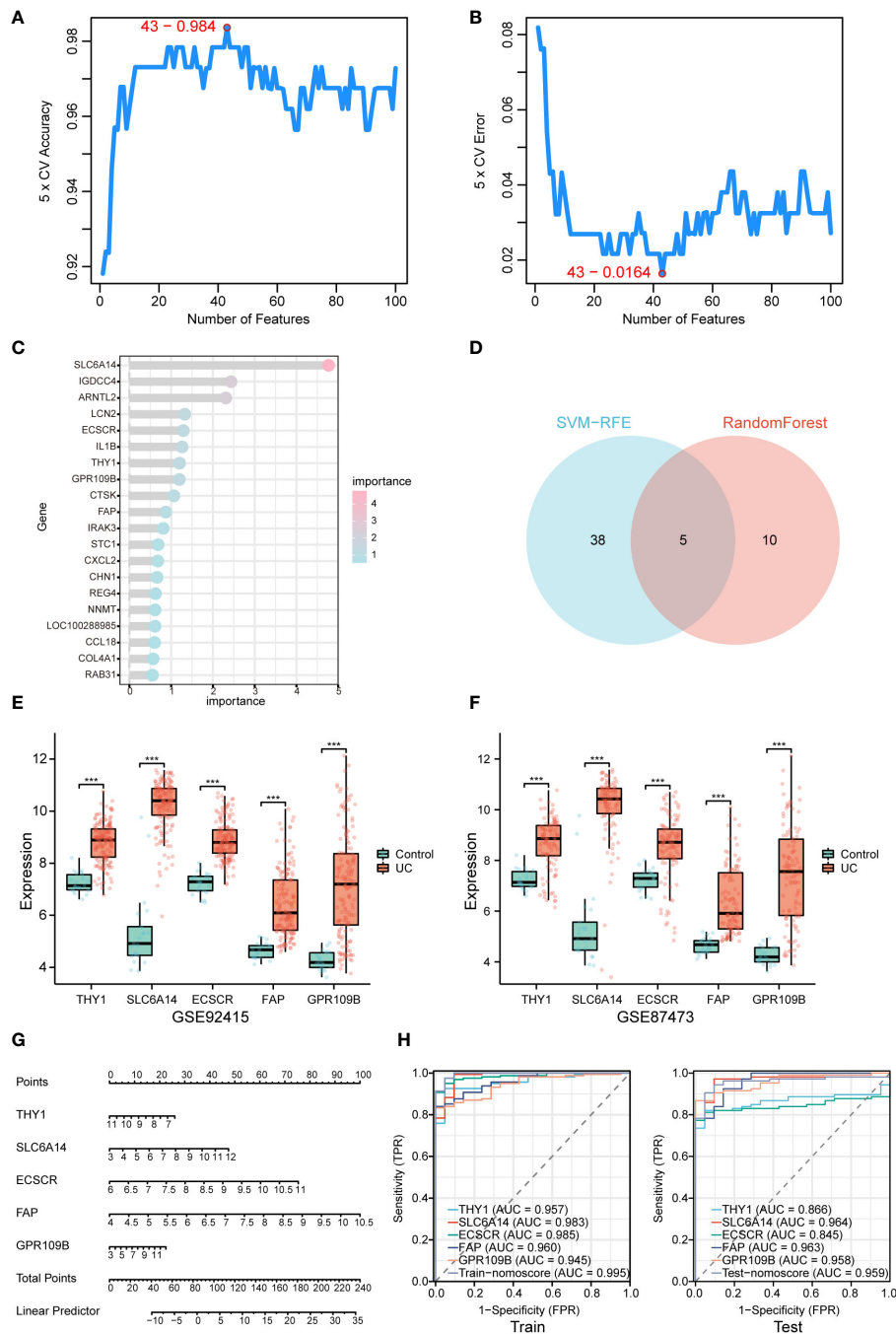


FIGURE 4 Screening for signature genes. **(A, B)** SVM-RFE algorithm was used to select the genes. **(C)** The genes were selected based on the RandomForest. **(D)** Venn diagram for two algorithmic. **(E, F)** Signature genes expression in GSE92415 dataset and GSE87473 dataset. **(G)** Nomogram for signature genes. **(H)** The ROC curve of the nomogram in the model set (GSE92415) and internal validation data (GSE87473). ***P < 0.001.

cytokine and cytokine receptors, while genes linked to low ECSCR expression showed significant enrichment in ascorbate and aldarate metabolism, as well as butanoate metabolism. Similarly, genes associated with high FAP expression exhibited significant enrichment in Hematopoietic cell lineage and primary immunodeficiency. In contrast, genes associated with low FAP expression displayed significant enrichment in butanoate metabolism and the citrate cycle (TCA cycle). Lastly, genes associated with high GPR109B expression displayed significant

enrichment in RA and viral protein interaction with cytokine and cytokine receptors. Conversely, genes associated with low GPR109B expression showed significant enrichment in butanoate metabolism and the citrate cycle (TCA cycle).

We conducted an analysis to compare GSVA scores in the Hallmark pathway between normal controls and UC patients. The results revealed a significant and predominant upregulation of the Hallmark pathway in UC patients. Specifically, upregulation was observed in KRAS signaling, IL2-STAT5 signaling, interferon

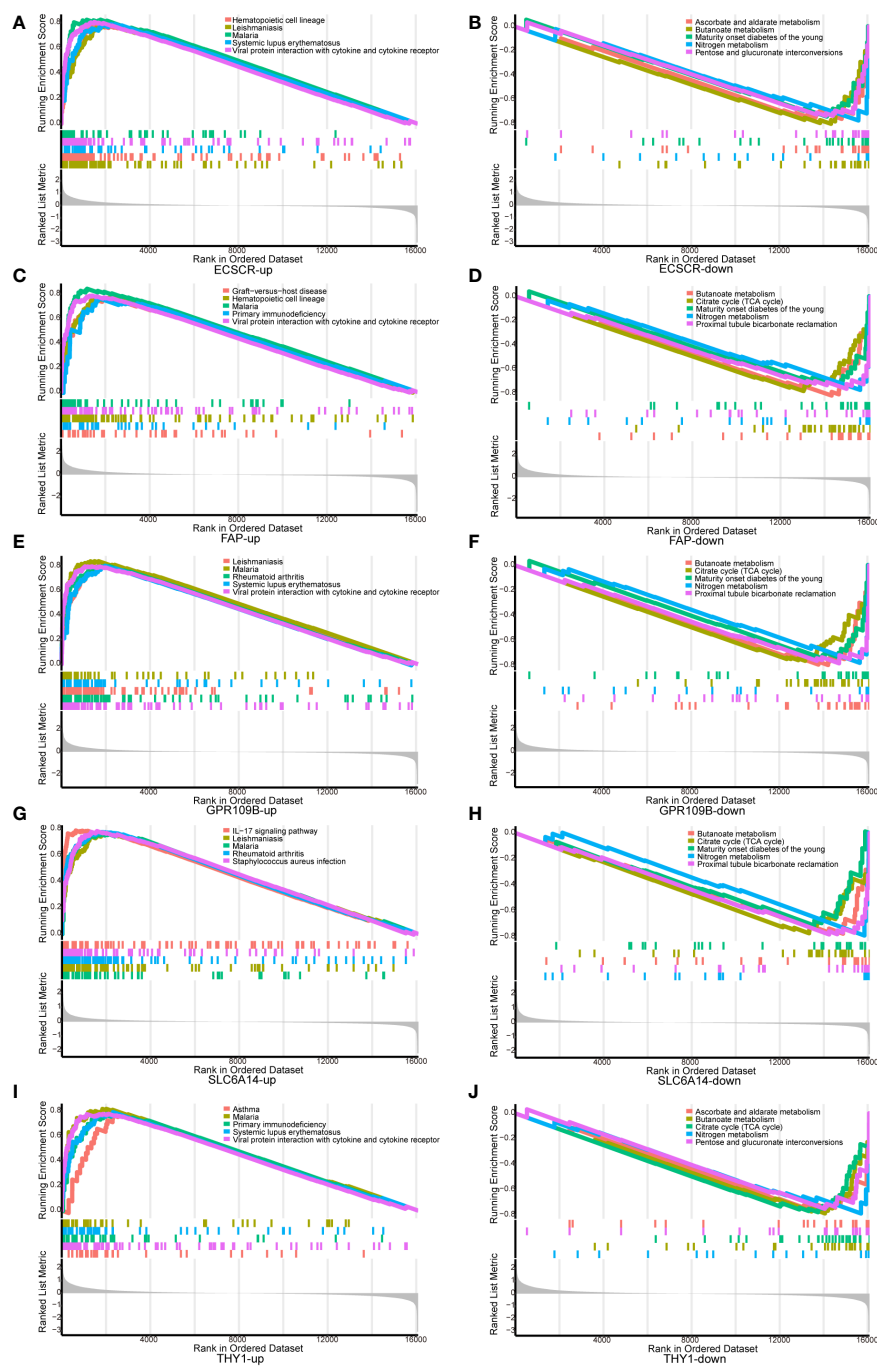


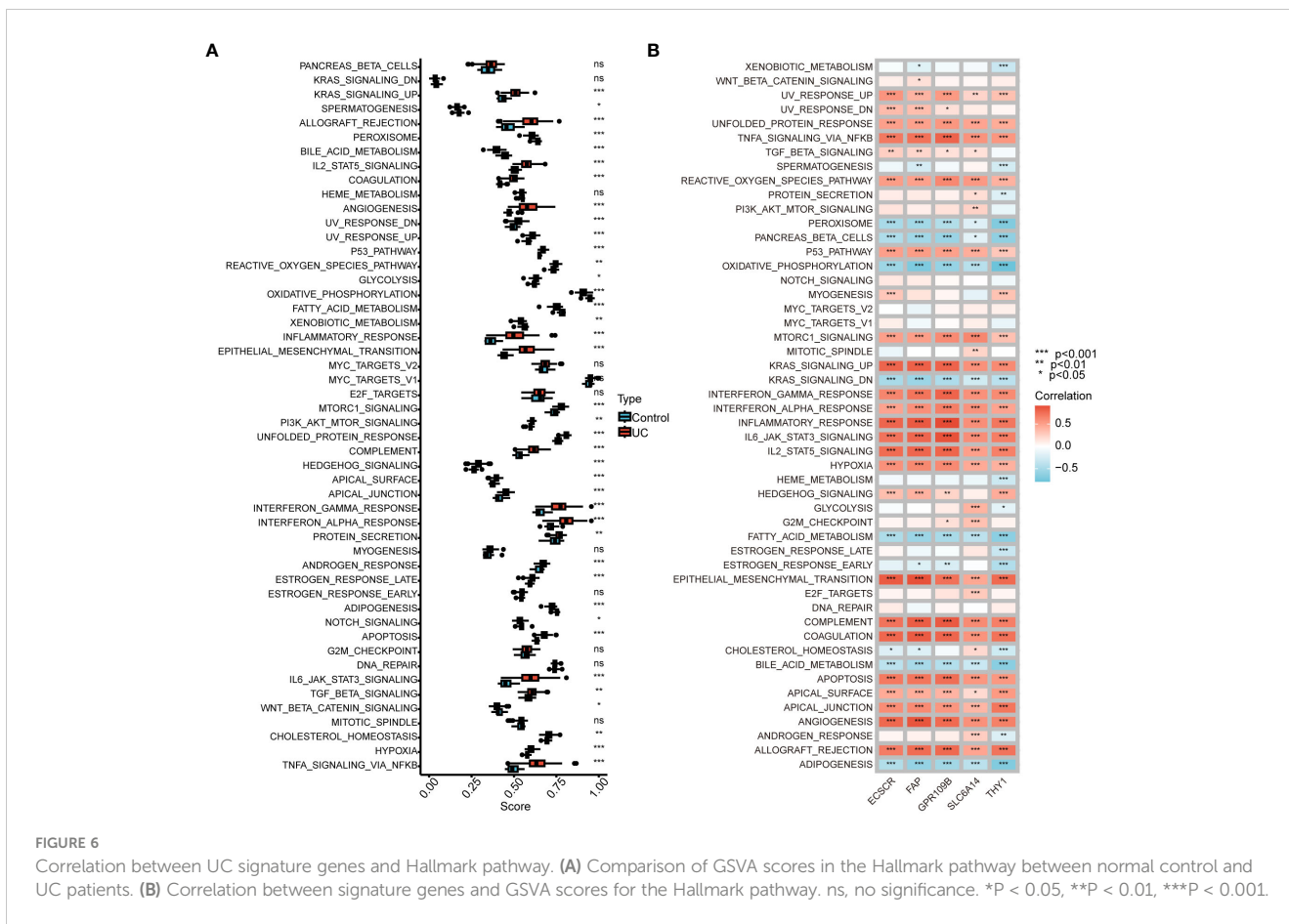
FIGURE 5
GSEA analysis of signature genes. (A, B) ECSCR, (C, D) FAP, (E, F) GPR109B, (G, H) SLC6A14, (I, J) THY1.

gamma response, interferon alpha response, IL6-JAK-STAT3 signaling, and TNFA signaling via NFKB. Conversely, bile acid metabolism, oxidative phosphorylation, and fatty acid metabolism were significantly downregulated (Figure 6A). Furthermore, we explored the correlation between genes in our model and GSVA scores for the Hallmark pathway. The analysis demonstrated a positive correlation between gene expression in the model and the Hallmark pathway upregulated in UC patients, as well as a negative correlation with the Hallmark pathway downregulated in UC patients (Figure 6B). These findings provide additional support

for the involvement of genes in our model in the progression of UC through various pathways.

Immune infiltration and correlation

To more precisely identify the colon immune cells associated with UC, the levels of 22 immune cell types were assessed in colon samples using CIBERSORT (Figure 7A). In comparison to healthy controls, UC patients exhibited elevated levels of M0 and M1



macrophages, along with decreased levels of resting CD4⁺ T memory cells and activated dendritic cells.

The association between immune cells and gene expression levels in the model was examined using Pearson correlation analysis (Figure 7B). The genes demonstrated a negative correlation with CD8⁺ T cells, resting memory CD4⁺ T cells, regulatory T cells, and M2 macrophages. Conversely, they exhibited a positive correlation with activated memory CD4⁺ T cells, M0 macrophages, M1 macrophages, activated mast cells, and neutrophils. This indicates that the varied expression of biomarkers has an impact on immune infiltration in UC.

Signature gene expression is strongly associated with different types of UC and drug efficacy

In the GSE11223 dataset, UC patients were stratified based on the duration of illness. Among the genes examined, namely THY1, SLC6A14, and FAP, significantly higher expression levels were observed in long-term UC (≥ 10 years). Conversely, in short-term UC (< 10 years), only THY1 exhibited elevated expression. Notably, ECSCR demonstrated increased expression in short-term UC but not in long-term UC (Figure 8A).

In the GSE107499 dataset, variations in the expression of colonic tissues were identified between inflamed and uninfamed states in UC (Figure 8B). The inflamed tissues exhibited a significant upregulation of signature genes compared to uninfamed tissues.

The GSE53306 dataset explores changes in gene expression among normal, active, and inactive colon tissues in patients with UC (Figure 8C). THY1, SLC6A14, and ECSCR exhibited a significant increase in both active and inactive UC tissues compared to normal tissue. However, there was no significant difference in their expression levels between active and inactive UC.

GSE206285 was to assess the expression profile of baseline biopsy samples from patients with moderate-to-severe UC treated with the IL-12/IL-23 inhibitor ustekinumab (Ust). THY1 levels were significantly lower in the responsive and mucosal healed UC patients compared to the non-responsive group before treatment with Ust. There was a slight decrease in SLC6A14 and ECSCR levels (Figures 8D, E).

The GSE92415 dataset comprises expression profiles from biopsy samples of UC patients treated with golimumab (GLM). Before the initiation of GLM treatment, patients with active UC exhibited higher expression levels of THY1, SLC6A14, and ECSCR compared to healthy controls. Following GLM treatment, although the expression of THY1 was reduced in the clinical remission group,

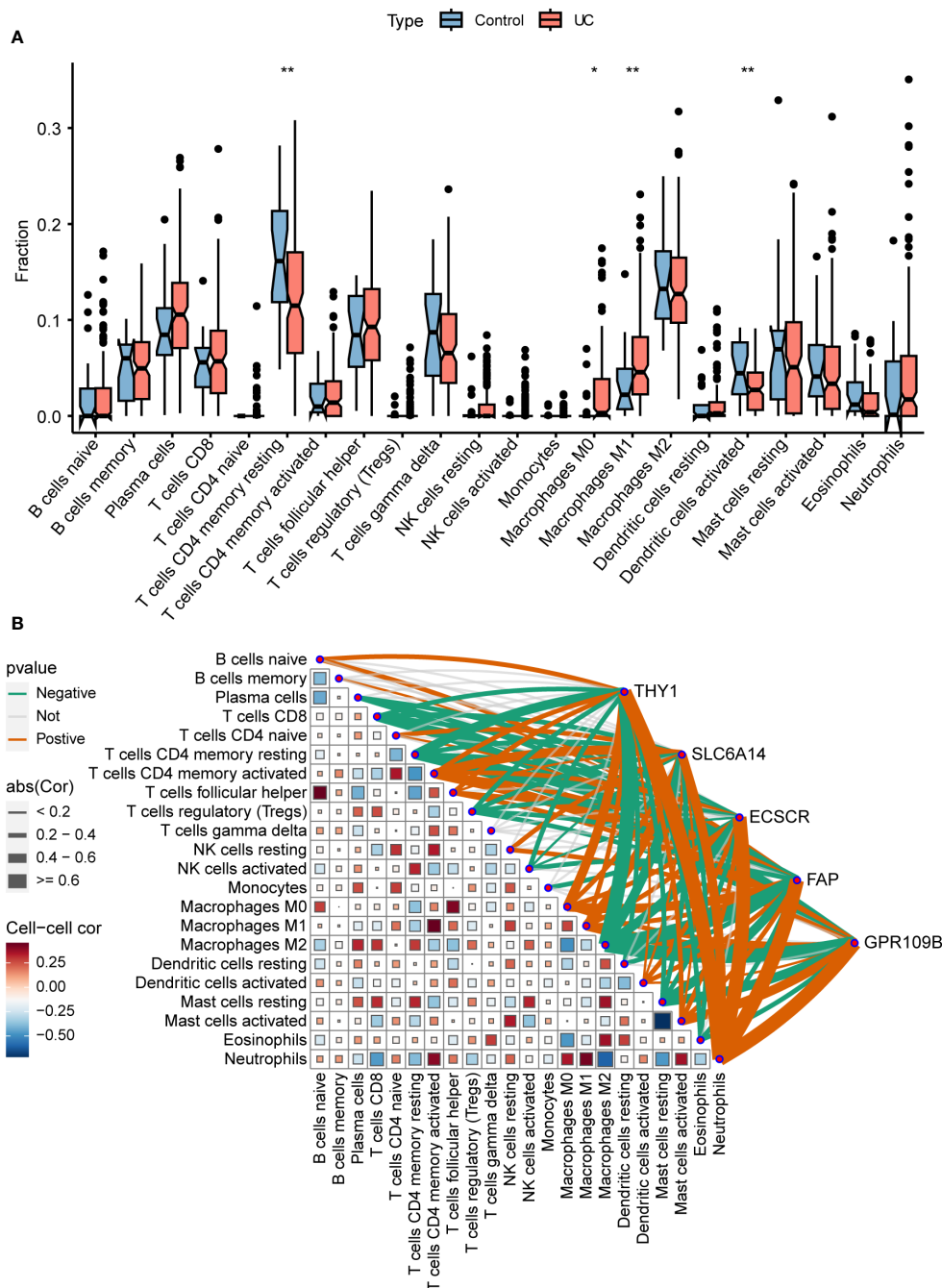


FIGURE 7 Immune infiltration and correlation. **(A)** Differences in infiltrating immune cells between normal control and UC. **(B)** Correlation of signature genes and 22 immune cell types. * $P < 0.05$, ** $P < 0.01$.

the expression levels of THY1, SLC6A14, and ECSCR did not fully revert to those observed in healthy controls (Figures 8F–H).

The GSE92415 cohort contains information on the Mayo score for all samples. The results showed that according to the Mayo score > 5 was the high scoring group and the rest was the low scoring group, in the high scoring group the gene expression was significantly upregulated in the model. Correlation analysis further confirmed that gene expression in the model was positively correlated with Mayo score (Supplementary Figure S1).

ELISA of clinical serum samples showed that ECSCR was upregulated in the serum of UC patients and positively correlated with neutrophil levels (Figures 8I, J).

Discussion

UC is characterized by a chronic, recurring IBD with a higher incidence rate (33). The low early diagnosis rate of UC, attributed to

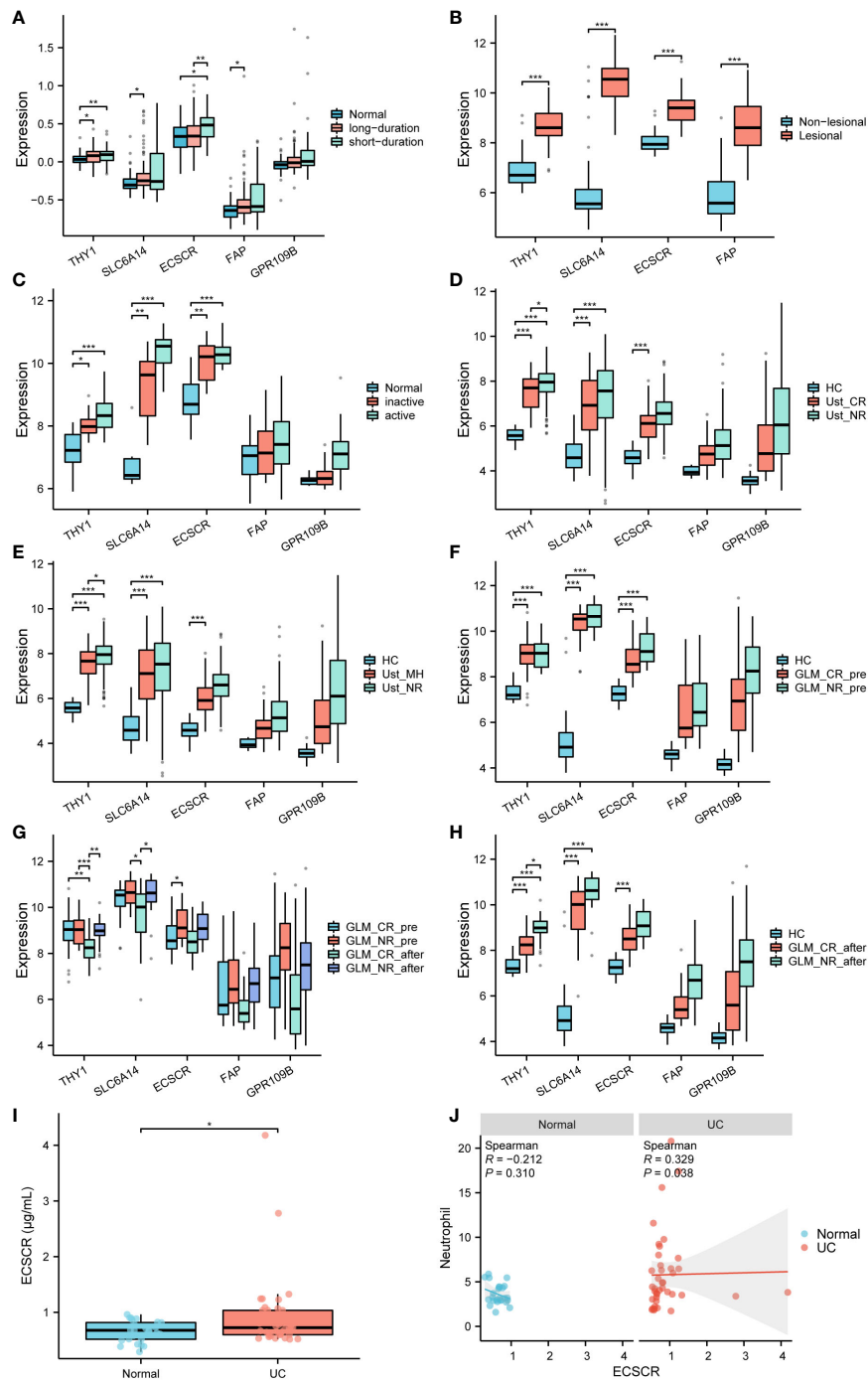


FIGURE 8

External validation of signature genes. **(A)** Differential expression of signature genes in normal control, long duration UC and short duration UC. **(B)** Differential expression of signature genes in lesional and non-lesional colonic tissues of UC patients. **(C)** Differential expression of signature genes in normal control, inactive and active UC. **(D, E)** Differential expression of signature genes in the colonic mucosal of HC (healthy control), Ust_CR (UC patients in clinical remission before ustekinumab therapy), Ust_NR (UC patients not responding before ustekinumab therapy), and Ust_MH (UC patients in mucosal healing before ustekinumab therapy). **(F–H)** Differential expression of signature genes in the colonic mucosa of healthy controls, UC patients in the non-response and response groups before and after GLM treatment. **(I)** Contents of ECSCR in serum of normal and UC patients were determined by ELISA. **(J)** Correlation between ECSCR and neutrophil levels. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

the absence of reliable biomarkers, leads to a delay in achieving initial remission of the disease (34). In order to enhance mucosal healing and remission rates, the identification of novel and effective diagnostic biomarkers accurately reflecting UC is imperative. This study employs RRA, WGCNA, and machine learning techniques to identify key

signature genes. Validation of these signature genes is conducted using multiple patient datasets, encompassing disease onset and medication details, along with clinical samples. The outcomes of this research are anticipated to contribute significantly to the discovery of new biomarkers for the diagnosis and treatment of UC.

In this study, 212 intersecting genes were identified through RRA and WGCNA analyses. Subsequent GO analysis revealed that these DEGs were associated with leukocyte migration, humoral immune response, cytotoxicity, and leukocyte chemotaxis, suggesting a pivotal role of leukocytes in the progression of UC. Neutrophils, crucial leukocyte cells, were observed in the epithelium, indicating histological UC activity and serving as an early core event in UC (35). These neutrophils coexisted with inflammation. The KEGG analysis demonstrated significant enrichment of genes related to the IL-17 signaling pathway, TNF signaling pathway, PI3K-Akt signaling pathway, RA and other immune-related pathways. Similarly, GSVA analysis indicated a significant activation of these pathways in UC patients, consistent with previous studies (7, 18). Intriguingly, these intersecting genes are not only associated with UC pathogenesis but also with RA. A large-sample meta-analysis suggested that individuals with inflammatory bowel inflammation are at a higher risk of developing RA (36, 37), indicating shared pathogenic factors for both UC and RA at the molecular level.

Five biomarkers (THY1, SLC6A14, ECSCR, FAP, and GPR109B) significantly associated with UC have been identified through machine learning analysis. A prediction model was developed, and ROC analysis demonstrated the model's and its UC signature gene's excellent discriminatory ability in distinguishing UC samples from healthy control samples. Furthermore, it is suggested that the UC signature gene may be closely linked to UC, providing a foundation for understanding the disease's etiology and discovering potential new therapies. THY1 is highly expressed in synovial fibroblasts in RA, which can invade and degrade cartilage by secreting inflammatory cytokines and chemokines, while stimulating osteoclasts, leading to bone erosion (38). A correlation between RA and UC (OR 1.082; 95% CI 1.002–1.168; $P = 0.044$) was found by Mendelian randomization analysis (39). In addition, THY1, implicated in cell adhesion during inflammation, exhibits aberrant methylation associated with UC (40). These findings suggest that THY1 may be involved in the comorbid processes of RA and UC. SLC6A14, also known as a Na^+/Cl^- driven amino acid transporter B, is up-regulated in both rectal and caecal mucosa in UC patients (41). Studies indicate that SLC6A14 expression is upregulated in UC patients, potentially contributing to colonic inflammation by regulating glutamine and nitric oxide synthase 2 and may also contribute to UC via the C/EBP β -PAK6 axis Iron death in epithelial cells (42–44). Moreover, there has been a suggestion that SLC6A14 contributes to the death of UC cells by controlling NLRP3 (45). FAP is a significant marker of cancer-associated fibroblasts and is closely linked to colorectal cancer invasion and metastasis. Additionally, the extent, duration, and severity of inflammation in UC are tied to a higher risk of colitis-associated colorectal cancer (CAC) (13). Furthermore, FAP levels are unusually high in UC. Further investigation is needed to determine if FAP plays a crucial role in the progression of UC to CAC. GPR109B, abundantly expressed in human neutrophils, plays a role in inflammatory processes, including UC. Targeting GPR109B in therapeutic strategies may prove beneficial in

diseases characterized by inflammation, such as UC (46). Furthermore, the five aforementioned biomarkers are positively associated with UC activation pathways and negatively correlated with UC down-regulated signaling pathways, supporting their significance in UC development and progression.

Furthermore, notable variations in the composition of immune cells were observed between UC and control samples. UC patients exhibited a higher abundance of M1 macrophages compared to controls, while there was no difference in M2 macrophages. This observation aligns with the understanding that UC is primarily associated with the presence of pro-inflammatory M1 macrophages, but not anti-inflammatory M2 macrophages (47). Additionally, it was found that resting CD4^+ T memory cells and activated dendritic cells were significantly lower in UC patients, contradicting previous findings (48). These differences may be attributed to the utilization of different datasets or the presence of unbalanced data in prior studies. Surprisingly, five biomarkers were found to have a significant positive correlation with macrophage M0 and M1 infiltration in the model. However, there is a scarcity of studies investigating the impact of these genes on immune cells in UC patients. Moreover, the genes in the model are positively correlated with neutrophils. It has been demonstrated that large numbers of neutrophils infiltrate the UC colonic mucosa, releasing serine proteases, matrix metalloproteinases, and myeloperoxidases through the production of reactive oxygen species, which directly cause tissue damage and produce typical crypt abscesses (13). Additionally, ELISA of clinical serum samples showed that ECSCR was upregulated in the serum of UC patients and positively correlated with neutrophil levels. These findings suggest that genes in the model are closely associated with neutrophil infiltration in UC. Our findings offer novel insights into the potential role of these genes in UC immunomodulation. Therefore, further investigation into the function of immune cells in the progression of UC is warranted.

Characterized gene expression has been confirmed to be strongly associated with various types of UC and the effectiveness of drugs in different datasets. Additionally, the diagnostic efficacy of ECSCR for UC has been further validated using clinical samples. Our hope is that these findings will present novel strategies for the diagnosis and treatment of UC. However, it is important to recognize certain limitations in this study. Firstly, the clinical data used were obtained from public databases, and the clinical information of the samples was incomplete, which hindered the exploration of the correlation between these characterized genes and clinical features. Secondly, no *in vivo* or *in vitro* experiments were conducted for validation. Therefore, further studies will be necessary to provide compelling evidence for our results.

In summary, novel targets such as THY1, SLC6A14, ECSCR, FAP, and GPR109B have been discovered in our findings, which could potentially play a role in the development of UC and serve as reliable diagnostic biomarkers for UC. Additionally, these new targets exhibit strong associations with various signaling pathways and immune cells involved in UC, thereby offering fresh insights into the underlying mechanisms of this condition.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Ethics statement

The Ethics Committee of The First Affiliated Hospital of Soochow University reviewed and approved the study (ID:2024296).

Author contributions

BF: Data curation, Investigation, Writing – original draft. YZ: Conceptualization, Formal analysis, Methodology, Writing – review & editing. LQ: Funding acquisition, Project administration, Resources, Writing – original draft. QT: Conceptualization, Resources, Visualization, Writing – original draft. ZZ: Methodology, Software, Writing – original draft. SZ: Resources, Validation, Visualization, Writing – original draft. JQ: Investigation, Software, Supervision, Writing – original draft. XZ: Formal analysis, Project administration, Resources, Writing – review & editing. CH: Conceptualization, Formal analysis, Validation, Writing – review & editing. YL: Funding acquisition, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work

References

- Shi L, Han X, Li JX, Liao YT, Kou FS, Wang ZB, et al. Identification of differentially expressed genes in ulcerative colitis and verification in a colitis mouse model by bioinformatics analyses. *World J Gastroenterol.* (2020) 26:5983–96. doi: 10.3748/wjg.v26.i39.5983
- de Negreiros LMV, Pascoal LB, Genaro LM, Silva JF, Rodrigues BL, Camargo MG, et al. Pouchitis: insight into the pathogenesis and clinical aspects. *Am J Transl Res.* (2022) 14:4406–25.
- Jiang S, Miao Z. High-fat diet induces intestinal mucosal barrier dysfunction in ulcerative colitis: emerging mechanisms and dietary intervention perspective. *Am J Transl Res.* (2023) 15:653–77.
- Shivashankar R, Tremaine WJ, Harmsen WS, Loftus EV Jr. Incidence and prevalence of crohn's disease and ulcerative colitis in olmsted county, minnesota from 1970 through 2010. *Clin Gastroenterol Hepatol.* (2017) 15:857–63. doi: 10.1016/j.cgh.2016.10.039
- Nambu R, Arai K, Kudo T, Murakoshi T, Kunisaki R, Mizuochi T, et al. Clinical outcome of ulcerative colitis with severe onset in children: a multicenter prospective cohort study. *J Gastroenterol.* (2023) 58:472–80. doi: 10.1007/s00535-023-01972-1
- Guo J, Chen N, Tan F, Zhou J, Xiang H, Luo Y, et al. iTRAQ-based proteomic analysis of imiquimod in the treatment of ulcerative colitis. *Am J Transl Res.* (2023) 15:4454–66.
- Cheng F, Li Q, Wang J, Zeng F, Wang K, Zhang Y. Identification of differential intestinal mucosa transcriptomic biomarkers for ulcerative colitis by bioinformatics analysis. *Dis Markers.* (2020) 2020:8876565. doi: 10.1155/2020/8876565
- Feuerstein JD, Cheifetz AS. Ulcerative colitis: epidemiology, diagnosis, and management. *Mayo Clin Proc.* (2014) 89:1553–63. doi: 10.1016/j.mayocp.2014.07.002
- Kou F, Cheng Y, Shi L, Liu J, Liu Y, Shi R, et al. LCN2 as a potential diagnostic biomarker for ulcerative colitis-associated carcinogenesis related to disease duration. *Front Oncol.* (2021) 11:793760. doi: 10.3389/fonc.2021.793760
- Xu M, Kong Y, Chen N, Peng W, Zi R, Jiang M, et al. Identification of immune-related gene signature and prediction of ceRNA network in active ulcerative colitis. *Front Immunol.* (2022) 13:855645. doi: 10.3389/fimmu.2022.855645
- Deng B, Liao F, Liu Y, He P, Wei S, Liu C, et al. Comprehensive analysis of endoplasmic reticulum stress-associated genes signature of ulcerative colitis. *Front Immunol.* (2023) 14:1158648. doi: 10.3389/fimmu.2023.1158648
- Huang J, Zhang J, Wang F, Zhang B, Tang X. Comprehensive analysis of cuproptosis-related genes in immune infiltration and diagnosis in ulcerative colitis. *Front Immunol.* (2022) 13:1008146. doi: 10.3389/fimmu.2022.1008146
- Zhang C, Zhang J, Zhang Y, Song Z, Bian J, Yi H, et al. Identifying neutrophil-associated subtypes in ulcerative colitis and confirming neutrophils promote colitis-associated colorectal cancer. *Front Immunol.* (2023) 14:1095098. doi: 10.3389/fimmu.2023.1095098
- Brookes MJ, Whitehead S, Gaya DR, Hawthorne AB. Practical guidance on the use of faecal calprotectin. *Frontline Gastroenterol.* (2018) 9:87–91. doi: 10.1136/flgastro-2016-100762
- Huang Y, Liu J, Liang D. Comprehensive analysis reveals key genes and environmental toxin exposures underlying treatment response in ulcerative colitis

was supported by National Natural Science Foundation of China (81901632, and 82001576), Key Clinical Technology Research in Suzhou (SKY2023001), and the Primary Research & Development Plan of Jiangsu Province (BE2022736).

Acknowledgments

We thank all the families for participating in this research project and The Charlesworth Group for language editing.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2024.1426875/full#supplementary-material>

- based on in-silico analysis and Mendelian randomization. *Aging (Albany NY)*. (2023) 15:14141–71. doi: 10.18632/aging.v15i23
16. Li C, Wu Y, Xie Y, Zhang Y, Jiang S, Wang J, et al. Oral manifestations serve as potential signs of ulcerative colitis: A review. *Front Immunol*. (2022) 13:1013900. doi: 10.3389/fimmu.2022.1013900
17. Spencer EA, Bergstein S, Dolinger M, Pittman N, Kellar A, Dunkin D, et al. Single-center experience with upadacitinib for adolescents with refractory inflammatory bowel disease. *Inflammation Bowel Dis*. (2023). doi: 10.1093/ibd/izad300
18. Huang J, Zhang J, Wang F, Zhang B, Tang X. Revealing immune infiltrate characteristics and potential diagnostic value of immune-related genes in ulcerative colitis: An integrative genomic analysis. *Front Public Health*. (2022) 10:1003002. doi: 10.3389/fpubh.2022.1003002
19. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. (2015) 43:e47. doi: 10.1093/nar/gkv007
20. Zhang S, Lin K, Qiu J, Feng B, Wang J, Li J, et al. Identification of potential key autophagy-related genes in asthma with bioinformatics approaches. *Am J Transl Res*. (2022) 14:7350–61.
21. Feng B, Zhou T, Guo Z, Jin J, Zhang S, Qiu J, et al. Comprehensive analysis of immune-related genes for classification and immune microenvironment of asthma. *Am J Transl Res*. (2023) 15:1052–62.
22. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*. (2012) 28:573–80. doi: 10.1093/bioinformatics/btr709
23. 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
24. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Cambridge (Mass))*. (2021) 2:100141. doi: 10.1016/j.xinn.2021.100141
25. Lin K, Wang T, Tang Q, Chen T, Lin M, Jin J, et al. IL18R1-related molecules as biomarkers for asthma severity and prognostic markers for idiopathic pulmonary fibrosis. *J Proteome Res*. (2023) 22:3320–31. doi: 10.1021/acs.jproteome.3c00389
26. Deist TM, Dankers F, Valdes G, Wijsman R, Hsu IC, Oberije C, et al. Machine learning algorithms for outcome prediction in (chemo)radiotherapy: An empirical comparison of classifiers. *Med physics*. (2018) 45:3449–59. doi: 10.1002/mp.12967
27. Alakwaa FM, Chaudhary K, Garmire LX. Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data. *J Proteome Res*. (2018) 17:337–47. doi: 10.1021/acs.jproteome.7b00595
28. Wu J, Zhang H, Li L, Hu M, Chen L, Xu B, et al. A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: A population-based analysis. *Cancer Commun (London England)*. (2020) 40:301–12. doi: 10.1002/cac2.12067
29. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf*. (2011) 12:77. doi: 10.1186/1471-2105-12-77
30. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
31. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. (2015) 12:453–7. doi: 10.1038/nmeth.3337
32. Zhang Y, Zhou T, Tang Q, Feng B, Liang Y. Identification of glycosyltransferase-related genes signature and integrative analyses in patients with ovarian cancer. *Am J Clin Exp Immunol*. (2024) 13:12–25. doi: 10.62347/OSUA8322
33. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet*. (2017) 390:2769–78. doi: 10.1016/S0140-6736(17)32448-0
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. Kaplanski G, Marin V, Montero-Julian F, Mantovani A, Farnarier C. IL-6: a regulator of the transition from neutrophil to monocyte recruitment during inflammation. *Trends Immunol*. (2003) 24:25–9. doi: 10.1016/S1471-4906(02)00013-3
36. Chen Y, Chen L, Xing C, Deng G, Zeng F, Xie T, et al. The risk of rheumatoid arthritis among patients with inflammatory bowel disease: a systematic review and meta-analysis. *BMC Gastroenterol*. (2020) 20:192. doi: 10.1186/s12876-020-01339-3
37. Skejoe C, Hansen AS, Stengaard-Pedersen K, Junker P, Hoerslev-Pedersen K, Hetland ML, et al. T-cell immunoglobulin and mucin domain 3 is upregulated in rheumatoid arthritis, but insufficient in controlling inflammation. *Am J Clin Exp Immunol*. (2022) 11:34–44.
38. Wei K, Korsunsky I, Marshall JL, Gao A, Watts GFM, Major T, et al. Notch signalling drives synovial fibroblast identity and arthritis pathology. *Nature*. (2020) 582:259–64. doi: 10.1038/s41586-020-2222-z
39. Meisinger C, Freuer D. Rheumatoid arthritis and inflammatory bowel disease: A bidirectional two-sample Mendelian randomization study. *Semin Arthritis Rheumatol*. (2022) 55:151992. doi: 10.1016/j.semarthrit.2022.151992
40. Häslér R, Feng Z, Bäckdahl L, Spehlmann ME, Franke A, Teschendorff A, et al. A functional methylation map of ulcerative colitis. *Genome Res*. (2012) 22:2130–7. doi: 10.1101/gr.138347.112
41. Eriksson A, Flach CF, Lindgren A, Kvifors E, Lange S. Five mucosal transcripts of interest in ulcerative colitis identified by quantitative real-time PCR: a prospective study. *BMC Gastroenterol*. (2008) 8:34. doi: 10.1186/1471-230X-8-34
42. Chen Y, Yan W, Chen Y, Zhu J, Wang J, Jin H, et al. SLC6A14 facilitates epithelial cell ferroptosis via the C/EBPβ-PAK6 axis in ulcerative colitis. *Cell Mol Life Sci*. (2022) 79:563. doi: 10.1007/s00018-022-04594-7
43. Novak EA, Mollen KP. Mitochondrial dysfunction in inflammatory bowel disease. *Front Cell Dev Biol*. (2015) 3:62. doi: 10.3389/fcell.2015.00062
44. Kolios G, Valatas V, Ward SG. Nitric oxide in inflammatory bowel disease: a universal messenger in an unsolved puzzle. *Immunology*. (2004) 113:427–37. doi: 10.1111/j.1365-2567.2004.01984.x
45. Gu Q, Xia H, Song YQ, Duan J, Chen Y, Zhang Y, et al. SLC6A14 promotes ulcerative colitis progression by facilitating NLRP3 inflammasome-mediated pyroptosis. *World J Gastroenterol*. (2024) 30:252–67. doi: 10.3748/wjg.v30.i3.252
46. Irukayama-Tomobe Y, Tanaka H, Yokomizo T, Hashidate-Yoshida T, Yanagisawa M, Sakurai T. Aromatic D-amino acids act as chemoattractant factors for human leukocytes through a G protein-coupled receptor, GPR109B. *Proc Natl Acad Sci United States America*. (2009) 106:3930–4. doi: 10.1073/pnas.0811844106
47. Yang Y, Hua Y, Zheng H, Jia R, Ye Z, Su G, et al. Biomarkers prediction and immune landscape in ulcerative colitis: Findings based on bioinformatics and machine learning. *Comput Biol Med*. (2024) 168:107778. doi: 10.1016/j.combiomed.2023.107778
48. Hu W, Fang T, Zhou M, Chen X. Identification of hub genes and immune infiltration in ulcerative colitis using bioinformatics. *Sci Rep*. (2023) 13:6039. doi: 10.1038/s41598-023-33292-y