ORIGINAL ARTICLE

UDC: 004:61]:616.348/.349-002-07-037-08
DOI: https://doi.org/10.2298/VSP250131049S

Observational study on the potential mechanism of SiNi powder in the treatment of ulcerative colitis based on network pharmacology and machine learning

Opservaciona studija o potencijalnom mehanizmu delovanja SiNi praha u lečenju ulceroznog kolitisa zasnovana na farmakologiji mreže i mašinskom učenju

Sihong Shen*, Yongduo Yu†

***Liaoning University of Traditional Chinese Medicine, †The Second Affiliated Hospital, Shenyang, Liaoning Province, China**

Abstract

Background/ Aim. Chronic idiopathic ulcerative colitis (UC) damages and disrupts the intestinal mucosa. Diagnosing UC and differential diagnosis is tough. Its anti-inflammatory and immunosuppressive properties make SiNi powder (SNP) a popular treatment for inflammatory illnesses. The multi-target mechanism of SNP on UC is unknown. The aim of this study was to examine the potential mechanisms of SNP in UC treatment through network pharmacology and machine learning approaches, identify novel diagnostic biomarkers, and develop a predictive model for UC diagnosis. **Methods.** The Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP) assessed active constituents and target proteins. Using two public datasets (GSE87473 and GSE75214), differential analysis was conducted on the gene expression matrix of UC to find the intersection of differentially expressed genes and SNP-related targets. Hub genes were assessed using several machine-learning algorithms to create a prediction model. Single-cell analysis studies were used to diagnose genes and immune cells. NetworkAnalyst predicted upstream transcription factors, micro-ribonucleic acids, and the protein-compound network. **Results.** According to the TCMSP database, the SNP included 95 active constituents and 795 associated targets against UC. After identifying 79 overlapping genes, machine learning discovered five hub genes: *TRPV1*, *ABCG2*, *BACE2*, *MMP3*, and *LIPC*. Diagnostics were verified using external datasets. These genes were used to create a predictive model with a large area under the curve (AUC = 1,000) and an external validation dataset with 1,000 AUCs, demonstrating excellent accuracy of the predictive model and the hub genes. **Conclusion.** SNP and UC are associated, and hub genes were found to evaluate UC risk. This computational technique opens new avenues for UC biomarker and therapeutic target research, although further experimental validation is required to confirm and validate these results.

Key words:
colitis, ulcerative; medicine, chinese traditional; gene expression; network pharmacology; treatment outcome.

Apstrakt

Uvod/Cilj. Hronični idiopatski ulcerozni kolitis (UK) oštećuje i remeti crevnu mukozu. Dijagnostikovanje UK-a i uspostavljanje diferencijalne dijagnoze je teško. Antiinflamacijska i imunosupresivna svojstva SiNi praha (SNP) čine ga popularnim tretmanom za inflamacijske bolesti. Višestruki mehanizam delovanja SNP-a u lečenju UK-a nije poznat. Cilj rada bio je da se ispitaju potencijalni mehanizmi delovanja SNP-a u lečenju UK-a primenom pristupa farmakologije mreže i mašinskog učenja, identifikuju novi dijagnostički biomarkeri i razvije prediktivni model za dijagnozu UK-a. **Metode.** Baza podataka i platforma za analizu sistemske tradicionalne kineske medicine (*Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform* – TCMSP) procenila je aktivne sastojke i ciljne proteine. Korišćenjem dva javno dostupna skupa podataka (GSE87473 i GSE75214), sprovedena je diferencijalna analiza matrice ekspresije gena kod UK-a u cilju pronalaženja preklapanja diferencijalno ispoljenih gena i meta delovanja povezanih sa SNP-om. *Hub* geni procenjeni su korišćenjem nekoliko algoritama mašinskog učenja da bi se kreirao model predikcije. Studije analize na nivou pojedinačnih ćelija korišćene su za identifikaciju gena i imunskih ćelija. *NetworkAnalyst* je predvideo ushodne transkripcione faktore,

Correspondence to: Yongduo Yu, Liaoning University of Traditional Chinese Medicine, The Second Affiliated Hospital, Huanghe North Street, Yuhong District, Shenyang City, Liaoning Province, 200 031, China. E-mail: yongduoyu20@gmail.com, yongduoyu@163.com

mikro-ribonukleinske kiseline i mrežu protein-jedinjenja. **Rezultati.** Prema TCMSP bazi podataka, SNP je uključivao 95 aktivnih sastojaka i 795 povezanih meta delovanja protiv UK-a. Nakon identifikovanja 79 gena koji se preklapaju, mašinsko učenje je otkrilo pet *hub* gena: *TRPV1*, *ABCG2*, *BACE2*, *MMP3* i *LIPC*. Provera dijagnostike je izvršena korišćenjem eksternih skupova podataka. Ovi geni su korišćeni za kreiranje prediktivnog modela sa velikom površinom ispod krive [*area under the curve* (AUC) = 1.000] i skupa podataka za eksternu validaciju sa 1.000 AUC, demonstrirajući odličnu tačnost prediktivnog modela i *hub* gena. **Zaključak.** SNP i UK su povezani, a utvrđeni su *hub* geni za procenu rizika od UK-a. Ova računarska tehnika otvara nove puteve za istraživanje biomarkera UK-a i ciljeva terapijskog delovanja, iako je potrebna dalja eksperimentalna provera da bi se potvrdili i validirali ovi rezultati.

Ključne reči:
kolitis, ulcerozni; medicina, kineska, tradicionalna; geni, ekspresija; farmakologija, mrežna; lečenje, ishod.

Introduction

Inflammatory bowel disease (IBD) is a persistent and relapsing inflammatory disorder of unknown etiology that impacts the gastrointestinal system. Ulcerative colitis (UC), a subtype of IBD, is defined by persistent inflammation of the colonic mucosa. Prevalent symptoms include diarrhea, stomach discomfort, the excretion of mucus and blood, and weight loss [1]. In addition to intestinal symptoms, UC often presents with extraintestinal manifestations that impact the liver and bile ducts, such as fatty liver and primary sclerosing cholangitis [2, 3]. The pathogenesis of UC remains incompletely elucidated; however, it is acknowledged as a complex IBD. The worldwide incidence and prevalence of UC are increasing rapidly, presenting a considerable threat to global public health. Despite advances in treatment options, effectively managing UC remains a significant challenge, particularly in achieving consistent clinical remission and preventing disease progression [4].

Traditional Chinese medicine (TCM) has promising advantages in UC treatment [5, 6]. Conventional Chinese medicine primarily modulates inflammatory cytokines, intestinal microbiota, and the immune mechanism while also protecting the intestinal mucosa [7]. Hence, it can play a role in treating UC. SiNi powder (SNP), a classical Chinese medicine formula, has demonstrated effectiveness and is reported to cause fewer adverse reactions in the treatment of inflammatory and bowel diseases [5, 8, 9]. Each component of SNP – *Radix Bupleuri, Aurantii Fructus Immaturus, Radix Paeoniae Alba,* and *Licorice* – may contribute distinctly to its therapeutic effects. Nonetheless, the exact mechanism by which SNP and its constituent components exert therapeutic benefits in UC has yet to be completely clarified. Network pharmacology, using public databases and accessible data, is an innovative, promising, and cost-effective method for identifying bioactive constituents, predicting drug action targets, and examining drug action processes through the lens of biological network equilibrium [10]. Moreover, in contrast to experimental medication approaches, network pharmacology prioritizes the multifaceted control of signaling pathways, making it particularly suitable for elucidating the mechanisms of conventional Chinese medicine, which has several chemical constituents and molecular targets [11].

In this study, we aimed to identify active constituents and their metabolites from selected medications using the Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP), based on oral bioavailability (OB) and drug similarity index. Then, the SNP-related differentially expressed genes (DEGs) of UC were screened by differential expression analyses using carefully selected public datasets. Subsequently, various machine learning algorithms were applied to identify key genes and develop a prediction model. The performance of the prediction model was validated using a nomogram and an external dataset. Finally, regulatory networks for transcription factor (TF) and micro-ribonucleic acid (RNA) – miRNA were built, along with a protein-compound network of hub genes. This will help future research on how these hub genes are controlled. While this computational approach provides valuable insights, experimental validation was still required to confirm the findings and assess their therapeutic relevance.

The aim of this study was to examine the potential mechanisms of SNP in UC treatment through a comprehensive bioinformatics approach that combines network pharmacology and machine learning, to identify and validate novel diagnostic biomarkers for UC, and to develop a robust prediction model for UC diagnosis.

Methods

Assessment of active constituents and target proteins

Chai Hu (CH, *Radix Bupleuri*), Zhishi (ZS, *Aurantii Fructus Immaturus*), Bai Shao (BS, *Paeoniae Radix Alba*), and Gan Cao (GC, *Licorice*) are included in SNP. The active constituents of these Chinese medications were identified using TCMSP (http://tcmspw.com/tcmsp) [10]. OB denotes the rate and extent of a drug's absorption into the systemic circulation. Drug-like (DL) features refer to the characteristics of a drug that has a certain functional category or exhibits analogous physical attributes. The drug half-life indicates the concentration of the medication in the bloodstream or body and is a crucial measure for determining the dosing interval, the delivered dosage, and drug accumulation [5]. The compounds exhibiting elevated action were further evaluated under the criteria of OB > 30%, DL > 0.18, Caco-2 permeability > -0.4, and half-life > 3 hrs, as previously documented [11]. The targeted proteins for each molecule were retrieved from the TCMSP database and standardized to a uniform gene nomenclature *via* the Universal Protein Resource (UniProt) protein database (http://www.uniprot.org/uploadlists/).

Collection and analysis of ulcerative colitis-related targets

To identify disease-specific targets, the GeneCards® database (https://www.genecards.org/) was used. It is a comprehensive platform that integrates information about all known human genes, including their roles in genome organization, protein expression, transcriptional regulation, inheritance patterns, and biological functions [12]. The database was queried specifically for UC-related targets using a systematic search strategy that incorporated multiple UC-associated keywords and synonyms to ensure comprehensive coverage of disease-relevant genes. The selection of UC-related targets followed a stringent filtering process based on relevance scores provided by the GeneCards® database. To increase the reliability of these results, emphasis should be placed on targets that have strong experimental evidence linking them to the development of UC. This approach allowed us to establish a high-confidence set of disease-specific targets for subsequent analysis. To identify potentially therapeutic targets of SNP in UC treatment, a systematic intersection analysis was performed between the UC-related targets and the previously identified SNP targets. This analysis was conducted using Venny 2.1.0, a precise tool for comparing multiple gene sets and identifying overlapping elements. The intersection analysis revealed targets that are both disease-relevant and potentially modulated by SNP components, suggesting possible therapeutic mechanisms. To visualize and analyze the complex relationships between the overlapping targets and their corresponding bioactive compounds, we employed Cytoscape 3.7.1. This powerful network analysis tool enabled us to construct and analyze interaction networks that illuminate the potential mechanisms of SNP in UC treatment. The visualization included both direct interactions and secondary connections, providing a comprehensive view of the potential therapeutic pathways.

Gene expression profiles and dataset selection

Our study utilized carefully selected public datasets from the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/). After a comprehensive review of available UC-related datasets, two bulk gene expression datasets were selected based on their sample size, data quality, and clinical annotation completeness: GSE87473 (106 UC, 21 normal) and GSE75214 (97 UC, 11 normal). These datasets were chosen specifically because they represent diverse patient populations and provide sufficient statistical power for robust analysis. The GSE87473 dataset was generated using the GPL13158 platform (HT_HG-U133_Plus_PM) Affymetrix HT HG-U133+ PM Array Plate, which offers comprehensive coverage of the human transcriptome. The GSE75214 dataset, based on the GPL6244 platform (HuGene-1_0-st) Affymetrix Human Gene 1.0 ST Array [transcript (gene) version], provides complementary data with different technical specifications, allowing for cross-platform validation of these findings.

Data preprocessing followed a rigorous protocol using the R package "GEOquery". The quality control process included several critical steps, as follows: a) probes successfully annotated with gene symbols; b) probes removed due to lacking gene symbols; c) probes removed because they matched multiple symbols; d) the number of unique gene symbols remains after resolving duplicates by selecting the maximum expression value.

To enhance the understanding of cellular heterogeneity in UC, single-cell RNA sequencing data were incorporated from GSE214695, which includes detailed transcriptional profiles from six UC rectum samples. This dataset provides crucial insights into cell-type-specific gene expression patterns and their potential roles in disease progression.

The integration of both bulk and single-cell RNA sequencing data provides complementary perspectives on UC pathogenesis, allowing for a more comprehensive analysis of disease mechanisms and potential therapeutic targets. While these publicly available datasets have inherent limitations, careful selection and preparation techniques improve the reliability of the results.

Differentially expressed gene analysis and functional analysis

Differential gene expression analysis was performed using the R Package Linear Models for Microarray Data (LIMMA), which employs robust linear modeling and empirical Bayes methods particularly suited for microarray data analysis. To ensure statistical rigor while maintaining biological relevance, a dual-threshold approach was implemented for identifying DEGs between UC and healthy controls in the GSE87473 dataset. Genes were considered significantly differentially expressed when meeting both statistical and biological significance criteria: FDR < 0.05 to control for multiple testing and an absolute $\log_2$ fold change ($|\log_2 FC| > 1$ to ensure biological meaningfulness of the expression differences. For a comprehensive functional interpretation of SNP-related DEGs, pathway and functional enrichment analyses were conducted using the Metascape database. This platform was selected for its integration of multiple authoritative databases, including Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome, and BioCarta. The enrichment analysis encompassed three major GO categories: biological process (BP), cellular component, and molecular function, alongside KEGG pathway analysis. Statistical significance was determined using an adjusted *p*-value threshold of 0.05, with correction for multiple testing using the Benjamini-Hochberg method. To ensure the robustness of the findings, functional categories and pathways were focused on at least three genes and required a minimum overlap of 10% between the gene set and each functional category. This comprehensive analytical approach allowed us to identify both statistically significant and biologically relevant gene expression changes while providing insights into the

functional implications of SNP-related genes in UC pathogenesis.

Screening and validation of diagnostic markers

A multi-algorithm machine learning approach was employed to identify robust diagnostic biomarkers for UC, incorporating three complementary methodologies to minimize algorithm-specific biases and enhance the reliability of the findings. The first approach utilized Random Forests, implemented through the "randomForest" R package, which can capture non-linear relationships to achieve optimal performance. Key parameters were tuned, including the number of trees (ntree) and the number of variables sampled at each split (mtry), allowing for the modeling of complex interactions between genes. The importance of each feature was assessed using both the mean decrease in accuracy and the mean decrease in the Gini index. The second method employed the Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression, implemented using the "glmnet" R package. This approach is particularly effective for handling high-dimensional data while preventing overfitting through L1 regularization. The optimal regularization parameter lambda (λ) was determined through 10-fold cross-validation, selecting the value that minimized partial likelihood deviance. The minimal λ value was specifically chosen to balance between model parsimony and predictive accuracy. The third approach utilized Support Vector Machine-Recursive Feature Elimination (SVM-RFE), which iteratively removes features based on their contribution to the classification boundary. The algorithm was implemented with a linear kernel and performed recursive feature elimination within a nested cross-validation framework to prevent selection bias. To ensure robust biomarker selection, attention was focused on genes that were consistently identified as significant across all three algorithms. This conservative approach helps mitigate algorithm-specific biases and increases the likelihood of identifying genuinely important biological signals rather than technical artifacts. For external validation, the independent dataset GSE75214 was utilized, which represents a different patient cohort and technical platform. The performance of the selected biomarkers was evaluated through multiple metrics: a) receiver operating characteristic (ROC) curves and corresponding area under the curve (AUC) values to assess discriminative ability; b) calibration curves to evaluate the agreement between predicted and observed probabilities; c) decision curve analysis to assess clinical utility across different threshold probabilities. Additionally, sensitivity analyses were performed to evaluate the robustness of the findings to various analytical choices, including different cross-validation schemes and parameter settings. The final nomogram was constructed based on the validated biomarkers and underwent rigorous calibration assessment. This comprehensive validation strategy addresses potential overfitting concerns and provides strong evidence for the generalizability of the findings across different patient populations and technical platforms.

Single-cell RNA sequencing analysis

To investigate cell-type-specific gene expression patterns in UC, comprehensive single-cell RNA sequencing analyses were performed using the Seurat R package version 4.0.2 (PMID: 33835452). The analysis pipeline incorporated stringent quality control measures to ensure data reliability while maximizing biological signal retention. Quality control thresholds were established based on the distribution of key technical metrics across all cells. Three-tier filtering approach was implemented: a) cell filtering based on RNA content (200–5,000 total RNAs) to exclude potential doublets (high RNA count) and empty droplets or dying cells (low RNA count); b) mitochondrial content filtering [excluding cells with > 10% mitochondrial unique molecular identifier (UMI) rate] to remove dying or stressed cells; c) gene filtering to remove mitochondrial genes, which can introduce technical artifacts due to variable cell stress levels.

Following quality control, data normalization was performed using the "NormalizeData" function in Seurat, employing a global-scaling normalization method that normalizes gene expression measurements for each cell by total expression and log-transforms the result. This approach accounts for differences in sequencing depth between cells while preserving biological variation. For dimensionality reduction, highly variable genes were identified using a variance-stabilizing transformation approach, selecting the top 2,000 genes that exhibited high cell-to-cell variation. Principal Component Analysis (PCA) was then performed on the scaled data of these variable genes. The optimal number of principal components (top 15) for downstream analysis was determined through multiple methods, including elbow plots and jackstraw analysis. Cell clustering was performed using the Uniform Manifold Approximation and Projection algorithm, which maintains both the local and global structure of the high-dimensional data. To ensure robust cluster identification, multiple resolution parameters and assessed cluster stability were employed using the Silhouette coefficient. Cell type annotation was conducted using the SingleR R package (PMID: 30643263), which leverages reference transcriptomic datasets to automatically annotate cell types. Manual curation of established cell-type-specific markers validated these annotations. Gene expression patterns were visualized using the "featureplot" function, with expression levels normalized and scaled for optimal visualization. This comprehensive single-cell analysis approach provides detailed insights into the cellular heterogeneity of UC tissue and the cell-type-specific expression patterns of the identified biomarkers.

Regulatory network construction and analysis

To elucidate the complex regulatory mechanisms underlying UC pathogenesis and the therapeutic effects of SNP, multiple regulatory networks were constructed using NetworkAnalyst 3.0 [(https://www.networkanalyst.ca/), (PMID:30931480)]. This comprehensive analysis encompassed three distinct but interconnected regulatory layers, each explored in detail below.

First, TF interactions with the characteristic genes were analyzed. TFs, as key regulators of gene expression, can bind to specific DNA sequences and modulate transcription. The analysis incorporated both direct TF-gene interactions and indirect regulatory relationships through intermediate molecules. To ensure biological relevance, interactions were filtered based on experimental evidence from curated databases. Second, miRNA-mediated regulation was investigated by constructing a diagnostic marker – miRNA coregulatory network. The miRNAs, as endogenous short non-coding RNAs, play crucial roles in post-transcriptional regulation through mRNA degradation or translation inhibition. Experimentally validated miRNA-target interactions were integrated from multiple databases to build a comprehensive regulatory network. Third, protein-drug interactions were examined to understand potential therapeutic mechanisms. The characteristic genes were analyzed in the context of known drug-target interactions, generating a diagnostic marker – drug/chemical coregulatory network. This analysis helps identify potential therapeutic compounds and predict drug sensitivity based on structural features.

Statistical analysis and data processing

All computational analyses were performed using the R programming language, version 4.1.1 (https://www.r-project.org/). A comprehensive statistical framework was implemented to ensure robust and reproducible results.

Group comparisons were conducted using non-parametric Wilcoxon tests, following an assessment of data distributions for normality. Results are presented as mean ± standard deviation, with statistical significance interpreted through multiple thresholds: not significant ($p \geq 0.05$); significant (*) ($p < 0.05$); highly significant (**) ($p < 0.01$); very highly significant (***) ($p < 0.001$); extremely significant (****) ($p < 0.0001$). For correlation analyses, Pearson correlation coefficients were calculated to assess relationships between continuous variables. The choice of Pearson correlation was based on preliminary assessments of linear relationships and data distributions. Statistical significance for correlations was set at $p < 0.05$. To control multiple testing where applicable, the Benjamini-Hochberg false discovery rate (FDR) was employed for correction. Effect sizes were calculated alongside *p*-values to provide a complete picture of the biological significance of the findings. All statistical tests were two-sided, and confidence intervals were set at 95%. Data visualization was performed using ggplot2 and other specialized R packages, with consistency in color schemes and plotting parameters maintained throughout the analysis.

Results

Identification and characterization of SNP active constituents and their targets

Through systematic analysis of the TCMSP database using multiple pharmacological parameters (OB, DL, Caco-2 permeability, and half-life), 95 active ingredients were identified in SNP that met the stringent selection criteria. These active ingredients were associated with 795 potential molecular targets, providing a comprehensive framework for understanding the therapeutic mechanisms of SNP. To identify disease-relevant targets, differential expression analysis was performed between UC and normal samples using the GSE87473 dataset. This analysis revealed 79 significantly DEGs that overlapped with SNP targets, suggesting potential therapeutic mechanisms. The differential expression patterns of these genes are visualized in a volcano plot (Figure 1A) and heatmap (Figure 1B), demonstrating a clear separation between UC and normal samples. To understand the contribution of individual components, the relationships were mapped between the four primary constituents of SNP (*Radix Bupleuri*, *Aurantii Fructus Immaturus*, *Radix Alba*, and *Licorice*) and their respective targets using Cytoscape visualization (Figure 1C). The expression profiles of all 79 SNP-related DEGs in GSE87473 are detailed in Figure 1D, revealing distinct patterns of regulation in UC compared to normal tissue. To elucidate the functional relationships among these targets, a protein-protein interaction network was constructed (Figure 1E). This network analysis revealed several highly connected nodes, suggesting key regulatory hubs in SNP's mechanism of action. The network structure indicates potential synergistic effects between different components of SNP in modulating UC-related pathways. Functional enrichment analysis using GO and KEGG pathways provided insights into the BPs affected by SNP-related DEGs. Key enriched pathways included regulation of inflammatory response, suggesting direct relevance to UC pathogenesis. Additional enriched processes encompassed circulatory system regulation, hormone response pathways, nutrient sensing, and smooth muscle cell proliferation (Figure 1F–G). These findings indicate that SNP may act through multiple complementary mechanisms to ameliorate UC symptoms.

This comprehensive analysis reveals that SNP's therapeutic effects likely arise from the coordinated action of multiple active ingredients targeting various disease-relevant pathways. The identification of specific molecular targets and pathways provides a foundation for understanding SNP's mechanism of action in UC treatment.

To identify robust diagnostic biomarkers for UC, a comprehensive machine-learning strategy was implemented, incorporating complementary algorithms. The SVM-RFE algorithm initially identified 50 candidate genes with potential diagnostic values (Figure 2A). In parallel, LASSO regression analysis, which excels at handling high-dimensional data while preventing overfitting, identified seven genes with strong predictive capabilities (Figure 2B–C).

Through intersection analysis of these independently identified gene sets, five robust core biomarkers were discovered: *TRPV1*, *ABCG2*, *BACE2*, *MMP3*, and *LIPC* (Figure 2D). Each of these genes demonstrated strong biological relevance to UC pathogenesis.

The diagnostic potential of these biomarkers was rigorously evaluated using ROC curve analysis. Each biomarker demonstrated strong discriminative ability with AUC values

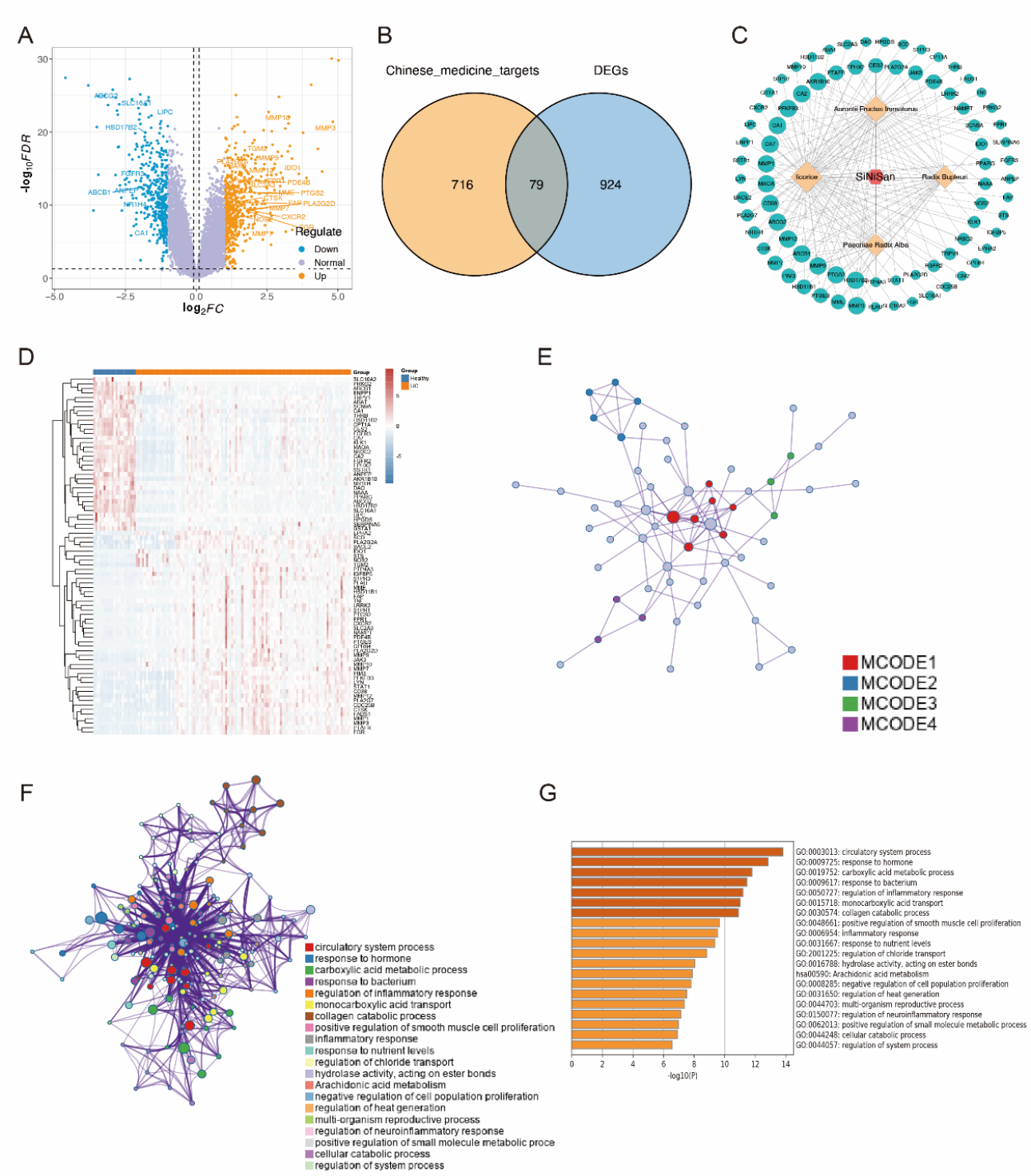

Fig. 1 – Identification and validation of hub genes using machine learning approaches: A) volcano plot displaying differentially expressed genes (DEGs) identified between ulcerative colitis (UC) and normal colon samples within the training dataset (GSE87473); B) intersection of Chinese medicine targets and DEGs; C) compound-target network illustrates the relationship between the four constituent herbs of SNP and the 79 overlapping genes network constructed using Cytoscape 3.7.1.; D) heatmap displaying the expression profiles of all 79 overlapping SNP-related DEGs across UC and normal samples from the GSE87473 dataset; E) protein-protein interaction network constructed for the 79 overlapping genes using the STRING database *via* Metascape and visualized in Cytoscape; F) bar chart illustrating the results of Gene Ontology Biological Process (BP) enrichment analysis for the 79 overlapping genes, performed using Metascape; G) bar chart illustrates the results of Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis for the 79 overlapping genes, performed using Metascape.
-log$_{10}$ FDR – negative logarithm base 10 of the false discovery rate; log$_2$ FC – logarithm base 2 fold change; MCODE – Molecular Complex Desection; log$_{10}$ (*p*) – negative base-10 logarithm of a *p*-value; STRING – Search Tool for the Retrieval of Interacting Genes/Proteins.

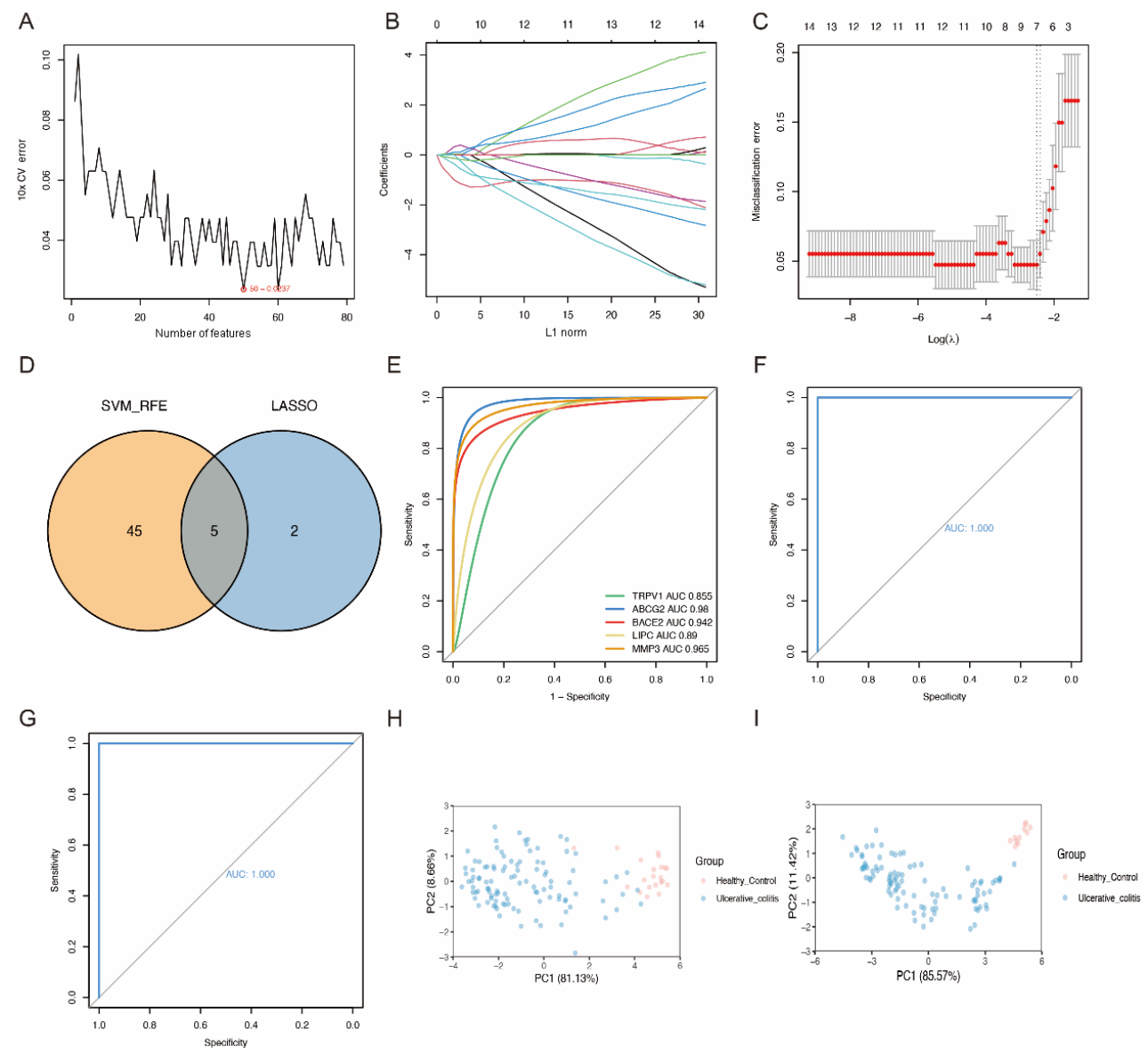

Fig. 2 – Identification and interaction network analysis of candidate diagnostic biomarkers: A) plot showing the cross-validation error curve as a function of the number of features considered during Support Vector Machine-Recursive Feature Elimination (SVM-RFE) analysis applied to the 79 overlapping genes in the training dataset (GSE87473); B) Least Absolute Shrinkage and Selection Operator (LASSO) regression coefficient profiles for the 79 overlapping genes in the training dataset (GSE87473). Each colored line represents the coefficient path for a single gene as the regularization parameter lambda (λ) changes; C) plot depicting the selection of the optimal tuning parameter in the LASSO model using 10-fold cross-validation. The λ value corresponding to the minimum partial likelihood deviances typically chosen; D) Venn diagram illustrates the intersection of candidate genes identified by LASSO regression (seven genes) and SVM-RFE (refined list). The overlap yields the final set of five robust core hub genes: *TRPV1*, *ABCG2*, *BACE2*, *MMP3*, and *LIPC*; E) ROC curves evaluating the diagnostic performance of each of the five individual hub genes in distinguishing UC from normal samples in the training dataset (GSE87473). The AUC value is provided for each gene, indicating its individual discriminative ability; F) ROC curve evaluating the performance of the diagnostic model built using the combined expression of the five hub genes in the training dataset; G) ROC curve evaluating the performance of the five-gene diagnostic model in the independent external validation dataset (GSE75214); H) Principal Component Analysis (PCA) plot based on the expression levels of the five hub genes in the training dataset (GSE87473). Blue dots represent normal samples, and red dots represent UC samples, demonstrating clear separation between the groups based on these genes; I) PCA plot based on the expression levels of the five hub genes in the validation dataset (GSE75214).

10 × CV error – 10-fold cross-validation error; L1 norm – LASSO regression normal; log – logarithm; AUC – area under the curve; ROC – receiver operating characteristic; PC – principal component; UC – ulcerative colitis.

exceeding 0.85 (Figure 2E), suggesting high sensitivity and specificity for UC detection. To enhance clinical applicability, a diagnostic nomogram was developed using the Rms package, which integrates the expression patterns of all five biomarkers.

Crucially, these findings were validated using an independent dataset. The diagnostic model maintained exceptional performance in both the training cohort (GSE87473) and the validation cohort (GSE75214), achieving perfect discrimination with AUC values of 1,000 in both datasets (Figures 2F–G). This remarkable consistency across different patient populations supports the robustness of the biomarker selection.

PCA further confirmed the diagnostic value of these biomarkers, revealing distinct clustering patterns between UC and normal samples in both GSE87473 and GSE75214 datasets (Figure 2H–I). This clear separation in gene expression patterns provides additional evidence for the biological relevance and diagnostic utility of the identified biomarkers.

These results demonstrate not only the statistical robustness of the machine-learning approach but also the biological significance of the identified biomarkers. The consistent performance across multiple validation steps suggests potential clinical utility for UC diagnosis and monitoring. However, further experimental validation was recognized as valuable for confirming the mechanistic roles of these biomarkers in UC pathogenesis.

To elucidate the complex regulatory mechanisms governing the candidate diagnostic biomarkers, comprehensive molecular interaction networks were constructed and analyzed. This multi-layer analysis encompassed miRNA regulation, TF control, and drug-target interactions, providing insights into potential therapeutic interventions.

Analysis of miRNA-mediated regulation revealed a complex regulatory network comprising 193 miRNAs interacting with the five potential biomarkers (Figure 3A). Of particular interest, three miRNAs emerged as potential master regulators: hsa-mir-1-3p, hsa-let-7b-5p, and hsa-mir-124-3p, each demonstrating the capacity to modulate multiple candidate diagnostic genes simultaneously. These findings suggest a coordinated post-transcriptional regulatory mechanism that may be crucial in UC pathogenesis.

TF analysis identified 32 TFs involved in regulating the candidate diagnostic genes (Figure 3B). Among these, GATA2 emerged as a particularly significant regulator, demonstrating potential simultaneous control over four of the five biomarkers: MMP3, TRPV1, ABCG2, and BACE2. This finding suggests that GATA2 might serve as a master regulator in the transcriptional control of UC-related genes and could represent a potential therapeutic target.

To explore potential therapeutic implications, drug-gene interactions were analyzed using the NetworkAnalyst database. This analysis revealed a comprehensive interaction network involving 369 potential target drugs/compounds and the five biomarkers (Figure 3C). Several compounds demonstrate notable interaction patterns with multiple biomarkers, including: a) benzo(a)pyrene, known for its effects on inflammatory pathways; b) cyclosporine, an established immunomodulator; c) estradiol, important in inflammatory response regulation; d) quercetin, a flavonoid with anti-inflammatory properties.

The interactions of these compounds with multiple biomarkers suggest potential therapeutic mechanisms, possibly explaining some of the observed clinical effects in UC treatment. The extensive network of drug-gene interactions provides a valuable resource for drug repurposing efforts and the development of novel therapeutic strategies.

The identification of these regulatory networks and potential therapeutic compounds offers new insights into UC pathogenesis and treatment. However, experimental validation would be necessary to confirm the functional significance of these regulatory relationships and the therapeutic potential of the identified compounds.

To understand the cellular context of the identified biomarkers, a comprehensive single-cell RNA sequencing analysis was performed using the UC dataset GSE214695.

Following quality control and data integration, 28,262 cells were analyzed from UC tissue samples. Stringent quality control measures were applied to ensure data reliability, including the removal of cells with aberrant RNA content (< 200 or > 5,000 total RNAs) and high mitochondrial content (> 10% UMI rate), which could indicate cellular stress or technical artifacts. Dimensionality reduction and clustering analysis, based on the top 2,000 highly variable genes and first 15 principal components, revealed 10 distinct major cell populations in the UC microenvironment (Figure 4A): a) adaptive immune system cells (B cells, $CD4^+$ T cells, $CD8^+$ T cells); b) innate immune system cells (monocytes, natural killer cells, dendritic cells, neutrophils, macrophages); c) structural cells (fibroblasts, epithelial cells).

Cell-cell interaction analysis revealed complex communication networks within the UC microenvironment (Figure 4B–C). B cells and epithelial cells emerged as major cellular hubs, suggesting their central role in disease pathogenesis. The interaction strength analysis highlighted particularly strong connections between immune system cells and epithelial cells, indicating active immune-epithelial crosstalk in UC tissue. Expression analysis of these diagnostic biomarkers revealed distinct cell-type-specific patterns (Figure 4D): a) *ABCG2* and *BACE2* showed the highest expression in epithelial cells, suggesting their involvement in epithelial barrier function and homeostasis; b) *MMP3* was predominantly expressed in fibroblasts, consistent with its role in tissue remodeling and extracellular matrix modification.

The distribution of diagnostic biomarkers across various cell types (Figure 4E) provides crucial insights into their functional roles in UC pathogenesis. This cell-type-specific expression pattern suggests that the identified biomarkers may be involved in multiple aspects of UC pathophysiology, from epithelial barrier dysfunction to immune response regulation.

These single-cell findings not only validate the biomarker selection but also provide important context for understanding their roles in UC pathogenesis. The cell-type-specific expression patterns suggest potential cellular mechanisms through which SNP might exert its therapeutic effects.

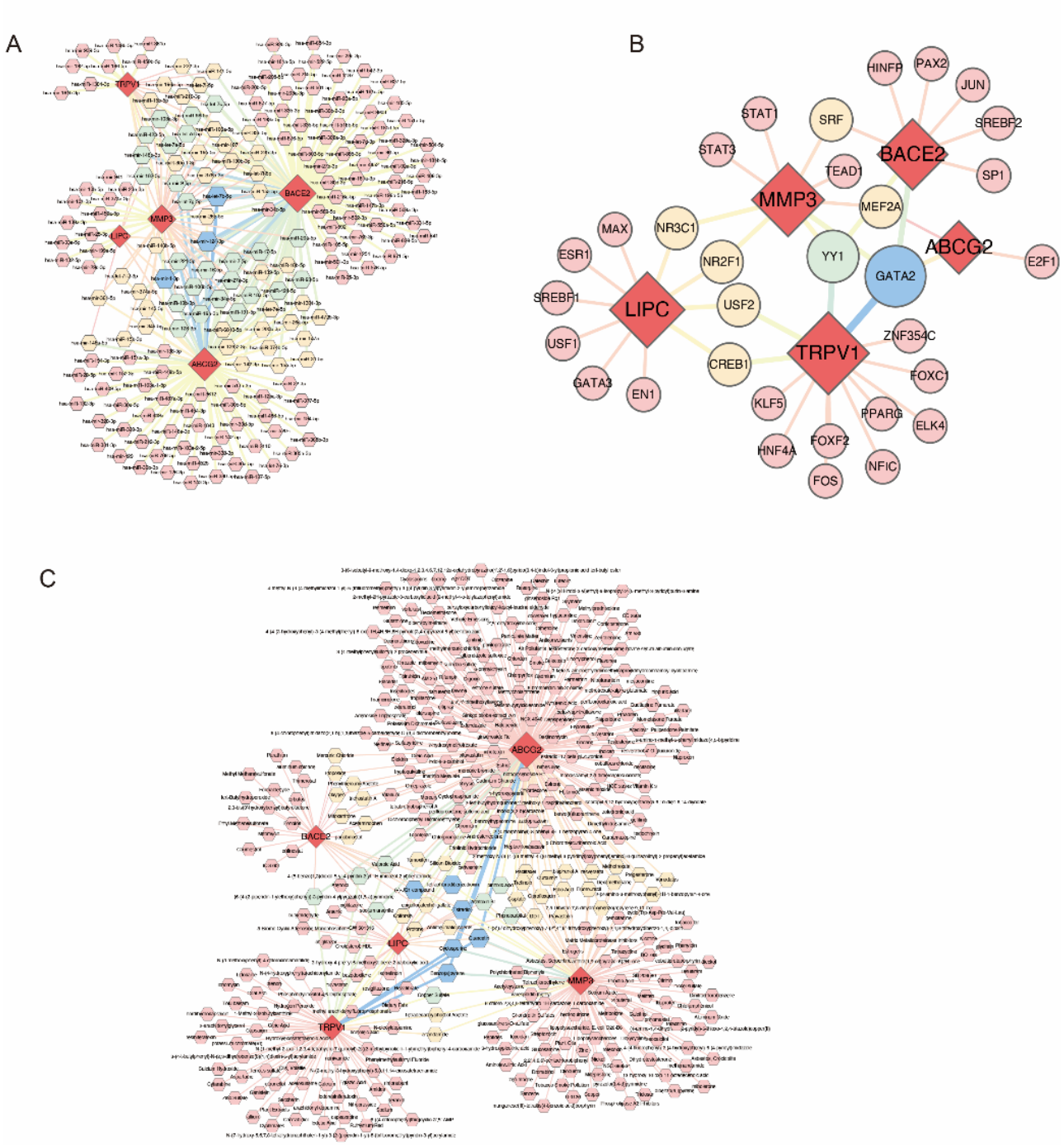

Fig. 3 – Cell-type-specific expression analysis using single-cell transcriptomics: A) predicted micro-ribonucleic acid (miRNA)-hub gene regulatory network. Larger nodes represent the five hub genes, while smaller nodes represent the 193 potentially interacting miRNAs identified from databases. Edges indicate predicted miRNA-target gene interactions; B) predicted transcription factor (TF)-hub gene regulatory network. Larger nodes represent the five hub genes, while smaller nodes represent the 32 potentially interacting TFs identified from databases such as JASPAR. Edges indicate predicted regulatory relationships (e.g., TF binding to promoter regions); C) predicted drug/compound-hub gene interaction network. Larger nodes represent the five hub genes, while smaller nodes represent 369 potentially interacting drugs or chemical compounds identified from databases such as DrugBank. Edges indicate known or predicted interactions between the compounds and the hub gene proteins.

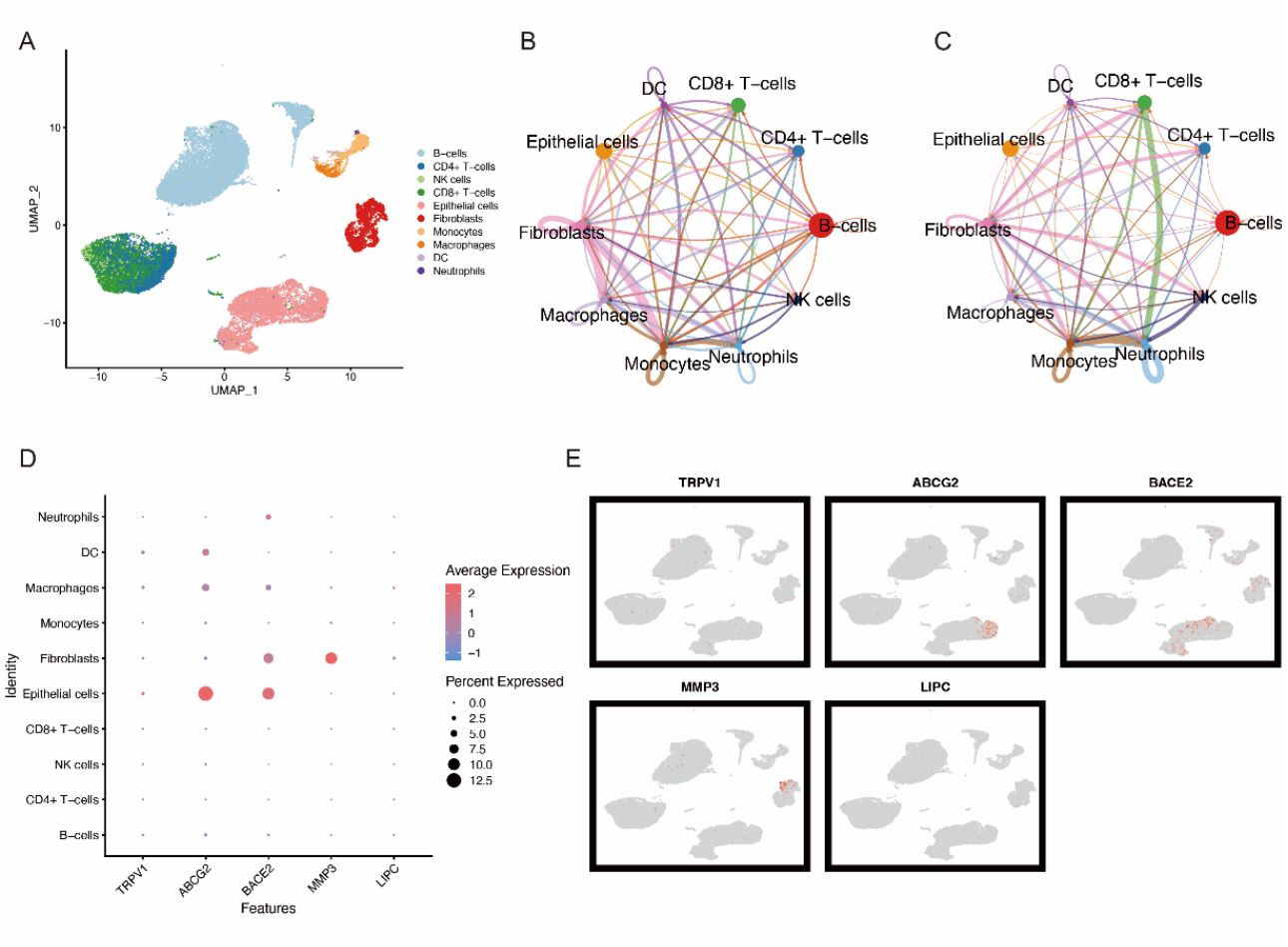

Fig. 4 – Single-cell transcriptome analysis of ulcerative colitis (UC)-related diagnostic biomarkers and cell-type interactions: A) Uniform Manifold Approximation and Projection (UMAP) plot visualizing the cellular landscape of UC rectum tissue. Each point represents a single cell, colored according to its assigned cell type identity based on marker gene expression; B) circle plots summarize the overall predicted cell-cell communication network among the identified cell types. The thickness or color intensity of lines connecting cell types indicates the relative strength or number of inferred interactions; C) network plot detailing the strength of predicted interactions for specific signaling pathways or ligand-receptor pairs between different cell types; D) dot plots show the expression levels of individual hub genes; E) UMAP plots summarize the normalized expression distribution of all five hub genes across the major annotated cell types.
NK – natural killer; DC – dendritic cells.

Discussion

UC represents a significant challenge in gastroenterology, primarily affecting adults aged 30–40 years with substantial impacts on quality of life and work productivity [12, 13]. While considerable progress has been made in understanding UC, both diagnosis and management remain challenging. Currently, diagnosis relies heavily on invasive colonoscopy procedures, highlighting the urgent need for reliable, non-invasive diagnostic biomarkers [14]. Moreover, existing treatment options often show variable efficacy and may be associated with significant side effects, creating a clear need for alternative therapeutic approaches.

In this context, the study investigated SNP, a TCM formulation with documented anti-inflammatory properties and effectiveness in treating inflammatory and bowel diseases [15]. Combining network pharmacology with machine learning approaches aimed to address three critical gaps in current knowledge: the identification of novel diagnostic biomarkers, the elucidation of SNP's mechanism of action, and the potential for personalized treatment approaches based on molecular profiles.

However, several important limitations of this approach were acknowledged. First, these findings are primarily based on computational analyses of public datasets, and while statistically robust, they require experimental validation. Second, the gene expression datasets used, though carefully selected, may not fully represent the diversity of UC patient populations. Third, while the machine learning approach identified promising biomarkers, their clinical utility needs to be validated in prospective studies.

The network pharmacology analysis revealed important insights into the mechanism of action of SNP. A detailed interaction network was created by carefully studying active

parts and the possible targets they could interact with. This network shows how SNP parts are connected and how they affect biology. This analysis identified 79 genes that were both targeted by SNP and differentially expressed between healthy and UC samples in the GEO database, suggesting potential therapeutic mechanisms.

The functional enrichment analysis through GO and KEGG pathways revealed several key BPs affected by these 79 SNP-related genes. Of particular significance was the enrichment of inflammatory response pathways, which aligns with the known pathophysiology of UC [16]. The enriched pathways included: regulation of inflammatory response, immune system modulation, cellular stress response, tissue repair and regeneration, and mucosal barrier function.

However, these computational findings are promising and acknowledge that they represent predictions that require experimental validation. The complex nature of both UC pathogenesis and TCM formulations means that additional mechanisms may exist beyond those identified in the analysis. Future studies should focus on experimental verification of these predicted pathways, particularly the specific roles of individual SNP components in modulating inflammatory responses.

Additionally, the interaction between different components of SNP (*Radix Bupleuri*, *Aurantii Fructus Immaturus*, *Radix Alba*, and *Licorice*) may produce synergistic effects that are not fully captured by the current analysis. Understanding these synergistic interactions could be crucial for optimizing therapeutic approaches.

Through the application of complementary machine learning approaches (SVM-RFE and LASSO) to the set of 79 DEGs, five robust core biomarkers were identified with potential diagnostic significance: *TRPV1*, *ABCG2*, *BACE2*, *MMP3*, and *LIPC* [17–21]. Each of these biomarkers demonstrates distinct biological roles and potential mechanistic connections to UC pathogenesis.

TRPV1 functions as a polymodal sensory transducer, responding to various stimuli, including heat, acidic pH, and both endogenous and exogenous inflammatory mediators [17]. Its role in UC is particularly interesting, as evidenced by research demonstrating its involvement in the protective effects of propolis on colonic tissue in UC patients [18]. The sensitivity of TRPV1 to multiple inflammatory signals makes it a potentially valuable diagnostic indicator of disease activity.

ABCG2, an important membrane transporter, has emerged as a key player in various cellular stress responses, including inflammation, tissue damage, and hypoxia [19, 20]. While its precise inhibitory mechanism in UC remains to be fully elucidated, studies have demonstrated reduced *ABCG2* expression in UC patients, with this reduction showing a negative correlation with specific functional miRNAs [21]. This relationship suggests a potential regulatory network that could be therapeutically targeted.

BACE2's significance in UC pathogenesis is highlighted by its regulation through the JAK2/STAT5 signaling pathway [22]. Its influence on interleukin (IL)-1R2 activity provides a mechanistic link to UC development, suggesting potential therapeutic applications [23]. The connection to established inflammatory pathways strengthens its potential as both a diagnostic marker and therapeutic target.

MMP3, a member of the zinc-dependent endopeptidase family, demonstrates particularly promising clinical utility. Its expression spans multiple cell types relevant to UC pathogenesis, including immune cells, connective tissues, and endothelial cells. The direct correlation between serum MMP3 levels and disease activity in pediatric UC patients, as demonstrated by Kofla-Dłubacz et al. [21], suggests its potential as a practical biomarker for monitoring disease progression.

The *LIPC* gene, encoding a 65-kilodalton (kD) glycoprotein, presents an intriguing connection between metabolic regulation and UC pathogenesis. Its association with metabolic and circulatory system disorders suggests potential mechanisms linking systemic metabolism to intestinal inflammation [24, 25]. The gene's structure, spanning 35 kilobases (kb) with nine exons and eight introns, provides multiple potential regulatory points that could be targeted therapeutically.

This multi-biomarker panel represents distinct but potentially interconnected aspects of UC pathophysiology, from inflammatory signaling (TRPV1) to tissue remodeling (MMP3) and metabolic regulation (LIPC). However, these experimental validations acknowledged that the utility of these biomarkers in clinical settings is essential. Future studies should focus on validating these markers in larger, diverse patient cohorts, investigating potential interactions between these biomarkers, developing practical diagnostic assays, and exploring their potential as therapeutic targets.

The current landscape of UC diagnosis presents significant challenges, primarily due to the absence of reliable, non-invasive biomarkers and robust predictive models [26, 27]. Although traditional diagnostic methods such as colonoscopy remain the gold standard, they are invasive and may fail to capture the molecular complexity of the disease. Recent advances in machine learning approaches have opened new avenues for improving UC diagnosis and prediction [28], with techniques such as logistic regression demonstrating particular promise in analyzing complex biological datasets [29].

In this study, a sophisticated dual-algorithm approach was employed, combining LASSO and SVM-RFE methodologies to develop a diagnostic model based on the identified key genes. The model's performance was remarkable, achieving perfect discrimination with AUC values of 1,000 in both the training and validation cohorts across different algorithms. While these results are encouraging, they must be interpreted within the context of several important considerations.

The performance metrics, while statistically impressive, highlight the need for broader validation. The analysis, based on publicly available datasets, may not fully capture the heterogeneity of UC presentations across different patient populations. Furthermore, the perfect discrimination observed in the model, while mathematically sound, suggests the need for testing in larger, more diverse cohorts to ensure generalizability in real-world clinical settings.

A critical limitation of the current study is the absence of experimental validation. The molecular mechanisms predicted by computational analyses require verification through carefully designed cell-based experiments and animal models. Such validation would not only confirm the biological relevance of the identified biomarkers but also provide insights into their potential as therapeutic targets. Although supported by computational analysis, the relationship between SNP components and these molecular targets likewise requires experimental confirmation.

The translation of these findings into clinical practice presents additional challenges. The relationship between gene expression patterns and protein levels needs to be established, and the stability of these markers across different technical platforms must be verified. Moreover, the development of practical, cost-effective diagnostic assays based on these markers requires significant additional research and validation.

The findings also raise important questions about the mechanism of action of SNP in UC treatment. While the computational approach suggests specific molecular targets and pathways, the complex interactions between different components of this traditional medicine warrant further investigation. Understanding these interactions could provide valuable insights for optimizing treatment strategies and potentially developing more targeted therapeutic approaches.

While our integrated approach provided insights, particularly the single-cell analysis, which offered valuable cellular context for the identified biomarkers (e.g., ABCG2/BACE2 in epithelial cells, MMP3 in fibroblasts), we recognize its current scope has limitations. Further in-depth single-cell analyses, such as trajectory inference, detailed cell-cell communication focusing specifically on the hub genes, and the construction of cell-type-specific regulatory networks, are warranted in future studies to fully elucidate the mechanistic roles of these genes within distinct UC cell populations. Our current findings provide a solid foundation for these necessary future investigations. Similarly, the diagnostic model built using the five hub genes requires careful interpretation. While achieving perfect AUC scores of 1.000 in both training and validation datasets mathematically indicates strong separation within these specific cohorts, we absolutely agree that such performance necessitates extreme caution. Perfect scores in complex biological systems raise significant concerns about potential overfitting, even with external validation, or suggest the model might be highly specific to characteristics shared by samples within these particular datasets (GSE87473, GSE75214). Consequently, the selected genes, while highly discriminatory here, may not generalize perfectly to the broader spectrum of UC heterogeneity found in larger, more diverse real-world populations. Although methodological diligence was applied, including careful control to prevent data leakage and appropriate use of cross-validation during feature selection (e.g., for LASSO lambda tuning), the possibility of subtle biases persisting in retrospective data analysis cannot be entirely ruled out. Therefore, as acknowledged throughout our discussion and limitations, these findings, particularly the remarkable AUC values, demand rigorous external confirmation. We strongly emphasize the critical need for validation in larger, prospective, multi-center clinical cohorts encompassing diverse patient characteristics before any conclusions about the model's real-world clinical utility can be drawn. Furthermore, regarding comparison with other state-of-the-art methods, we acknowledge that this study did not perform a direct quantitative benchmarking of our five-gene signature against other previously published UC diagnostic models, representing another limitation and an avenue for future comparative studies.

Moving forward, these limitations and considerations point to several crucial directions for future research. Large-scale validation studies in diverse patient populations will be essential for confirming the clinical utility of the identified biomarkers. Experimental studies focusing on molecular mechanisms will help establish the biological basis for computational predictions. Finally, the development of practical diagnostic tools based on these findings will require careful consideration of both technical and clinical implementation challenges.

The study has identified a signature of five hub genes (*TRPV1*, *ABCG2*, *BACE2*, *MMP3*, and *LIPC*) that demonstrate a significant correlation with SNP's therapeutic effects in UC. Not only do these genes serve as potential diagnostic biomarkers, but they also provide insights into the molecular mechanisms underlying both UC pathogenesis and SNP's therapeutic action.

These hub genes represent diverse BPs relevant to UC pathophysiology. *TRPV1*'s role in inflammatory signaling and pain sensation suggests its involvement in both symptom manifestation and disease progression [17, 18]. *ABCG2*'s function in cellular stress response and barrier protection points to its potential role in maintaining intestinal homeostasis [19, 20]. *BACE2*'s connection to the JAK2/STAT5 pathway highlights the involvement of key inflammatory signaling cascades [22, 23]. *MMP3*'s role in tissue remodeling and its correlation with disease activity make it particularly promising as a monitoring biomarker [21]. *LIPC*'s involvement in metabolic regulation suggests a previously underappreciated connection between metabolic processes and UC pathogenesis [24, 25].

The integration of these markers into a diagnostic model represents a step toward more precise and personalized approaches to UC management. However, the translation of these findings into clinical practice will require careful validation through experimental studies and clinical trials. Understanding the specific mechanisms by which SNP modulates these genes could lead to optimized treatment strategies and potentially the development of new therapeutic approaches.

Furthermore, these findings lay the groundwork for future investigations into both traditional medicine mechanisms and novel therapeutic strategies for UC. The identification of these molecular targets may facilitate the development of more targeted treatments and provide a scientific basis for understanding the multi-component, multi-target nature of TCM formulations such as SNP.

This study thus bridges traditional medicine and modern molecular biology, offering new perspectives on UC pathogenesis and treatment while highlighting the potential for integrating traditional and modern therapeutic approaches.

Limitations of the study

This study possesses several significant limitations inherent to its computational design and reliance on publicly available data. Foremost among these is the complete lack of experimental validation, which is the most critical constraint. All findings regarding SNP's potential mechanisms, the identified hub genes, and the diagnostic model's performance are derived solely from bioinformatics analyses and predictions. Confirmation through dedicated *in vitro*, *in vivo*, and prospective clinical studies is essential to establish the biological relevance, functional roles, diagnostic accuracy, and therapeutic potential of these findings. Furthermore, the study's conclusions depend heavily on the accuracy and completeness of the public databases utilized (e.g., TCMSP, GeneCards®, NetworkAnalyst resources) and the representativeness of the specific GEO datasets chosen. These datasets may possess inherent limitations, including data quality issues, technical variations, limited clinical annotation depth, and they may not fully capture the broad heterogeneity of the global UC patient population, thereby potentially limiting the generalizability of our results. While the achievement of perfect AUC scores (1,000) is mathematically noteworthy, it must be interpreted with caution, as it raises concerns about potential overfitting or model specificity, highlighting the urgent need for validation in diverse and independent cohorts. Additionally, the inherent complexity of SNP as a multi-component formula makes it challenging to computationally dissect synergistic effects or attribute outcomes definitively, and real-world variability is not captured. The depth of the single-cell analysis was also limited, providing cellular context but lacking deeper functional investigation. Finally, it is crucial to reiterate that this study identifies correlations and associations, not causal relationships, which can only be established through experimental work. These limitations collectively underscore that the study provides hypotheses and potential leads rather than confirmed mechanisms or clinically actionable tools.

Conclusion

This study bridges TCM and modern molecular biology, offering new perspectives on UC pathogenesis and SNP's therapeutic mechanisms. The identified biomarkers and regulatory networks provide valuable foundations for future experimental research and potential clinical applications. However, translation of these findings into clinical practice will require rigorous experimental validation, prospective clinical studies, and development of practical diagnostic assays.

Acknowledgement

We wish to express our gratitude to all research contributors who were not listed as co-authors for their valuable contributions to this study.

Conflicts of interest

The authors declare no conflict of interest.

Funding

This study was supported by the Liaoning Provincial Talent Project (No. XLYC2002002) and explores the mechanism of brain-gut interaction by UC-mediated 5-HT signaling system based on microbial metabolomics.

REFERENCES

1. *Hendrickson BA, Gokhale R, Cho JH.* Clinical aspects and pathophysiology of inflammatory bowel disease. Clin Microbiol Rev 2002; 15(1): 79–94.
2. *Levine JS, Burakoff R.* Extraintestinal manifestations of inflammatory bowel disease. Gastroenterol Hepatol (N Y) 2011; 7(4): 235–41.
3. *Gordon H, Burisch J, Ellul P, Karmiris K, Katsanos K, Allocca M,* et al. ECCO Guidelines on Extraintestinal Manifestations in Inflammatory Bowel Disease. J Crohns Colitis 2024; 18(1): 1–37.
4. *Fanizzi F, Allocca M, Fiorino G, Zilli A, Furfaro F, Parigi TL,* et al. Raising the bar in ulcerative colitis management. Therap Adv Gastroenterol 2024; 17: 17562848241273066.
5. *Zeng Y, Zhang JW, Yang J.* Optimal traditional Chinese medicine formulas in treating ulcerative colitis: Choose one or take it all? World J Clin Cases 2024; 12(32): 6570–4.
6. *Wang M, Fu R, Xu D, Chen Y, Yue S, Zhang S,* et al. Traditional Chinese Medicine: A promising strategy to regulate the imbalance of bacterial flora, impaired intestinal barrier and immune function attributed to ulcerative colitis through intestinal microecology. J Ethnopharmacol 2024; 318(Pt A): 116879.
7. *Qi Q, Liu YN, Jin XM, Zhang LS, Wang C, Bao CH,* et al. Moxibustion treatment modulates the gut microbiota and immune function in a dextran sulphate sodium-induced colitis rat model. World J Gastroenterol 2018; 24(28): 3130–44.
8. *Li L, Yang L, Yang L, He C, He Y, Chen L,* et al. Network pharmacology: a bright guiding light on the way to explore the personalized precise medication of traditional Chinese medicine. Chin Med 2023; 18(1): 146.
9. *Feng W, Zhu L, Shen H.* Traditional Chinese Medicine Alleviates Ulcerative Colitis via Modulating Gut Microbiota. Evid Based Complement Alternat Med 2022; 2022: 8075344.
10. *Liang Y, Li Y, Lee C, Yu Z, Chen C, Liang C.* Ulcerative colitis: molecular insights and intervention therapy. Mol Biomed 2024; 5(1): 42.
11. *Noor F, Tahir ul Qamar M, Ashfaq UA, Albutti A, Alwashmi ASS, Aljasir MA.* Network Pharmacology Approach for Medicinal Plants: Review and Assessment. Pharmaceuticals 2022; 15(5): 572.
12. *Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T,* et al. GeneHancer: genome-wide integration of enhancers

and target genes in GeneCards. Database (Oxford) 2017; 2017: bax028.
13. *Nasr S, Dahmani W, Jaziri H, Hammami A, Slama AB, Ameur WB,* et al. Exploring work productivity loss in patients with inflammatory bowel disease. Future Sci OA 2023; 9(8): FSO872.
14. *Liu S, Eisenstein S*. State-of-the-art surgery for ulcerative colitis. Langenbecks Arch Surg 2021; 406(6): 1751–61.
15. *Li S, Huang M, Wu G, Huang W, Huang Z, Yang X,* et al. Efficacy of Chinese Herbal Formula Sini Zuojin Decoction in Treating Gastroesophageal Reflux Disease: Clinical Evidence and Potential Mechanisms. Front Pharmacol 2020; 11: 76.
16. *Saez A, Herrero-Fernandez B, Gomez-Bris R, Sánchez-Martinez H, Gonzalez-Granado JM*. Pathophysiology of Inflammatory Bowel Disease: Innate Immune System. Int J Mol Sci 2023; 24(2): 1526.
17. *Csekő K, Beckers B, Keszthelyi D, Helyes Z*. Role of TRPV1 and TRPA1 Ion Channels in Inflammatory Bowel Diseases: Potential Therapeutic Targets? Pharmaceuticals (Basel) 2019; 12(2): 48.
18. *Quaglio AEV, Santaella FJ, Rodrigues MAM, Sassaki LY, Di Stasi LC*. MicroRNAs expression influence in ulcerative colitis and Crohn's disease: A pilot study for the identification of diagnostic biomarkers. World J Gastroenterol 2021; 27(45): 7801–12.
19. *Wu F, Zikusoka M, Trindade A, Dassopoulos T, Harris ML, Bayless TM,* et al. MicroRNAs are differentially expressed in ulcerative colitis and alter expression of macrophage inflammatory peptide-2 alpha. Gastroenterology 2008; 135(5): 1624–35.e24.
20. *Vassar R, Kuhn PH, Haass C, Kennedy ME, Rajendran L, Wong PC,* et al. Function, therapeutic potential and cell biology of BACE proteases: current status and future prospects. J Neurochem 2014; 130(1): 4–28.
21. *Kofla-Dłubacz A, Matusiewicz M, Krzesiek E, Noga L, Iwańczak B*. Metalloproteinase-3 and -9 as novel markers in the evaluation of ulcerative colitis activity in children. Adv Clin Exp Med 2014; 23(1): 103–10.
22. *Tian M, Qi Y, Zhang X, Wu Z, Chen J, Chen F*, et al. Regulation of the JAK2-STAT5 Pathway by Signaling Molecules in the Mammary Gland. Front Cell Dev Biol 2020; 8: 604896.
23. *Mora-Buch R, Dotti I, Planell N, Calderón-Gómez E, Jung P, Masamunt MC,* et al. Epithelial IL-1R2 acts as a homeostatic regulator during remission of ulcerative colitis. Mucosal Immunol 2016; 9(4): 950–9.
24. *Adolph TE, Meyer M, Jukic A, Tilg H*. Heavy arch: from inflammatory bowel diseases to metabolic disorders. Gut 2024; 73(8): 1376–87.
25. *Saltiel AR, Olefsky JM*. Inflammatory mechanisms linking obesity and metabolic disease. J Clin Invest 2017; 127(1): 1–4.
26. *Zhang S, Zhang G, Wang W, Guo SB, Zhang P, Wang F,* et al. An assessment system for clinical and biological interpretability in ulcerative colitis. Aging (Albany NY) 2024; 16(4): 3856–79.
27. *Vălean D, Zaharie R, Ţaulean R, Usatiuc L, Zaharie F*. Recent Trends in Non-Invasive Methods of Diagnosis and Evaluation of Inflammatory Bowel Disease: A Short Review. Int J Mol Sci 2024; 25(4): 2077.
28. *Kulkarni C, Liu D, Fardeen T, Dickson ER, Jang H, Sinha SR,* et al. Artificial intelligence and machine learning technologies in ulcerative colitis. Therap Adv Gastroenterol 2024; 17: 17562848241272001.
29. *Yousef M, Allmer J*. Deep learning in bioinformatics. Turk J Biol 2023; 47(6): 366–82.

Received on January 31, 2025
Revised on May 5, 2025
Revised on June 4, 2025
Accepted on June 11, 2025
Online First July 2025