

Isolation of polymethoxylated flavonoids from *Clerodendranthus spicatus* and an integrative network pharmacology study on their effect against ulcerative colitis

Xuejin Wang, Wei Li*

Department of Functional Food and Wine, Shenyang pharmaceutical university, Shenyang 110016, China

Abstract Ulcerative colitis (UC) is a chronic inflammatory bowel disease with limited therapeutic options. *Clerodendranthus spicatus*, a traditional medicinal plant, has demonstrated anti-inflammatory properties, yet its bioactive compounds and mechanisms in UC remain unclear. Five polymethoxylated flavonoids (compounds **1-5**) were isolated from *Clerodendranthus spicatus*, with three reported for the first time of this plant. An integrative approach combining network pharmacology, transcriptomics (GSE243625, GSE66407, GSE206171), weighted gene co-expression network analysis (WGCNA), machine learning (LASSO and SVM-RFE), gene set enrichment analysis (GSEA), and molecular docking was employed to elucidate therapeutic mechanisms. Network pharmacology revealed that the five compounds converged on the IL-17 and TNF- α signaling pathways. Machine learning identified five core inflammation-related genes (NPTX2, SIGLEC12, SPINK4, SOX12, ZNF575), which were validated across independent cohorts (combined AUC > 0.81). GSEA confirmed that all five genes were embedded in cytokine/JAK-STAT, TLR/NF- κ B, and TNF signaling networks. Molecular docking demonstrated favorable binding affinities (−8.0 to −4.8 kcal/mol) between the flavonoids and core targets. The five *Clerodendranthus spicatus* flavonoids exert anti-UC effects through dual mechanisms: modulating IL-17 and TNF- α pathways at the systems level and targeting core inflammatory genes at the molecular level, providing a scientific basis for developing natural product-derived UC therapeutics.

Keywords: Ulcerative colitis; *Clerodendranthus spicatus*; Network pharmacology; Machine learning

1 Introduction

Ulcerative colitis (UC) was a chronic relapsing inflammatory bowel disease characterized by continuous mucosal inflammation of the colon and rectum. This condition showed a rising global incidence and imposed a substantial burden on quality of life. The pathogenesis of UC proved multifactorial, as it involved genetic predisposition, environmental triggers, gut microbiota dysbiosis, and aberrant immune responses^[1]. Among the key molecular drivers, the interleukin-17 (IL-

17) and tumor necrosis factor- α (TNF- α) signaling pathways were identified as central orchestrators of intestinal inflammation in UC^[2]. IL-17, primarily secreted by Th17 cells, promoted neutrophil recruitment and induced pro-inflammatory cytokines such as IL-6 and IL-8. Additionally, it disrupted epithelial barrier integrity. In contrast, TNF- α amplified NF- κ B-dependent inflammatory gene expression and sustained chronic mucosal damage^[3]. Current therapeutic strategies included aminosalicylates, corticosteroids, immunosuppressants, and biologics targeting TNF- α or integrins. Although these treatments provided symptomatic relief for many patients, a substantial

* Corresponding author: Li Wei (wei_li1981@163.com).

These authors have no conflict of interest to declare.

proportion exhibited primary non-response, secondary loss of response, or intolerable adverse effects ^[4]. Furthermore, the complexity and heterogeneity of UC pathogenesis limited the efficacy of single-target therapies ^[5]. Consequently, this highlighted an urgent clinical need for novel multi-target agents with improved safety profiles.

Natural products, particularly plant-derived polyphenols, garnered increasing attention as promising sources of multi-target anti-inflammatory compounds. This interest stemmed from their ability to modulate multiple signaling pathways simultaneously while maintaining relatively low toxicity. *Clerodendranthus spicatus* (*C. spicatus*) commonly known as kidney tea plant or “Shen Cha”, represented a traditional medicinal herb widely distributed across Southeast Asia. For centuries, practitioners used this herb to treat urinary disorders, edema, nephritis, and various inflammatory conditions ^[6].

Phytochemical studies revealed that *C. spicatus* was rich in bioactive constituents. These included polymethoxylated flavonoids, diterpenoids, triterpenes, and phenolic acids. Among these components, the polymethoxylated flavonoids proved particularly noteworthy for their potent anti-inflammatory activities. These compounds were characterized by multiple methoxy substitutions on the flavone, flavanone, or chalcone scaffolds. Pharmacological investigations demonstrated that *C. spicatus* extracts exhibited significant anti-inflammatory effects through several mechanisms ^[7]. Moreover, *C. spicatus* showed beneficial effects in ameliorating gouty arthritis, protecting renal function, and modulating immune responses. These findings suggested its potential as a multi-target therapeutic agent for chronic inflammatory diseases ^[8].

In this study, we isolated and identified five polymethoxylated flavonoids (compounds **1-5**) from *C. spicatus*. Notably, three of these compounds were reported from this species for the first time of this plant. To elucidate their therapeutic mechanisms in UC, we employed an integrative approach that combined network pharmacology, transcriptomics, machine learning, functional enrichment analysis,

and molecular docking. Our objectives were twofold. First, we investigated whether these flavonoids converged on key inflammatory pathways implicated in UC pathogenesis at the systems level. Second, we identified and validated core inflammation-related genes that served as molecular targets for these compounds. This multi-level framework aimed to provide mechanistic insights into how *C. spicatus* polyphenols might ameliorate UC. Ultimately, it offered a scientific foundation for the development of natural product-derived therapeutics targeting inflammatory bowel disease.

2 Materials and methods

2.1 Isolation of flavonoids (1-5) from *C. spicatus*

Dried whole herbs of *C. spicatus* (19 kg) were extracted with 70% ethanol (5×, w/v). Each batch was macerated three times, with each maceration lasting 4 h. The combined extracts were concentrated under reduced pressure to recover ethanol, thus affording a 70% ethanol extract (4014.20 g). Subsequently, the extract was dissolved in water and successively partitioned with petroleum ether, ethyl acetate, and n-butanol. Each solvent extraction was repeated three times using equal volumes. The ethyl acetate fraction (250.72 g) was subjected to silica gel column chromatography and eluted with a petroleum ether-ethyl acetate gradient, yielding 30 fractions (Fr. 1-30).

Fr. 9 was further separated by ODS column chromatography using a MeOH-H₂O gradient; this process yielded Fr. e9-1 and Fr. e9-2 at 90% MeOH. Fr. e9-1 was then purified by semi-preparative HPLC with isocratic 90% MeOH-H₂O, affording compound **5** (3.1 mg, *t_R* 9 min) and compound **4** (5.3 mg, *t_R* 11 min). Fr. 16 was separated on an ODS column with a MeOH-H₂O gradient, producing Fr. e16-1 (20% MeOH), Fr. e16-2 (90% MeOH), and Fr. e16-3 (100% MeOH). Thereafter, Fr. e16-2 was purified by semi-preparative HPLC using isocratic 85% MeOH-H₂O, thus obtaining compound **1** (4.5 mg, *t_R* 12 min).

Fr. 18 was chromatographed on an ODS column with a MeOH-H₂O gradient, affording Fr. e18-1 (90%

MeOH). This subfraction was subsequently purified by semi-preparative HPLC (MeOH-H₂O, 90:10, isocratic), yielding compound **3** (5.5 mg, *t_R* 9 min). Fr. 28 was separated by ODS column chromatography using a MeOH-H₂O gradient; this separation produced Fr. e28-1 and Fr. e28-2 at 90% MeOH. Fr. e28-2 was then purified by semi-preparative HPLC with isocratic 85% MeOH-H₂O, affording compound **2** (2.1 mg, *t_R* 8 min).

2.2 Data preparation and target collection

The putative targets of the five flavonoids (compounds **1-5**) were predicted using SwissTargetPrediction (<https://swisstargetprediction.ch/>). For each compound, the canonical SMILES structure was submitted to the platform. Subsequently, the predicted human protein targets were collected and merged. Thereafter, duplicates were removed to generate a non-redundant compound-target set.

Three UC transcriptome datasets were retrieved from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). GSE243625 was designated as the primary dataset for downstream analyses. In contrast, GSE66407 and GSE206171 were employed as independent validation datasets.

2.3 Differential expression analysis of UC

Differential expression analysis was conducted in the primary UC dataset (GSE243625) using the limma package in R. Gene expression levels were compared between UC samples and healthy controls. Subsequently, differentially expressed genes (DEGs) were defined using stringent thresholds of $|\log_2FC| > 1$ and false discovery rate (FDR) < 0.05 . For visualization purposes, the DEG results were displayed using volcano plots and heatmaps.

2.4 Intersection of DEGs and compound targets and functional enrichment

The overlapping genes between the UC

DEGs and the predicted targets of compounds **1-5** were obtained to generate a common gene set. Subsequently, these common genes were subjected to Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses using the clusterProfiler package in R. Prior to enrichment analysis, gene symbols were converted to Entrez IDs using the org.Hs.eg.db package. Specifically, GO enrichment was performed using the enrichGO function. In parallel, KEGG pathway enrichment was conducted using the enrichKEGG function. Significance thresholds were set to $P < 0.05$ and $q < 0.05$ to ensure statistical rigor. Thereafter, the enrichment results were exported for further interpretation. Finally, bar plots and dot plots were generated using the enrichplot and ggplot2 packages.

2.5 WGCNA of the primary UC dataset

Weighted gene co-expression network analysis (WGCNA) was performed in R using the WGCNA package based on the primary UC dataset. Potential outlier samples were assessed by hierarchical clustering. Subsequently, eligible samples were retained for network construction. Additionally, clinical and phenotypic information was imported and visualized alongside the sample dendrogram as a trait heatmap.

An unsigned co-expression network was then constructed. First, the soft-thresholding power was determined using the pickSoftThreshold function. Following this, modules were identified using the blockwiseModules function with the following key settings: minimum module size = 50, deepSplit = 1, and module merging cut height = 0.6. Thereafter, module eigengenes were calculated and correlated with clinical traits to evaluate module-trait relationships. These relationships were summarized as a module-trait correlation heatmap. For the module showing the strongest association with the trait of interest, gene significance (GS) and module membership (MM) were computed. Finally, the GS-MM relationship was visualized using a scatter plot.

2.6 Machine learning

Machine learning approaches were applied to identify core genes associated with UC. First, candidate genes were defined as the intersection between the DEGs and the genes from the WGCNA module. Specifically, the WGCNA module showing the strongest correlation with the trait of interest was selected. Based on this candidate set, support vector machine-recursive feature elimination (SVM-RFE) was performed using five-fold cross-validation. This method ranked variables according to their importance. Subsequently, the optimal feature subset was determined based on the minimum classification error. Ultimately, the selected features were retained as core genes.

In parallel, least absolute shrinkage and selection operator (LASSO) regression was conducted using the glmnet package. A five-fold cross-validated LASSO model was fitted to determine the optimal penalty parameter (λ). Thereafter, genes with non-zero coefficients at the optimal λ were identified as LASSO-selected core genes. Finally, their coefficients were extracted for subsequent analyses.

2.7 ROC analysis

Receiver operating characteristic (ROC) analysis was performed to evaluate the diagnostic performance of the selected core genes in the two validation datasets (GSE66407 and GSE206171). First, the pROC package in R was employed for this analysis. ROC curves were generated for each core gene individually. Subsequently, a combined diagnostic model was constructed by fitting a logistic regression model based on the expression levels of all core genes. Thereafter, the area under the ROC curve (AUC) and its 95% confidence interval (CI) were calculated for the single-gene classifiers. Similarly, these metrics were also computed for the combined model. Finally, ROC curves were plotted to visualize and compare the classification performance across validation datasets.

2.8 GSEA enrichment analysis

Single-gene gene set enrichment analysis (GSEA) was conducted for each of the five core genes. First, based on the gene expression matrix, Spearman correlation coefficients between the target core gene and all other genes were calculated. Subsequently, these coefficients were used to generate a ranked gene list. The ranked list was then employed as input for GSEA implemented in the clusterProfiler package. Specifically, KEGG gene sets were obtained from the MSigDB collection (C2: CP: KEGG Medicus). These gene sets were subsequently imported for enrichment analysis. Thereafter, the enrichment results were exported for further interpretation. Finally, representative enrichment plots and summary visualizations were generated using the enrichplot and ggplot2 packages.

2.9 Correlation analysis

Correlation analysis was performed for the five core genes in two independent validation datasets. First, gene expression matrices were imported and subset to the five core genes. Subsequently, pairwise Spearman correlation coefficients were calculated across samples to construct a correlation matrix. For each gene pair, two-sided Spearman correlation tests were conducted to obtain *P* values. Thereafter, multiple testing correction was applied using the Benjamini-Hochberg method to calculate false discovery rates (FDRs). The pairwise correlation statistics were then exported for further evaluation. Finally, the correlation patterns were visualized as a heatmap using the ComplexHeatmap package. Specifically, correlation coefficients were annotated in each cell. Additionally, significance levels were indicated by FDR-based symbols.

2.10 Molecular docking

Molecular docking was performed between the five flavonoids (compounds **1-5**) and the five core

gene-encoded proteins, generating 25 ligand-protein docking pairs. Protein structures were obtained from the AlphaFold Protein Structure Database. Subsequently, ligand structures were prepared and docked to each protein using AutoDock Vina. Binding affinity scores predicted by Vina (kcal/mol) were employed as the primary metric to evaluate binding strength. Notably, lower affinity values indicated more favorable binding interactions. Finally, the docking affinities of all 25 pairs were summarized in a matrix and visualized as a heatmap to compare binding patterns across compounds and target proteins.

3 Results

3.1 Structural identification of compounds 1-5

Compound **1** (Fig. 1) was obtained as a yellow amorphous powder. Its ^1H NMR spectrum displayed a characteristic hydrogen-bonded hydroxyl signal at δ_{H} 12.77 (s, 5-OH) and an AA'BB' coupling system [δ_{H} 7.85 (2H, d, $J = 8.9$ Hz, H-2',6'), 7.02 (2H, d, $J = 8.9$ Hz, H-3',5')], indicating a 4'-substituted B-ring. Three methoxy singlets were observed at δ_{H} 3.97, 3.93, and 3.90. The ^{13}C NMR data were assigned as follows: ^{13}C NMR (150 MHz, CDCl_3): δ_{C} 164.2 (C-2), 104.3 (C-3), 182.8 (C-4), 153.4 (C-5), 132.7 (C-6), 158.8 (C-7), 90.7 (C-8), 153.2 (C-9), 106.3 (C-10), 123.7 (C-1'), 128.1 (C-2'), 114.6 (C-3'), 162.7 (C-4'), 114.6 (C-5'), 128.1 (C-6'), 61.0 (6-OCH₃), 56.4 (7-OCH₃), 55.6 (4'-OCH₃). By comparing these spectroscopic data with literature values, compound **1** was identified as 5-hydroxy-6,7,4'-trimethoxyflavone.

Compound **2** was isolated as a yellow amorphous powder. The ^1H NMR spectrum showed a chelated hydroxyl proton at δ_{H} 12.75 (s, 5-OH) and an ABX spin system on the B-ring [δ_{H} 7.53 (1H, dd, $J = 8.5$, 2.1 Hz, H-6'), 7.34 (1H, d, $J = 2.1$ Hz, H-2'), 6.98 (1H, d, $J = 8.5$ Hz, H-5')], along with four methoxy groups (δ_{H} 3.99, 3.98, 3.97, 3.93). The ^{13}C NMR data were assigned as follows: ^{13}C NMR (150 MHz, CDCl_3): δ_{C} 164.1 (C-2), 104.6 (C-3), 182.7 (C-4), 153.2 (C-5), 132.8 (C-6), 158.9 (C-7), 90.7 (C-8), 153.3 (C-9), 106.3 (C-10), 120.2 (C-1'), 111.3 (C-2'),

152.4 (C-3'), 149.5 (C-4'), 108.9 (C-5'), 123.9 (C-6'), 61.0 (6-OCH₃), 56.5 (7-OCH₃), 56.3 (3'-OCH₃), 56.2 (4'-OCH₃). Based on these data and comparison with reported values, compound **2** was assigned as 5-hydroxy-6,7,3',4'-tetramethoxyflavone.

Compound **3** was obtained as a yellow amorphous powder. The ^1H NMR spectrum revealed an AA'BB' coupling system [δ_{H} 7.82 (2H, dd, $J = 8.9$, 4.7 Hz, H-2',6'), 7.00 (2H, dd, $J = 8.9$, 4.6 Hz, H-3',5')] characteristic of a 4'-substituted B-ring, along with four methoxy singlets at δ_{H} 3.99, 3.98, 3.91, and 3.88. The ^{13}C NMR data were assigned as follows: ^{13}C NMR (100 MHz, CDCl_3): δ_{C} 161.3 (C-2), 107.1 (C-3), 177.4 (C-4), 96.3 (C-5), 157.7 (C-6), 140.4 (C-7), 152.7 (C-8), 154.6 (C-9), 112.9 (C-10), 124.0 (C-1'), 127.7 (C-2', 6'), 114.5 (C-3', 5'), 162.2 (C-4'), 56.4 (6-OCH₃), 62.3 (7-OCH₃), 55.6 (8-OCH₃), 61.6 (4'-OCH₃). Comparison with literature data identified compound **3** as 6,7,8,4'-tetramethoxyflavone.

Compound **4** was purified as a yellow amorphous powder with a specific rotation of $[\alpha]_{\text{D}}^{20} +12.36$ (c 0.1, MeOH). The ^1H NMR spectrum displayed characteristic flavanone H-2/H-3 coupling patterns [δ_{H} 5.34 (1H, dd, $J = 13.4$, 2.7 Hz, H-2), 3.04 (1H, dd, $J = 16.7$, 13.4 Hz, H-3 α), 2.77 (1H, dd, $J = 16.7$, 2.8 Hz, H-3 β)], along with an ABX system on the B-ring and five methoxy groups (δ_{H} 3.83-3.94). The ^{13}C NMR data were assigned as follows: ^{13}C NMR (150 MHz, CDCl_3): δ_{C} 79.5 (C-2), 45.6 (C-3), 189.6 (C-4), 154.4 (C-5), 137.7 (C-6), 159.6 (C-7), 96.5 (C-8), 159.9 (C-9), 109.3 (C-10), 131.2 (C-1'), 109.4 (C-2'), 149.4 (C-3'), 149.6 (C-4'), 111.2 (C-5'), 119.0 (C-6'), 61.7 (5-OCH₃), 61.5 (6-OCH₃), 61.5 (7-OCH₃), 56.1 (3'-OCH₃), 56.1 (4'-OCH₃). These data were consistent with those reported for 5,6,7,3',4'-pentamethoxyflavanone, thus confirming the identification of compound **4**.

Compound **5** was obtained as a yellow amorphous powder. The ^1H NMR spectrum featured characteristic trans-olefinic protons of a chalcone skeleton at δ_{H} 7.85 (1H, d, $J = 15.5$ Hz, H- α) and 7.81 (1H, d, $J = 15.5$ Hz, H- β). An ABX spin system was observed for the B-ring protons [δ_{H} 7.29 (1H, m, H-6), 7.25 (1H, m, H-2), 6.91 (1H, d, $J = 8.3$ Hz,

H-5)], along with five methoxy singlets (δ_{H} 3.95, 3.94, 3.93, 3.91, 3.84). The ^{13}C NMR data were assigned as follows: ^{13}C NMR (150 MHz, CDCl_3): δ_{C} 128.5 (C-1), 110.4 (C-2), 149.3 (C-3), 151.4 (C-4), 111.3 (C-5), 123.1 (C-6), 108.8 (C-1'), 162.8 (C-2'), 96.7 (C-3'), 160.1 (C-4'), 135.4 (C-5'), 155.0 (C-6'), 124.4 (C- α), 143.8 (C- β), 192.8 (C- β'), 56.0 (3-OCH₃), 56.1

(4-OCH₃), 56.2 (4'-OCH₃), 61.4 (5'-OCH₃), 62.1 (6'-OCH₃). These data matched those reported for 2'-hydroxy-3,4,4',5',6'-pentamethoxychalcone, leading to the identification of compound **5** as such.

Notably, compounds **3-5** were isolated from *C. spicatus* for the first time.

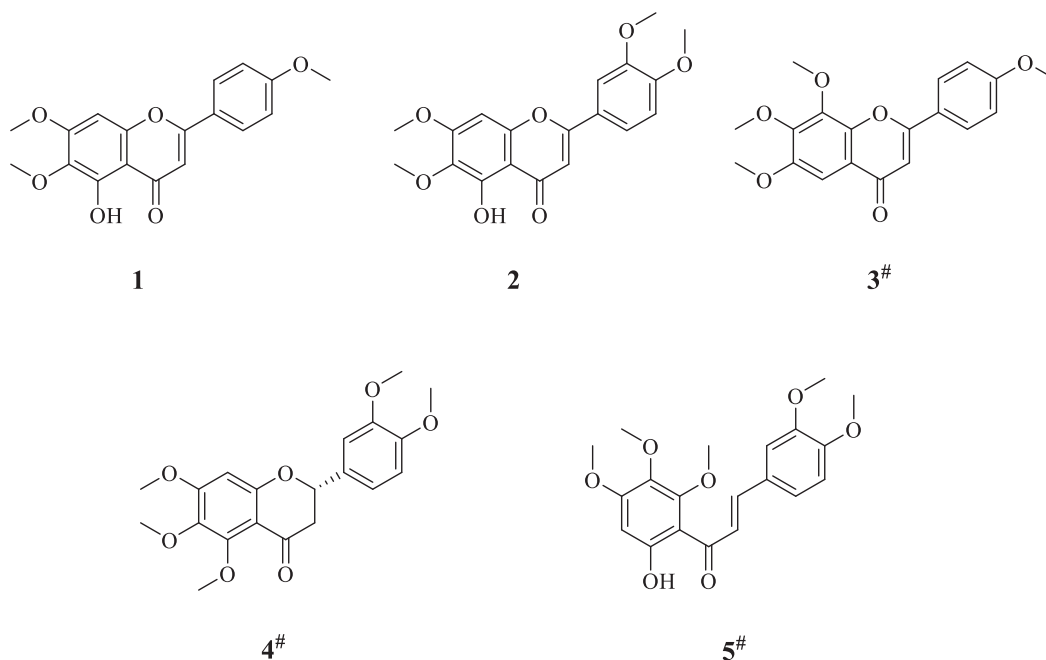


Fig. 1 The compounds isolated from the root barks of *C. spicatus*. ([#] Indicates compounds isolated from the *C. spicatus*. for the first time)

3.2 Network pharmacology analysis revealed IL-17 and TNF signaling as key pathways

As shown in Fig. 2A and 2B, differential expression analysis identified 849 DEGs between UC and control samples. These DEGs were visualized by a heatmap. Additionally, a volcano plot was generated to display the distribution pattern. SwissTargetPrediction yielded 245 putative targets for the five polyphenolic compounds from *C. spicatus*. Subsequently, these targets were intersected with the DEGs. This intersection produced 32 common genes,

as summarized in the Venn diagram (Fig. 2C). KEGG enrichment analysis of the 32 common genes revealed significant enrichment in inflammation-related pathways. In particular, the IL-17 signaling pathway and the TNF signaling pathway showed prominent enrichment (Fig. 2D). Taken together, these results suggested that the five polyphenolic compounds from *C. spicatus* might exert therapeutic effects against UC. Specifically, this activity appeared to be mediated primarily through modulation of IL-17- and TNF-mediated inflammatory responses.

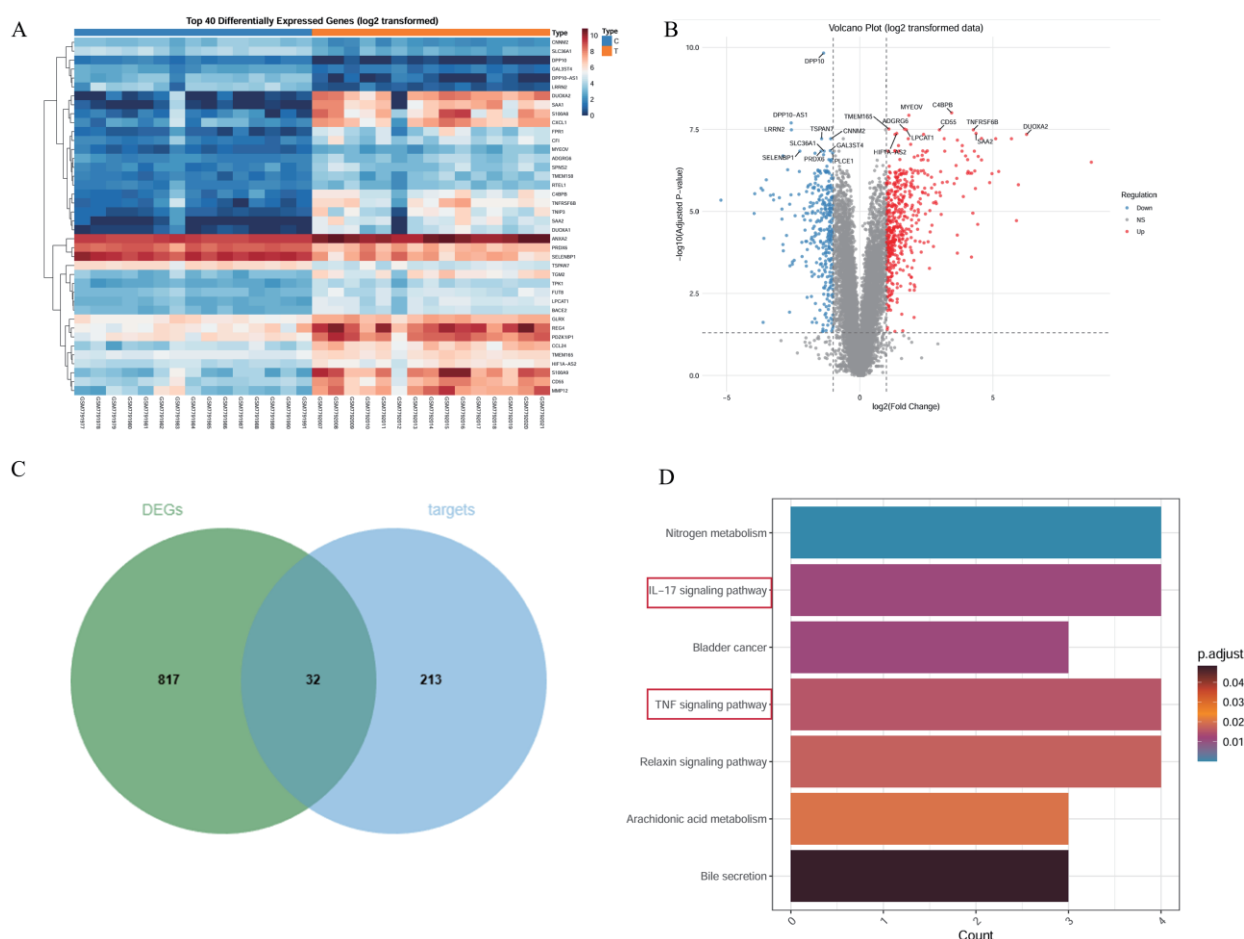


Fig. 2 Network pharmacology analysis, **A**: Heat maps analysis of DEGs; **B**: Volcano plot of DEGs; **C**: Venn diagram analysis between DEGs and compounds active targets genes; **D**: KEGG enrichment analysis of intersection genes

3.3 WGCNA identified the MEgrey module as the key UC-associated module

As shown in Fig. 3A, hierarchical clustering of genes based on expression similarity generated the module dendrogram. Subsequently, this dendrogram was used for WGCNA module detection. A total of 23 co-expression modules were identified. The module-trait heatmap (Fig. 3B) indicated the correlation between these modules and UC. Notably, the MEgrey module exhibited the strongest association with UC ($r = 0.85$, $P = 3 \times 10^{-9}$). Consequently, MEgrey

was selected as the key UC-related module for subsequent analyses. Gene significance (GS) and module membership (MM) analyses were performed within the MEgrey module (Fig. 3C). These analyses demonstrated a strong positive relationship ($\text{cor} = 0.84$, $P < 1 \times 10^{-200}$). Moreover, this finding supported the biological relevance of this module to UC. Finally, MEgrey module genes were intersected with DEGs. This intersection yielded 128 common genes (Fig. 3D). These genes were defined as candidate genes for downstream machine learning screening.

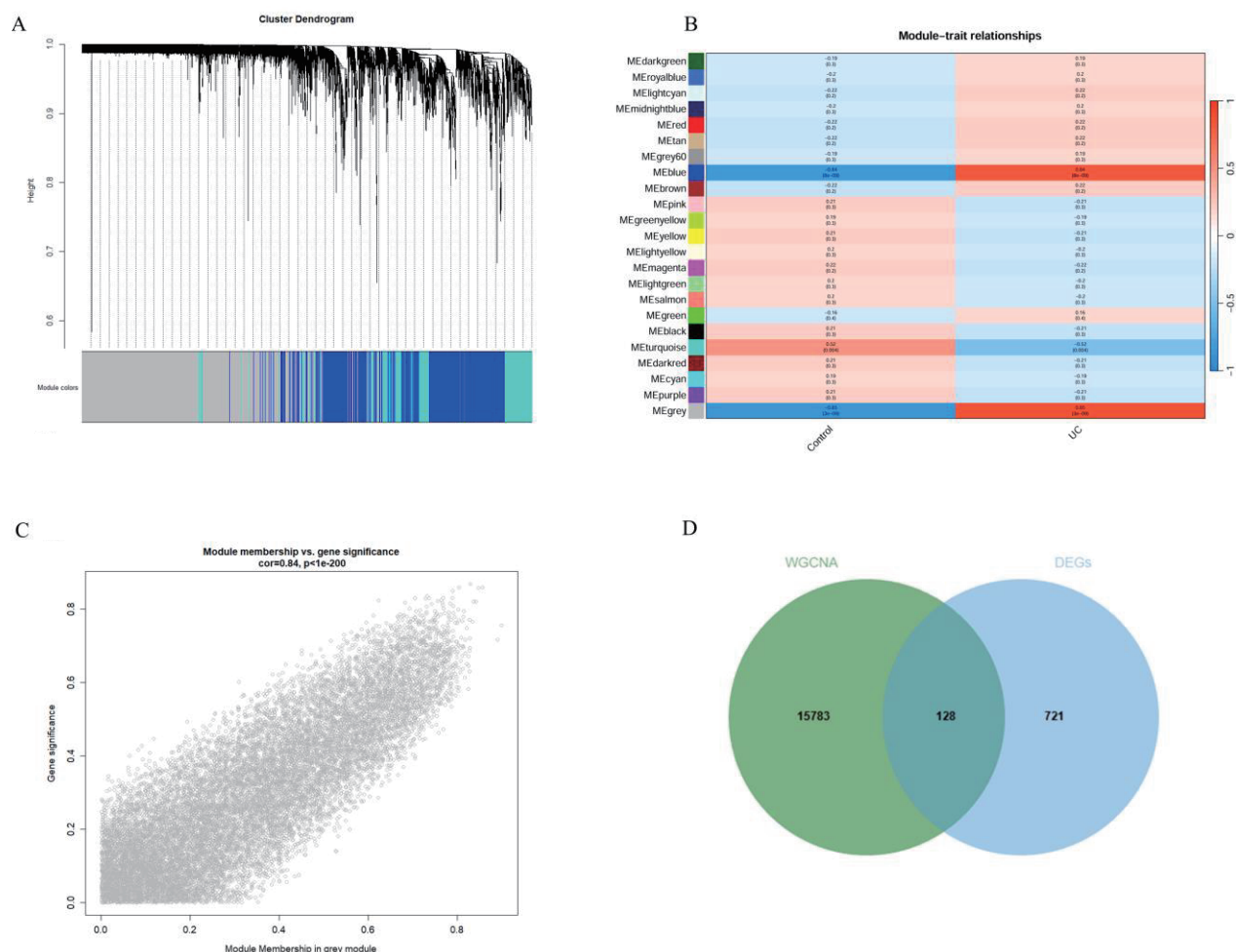


Fig. 3 WGCNA analysis, A: Hierarchical clustering dendrogram of genes and identified co-expression modules; B: Heat map depicting 23 modules and the correlation between UC and control groups; C: Correlation between the MEgrey gene module and IBD; D: Venn diagram of intersection genes between DEGs and MEgrey module genes

3.4 Machine learning identified five core genes

Machine learning was applied to further screen candidate genes. As shown in Fig. 4A and 4B, LASSO regression with five-fold cross-validation was employed. This method was used to determine the optimal penalty parameter and shrink regression coefficients. As a result, 10 genes with non-zero coefficients were identified. In parallel, SVM-RFE

was performed using five-fold cross-validation. The error and accuracy profiles across different feature numbers are presented in Fig. 4C and 4D. Based on this analysis, 13 genes were retained as the optimal feature subset. Subsequently, the overlap between the LASSO-derived and SVM-derived gene sets was calculated. This overlap comprised five shared genes (Fig. 4E). Ultimately, these genes were defined as the final core genes for subsequent analyses.

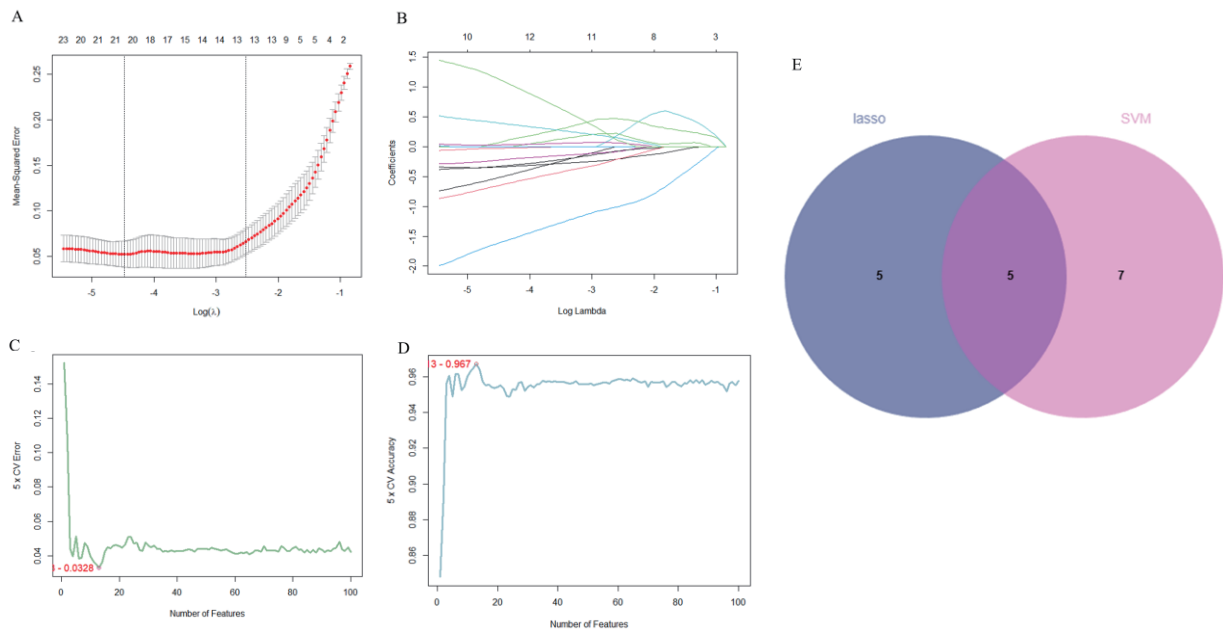


Fig. 4 Core gene screening, **A:** LASSO coefficient map of core genes; **B:** Verification of the optimal parameter (lambda) in the LASSO model; **C:** Error curve of the SVM algorithm; **D:** Accuracy curve of the SVM algorithm; **E:** Intersection of genes identified by SVM and LASSO algorithms

3.5 Validation of core genes in independent cohorts

In the two independent validation cohorts, the five core genes (NPTX2, SIGLEC12, SPINK4, SOX12, and ZNF575) exhibited both shared and cohort-specific correlation patterns. In GSE66407 (Fig. 5A), most pairwise correlations were weak. However, SIGLEC12 showed a modest positive correlation with SPINK4 (Spearman's $P = 0.25$, $FDR < 0.001$). In contrast, several negative correlations were evident. Specifically, SPINK4 was negatively correlated with SOX12 ($P = -0.26$, $FDR < 0.001$). Notably, SPINK4 showed a stronger negative correlation with ZNF575 ($P = -0.46$, $FDR < 0.001$). Additionally, SIGLEC12 was negatively correlated with ZNF575 ($P = -0.20$, $FDR < 0.001$). Conversely, NPTX2 showed minimal correlation with the remaining genes.

In GSE206171 (Fig. 5C), the correlation structure became more pronounced. NPTX2 was moderately and positively correlated with SIGLEC12 ($P = 0.34$, $FDR < 0.001$). Similarly, NPTX2 also correlated positively with SPINK4 ($P = 0.35$, $FDR < 0.001$). Furthermore, SIGLEC12 correlated positively

with SPINK4 ($P = 0.29$, $FDR < 0.001$). These findings suggested increased concordance among these genes in this cohort. Meanwhile, the negative association between SPINK4 and ZNF575 was consistently observed and slightly strengthened ($P = -0.51$, $FDR < 0.001$). Likewise, ZNF575 remained negatively correlated with SIGLEC12 ($P = -0.26$, $FDR < 0.01$). ZNF575 was also negatively correlated with NPTX2 ($P = -0.23$, $FDR < 0.01$). In comparison to GSE66407, SPINK4 and SOX12 showed only a weak negative correlation ($P = -0.12$). Conversely, SOX12 was negatively correlated with ZNF575 ($P = -0.21$, $FDR < 0.05$).

ROC analyses further supported the diagnostic value of the core genes in both cohorts (Fig. 5B and Fig. 5D). In GSE66407, the combined logistic model achieved an AUC of 0.812 (95% CI: 0.760-0.863). This model outperformed single-gene classifiers. Among individual genes, ZNF575 (AUC = 0.715) and SPINK4 (AUC = 0.709) showed the best performance. In contrast, SIGLEC12 demonstrated limited discrimination (AUC = 0.512). In GSE206171, the combined model showed improved performance (AUC =

0.861, 95% CI: 0.800–0.922). Among individual genes, ZNF575 (AUC = 0.788) and SPINK4 (AUC = 0.750) remained the strongest contributors. However, SOX12 showed comparatively lower accuracy (AUC = 0.573). Overall, the correlation analyses indicated

partially conserved co-expression relationships across cohorts. Moreover, the ROC results demonstrated that integrating the five genes consistently provided superior diagnostic performance compared with any single marker.

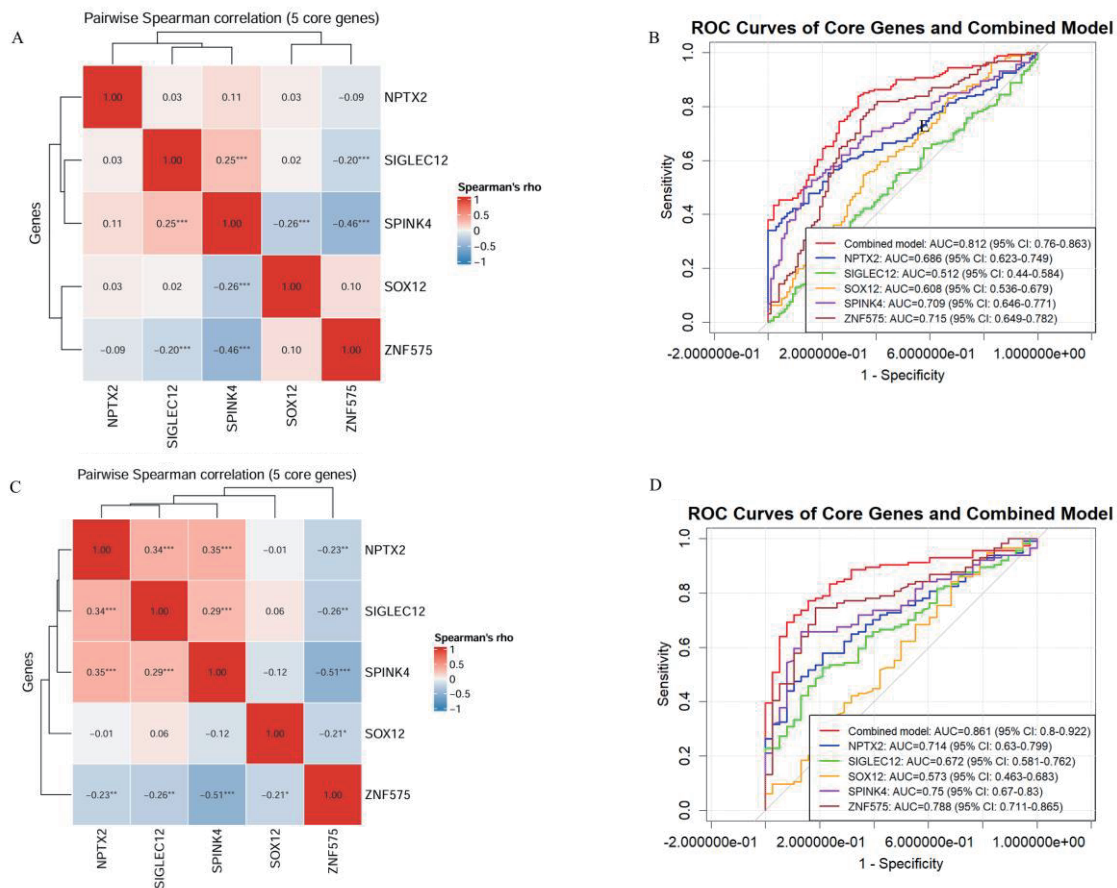


Fig. 5 Validation of core gene expression and diagnostic performance in two independent cohorts, A: Correlation heatmap of core genes in the GSE66407 cohort; B: ROC curves for the core genes in the GSE66407 cohort; C: Correlation heatmap of core genes in the GSE206171 cohort; D: ROC curves for the core genes in the GSE206171 cohort

3.6 GSEA analysis

Single-gene GSEA was conducted to clarify the inflammatory pathway context associated with each core gene. The NPTX2-associated signature (Fig. 6A) showed enrichment of cytokine-driven programs, particularly cytokine/JAK-STAT and IL-2 family-to-JAK/STAT signaling, together with innate inflammatory modules including TLR2/4-NF- κ B and TNF/NF- κ B signaling. For SIGLEC12 (Fig. 6B), enrichment was again observed in cytokine/JAK-STAT signaling, with additional signals related to IL-6

family-to-JAK/STAT and type I interferon pathways, and it also overlapped with TLR2/4-NF- κ B and TNF/NF- κ B inflammatory cascades. ZNF575 (Fig. 6C) was enriched for cytokine/JAK-STAT and IL-2 family-to-JAK/STAT signaling, together with chemokine- and PI3K-linked programs such as CCR/CXCR-PI3K-RAC and PI3K/NOX2-related signatures. The SOX12-related profile (Fig. 6D) remained centered on cytokine/JAK-STAT and IL-2 family-to-JAK/STAT signaling, while also highlighting CCR/CXCR-GNB-PI3K-RAC and PI3K/NOX2-associated inflammatory signatures. Similarly, SPINK4 (Fig. 6E) was enriched

for cytokine/JAK-STAT and IL-2 family-to-JAK/STAT signaling, and it further implicated chemokine-related CCR/CXCR-PI3K-RAC signaling and immune receptor-coupled programs (e.g., BCR/PLC γ /calcineurin-related signatures). Taken together, these

results indicated that all five core genes were closely embedded in cytokine-, NF- κ B-, and chemokine-associated inflammatory networks, supporting their relevance to UC pathogenesis and highlighting them as inflammation-related core genes in UC.

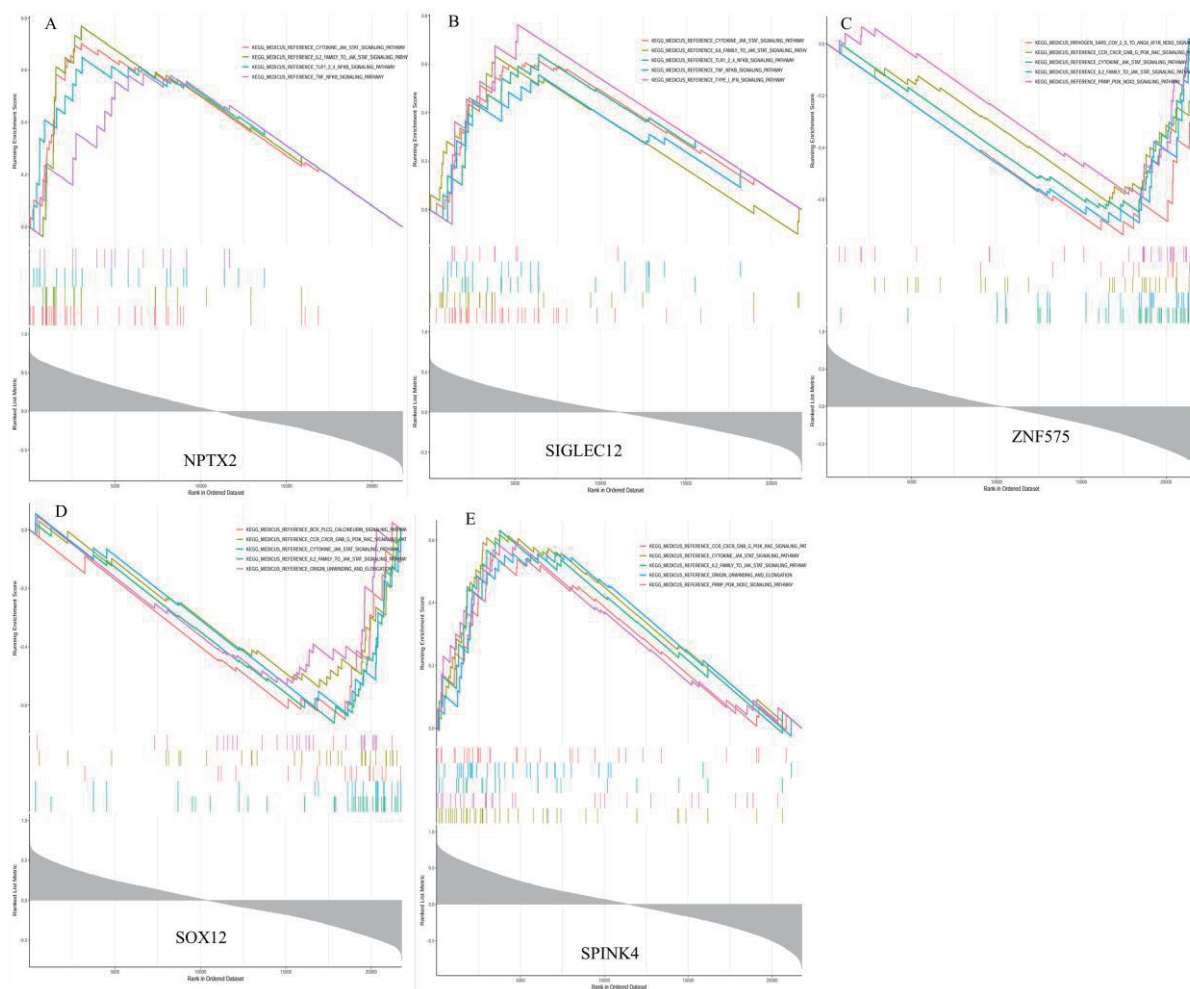


Fig. 6 GSEA analysis, A: GSEA analysis for NPTX2; B: GSEA analysis for SIGLEC12; C: GSEA analysis for SOX12; D: GSEA analysis for SPINK4; E: GSEA analysis for ZNF57

3.7 Molecular docking validated the binding of compounds 1-5 to UC core targets

Molecular docking was performed between the five polyphenolic compounds (compounds 1-5) and the five UC core targets (NPTX2, ZNF575, SPINK4, SIGLEC12, and SOX12) (Fig. 7). As summarized in the docking heatmap, all 25 compound-protein pairs were analyzed. These pairs exhibited favorable binding affinities, as indicated by negative Vina

scores. Specifically, the values ranged from -8.0 to -4.8 kcal/mol. Notably, compound 1 showed the strongest predicted interaction with NPTX2 (-8.0 kcal/mol). In contrast, the remaining pairs generally displayed moderate affinities around -6 kcal/mol. Overall, these docking results supported the potential of the five *C. spicatus* polyphenols to interact with UC pathogenesis-related core proteins. This finding provided structural evidence for their putative therapeutic effects.

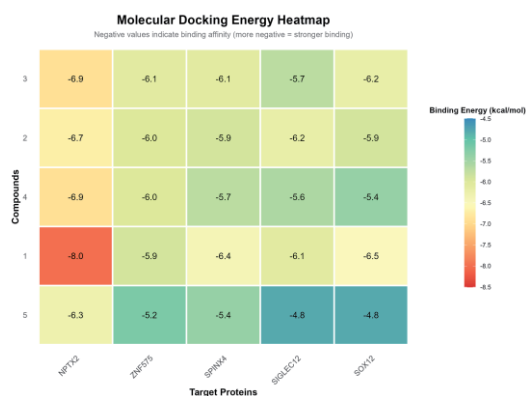


Fig. 7 Heatmap of molecular dynamics analysis

4 Discussion

Five polymethoxylated flavonoids (compounds **1-5**) were isolated from *C. spicatus* and were identified as 5-hydroxy-6,7,4'-trimethoxyflavone (**1**), 5-hydroxy-6,7,3',4'-tetramethoxyflavone (**2**), 6,7,8,4'-tetramethoxyflavone (**3**), 5,6,7,3',4'-pentamethoxyflavanone (**4**), and 2'-hydroxy-3,4,4',5',6'-pentamethoxychalcone (**5**) based on NMR data and comparison with reported values. Compounds **1-3** were flavones, compound **4** was a flavanone, and compound **5** was a chalcone. Structurally, compounds **1** and **2** shared a 5-hydroxy-6,7-dimethoxy substitution pattern on the A-ring and differed in the B-ring substitution (4'-substituted in **1** and 3',4'-disubstituted in **2**). Compound **3** differed from **1** and **2** by lacking the 5-hydroxy group and bearing a 6,7,8-trimethoxy pattern on the A-ring, while retaining a 4'-substituted B-ring. Notably, compounds **3-5** were isolated from *C. spicatus* for the first time.

UC is a chronic relapsing inflammatory bowel disease characterized by immune dysregulation, epithelial barrier dysfunction, and persistent mucosal inflammation [9]. Despite advances in therapeutic strategies, including biologics targeting TNF- α , integrins, and interleukins, challenges remain. A substantial proportion of patients experience inadequate responses, loss of response, or adverse effects. This underscores the need for novel therapeutic agents with multi-target efficacy and favorable safety profiles [10]. Plant-derived flavonoids present an attractive alternative, combining multi-

target modulation with inherently lower toxicity burdens.

The pathogenesis of UC is multifaceted. It involves genetic susceptibility, environmental triggers, gut microbiota dysbiosis, and aberrant immune responses. Among the key molecular drivers, IL-17 and TNF- α signaling pathways have emerged as central orchestrators of the inflammatory cascade in UC. IL-17, primarily secreted by Th17 cells, promotes neutrophil recruitment, induces pro-inflammatory cytokines, and disrupts intestinal barrier integrity. In contrast, TNF- α amplifies NF- κ B-dependent inflammatory gene expression and sustains chronic mucosal inflammation [11]. Consequently, therapeutic interventions targeting these pathways hold promise for attenuating UC pathology.

Our network pharmacology analysis revealed that the five *C. spicatus* flavonoids converged on two central inflammatory pathways: the IL-17 signaling pathway and the TNF signaling pathway. Specifically, KEGG enrichment of the 32 overlapping genes between compound targets and UC DEGs highlighted these pathways as the most significantly enriched. This suggested that the therapeutic effects of these flavonoids are mediated through simultaneous suppression of IL-17- and TNF- α -driven inflammatory responses.

This dual-pathway targeting is particularly relevant to UC pathogenesis, where IL-17 and TNF- α act synergistically to amplify intestinal inflammation. IL-17 drives epithelial damage and immune cell infiltration. Meanwhile, TNF- α sustains chronic

inflammation via NF- κ B activation and downstream production of pro-inflammatory mediators such as IL-6, IL-8, and COX-2. By intercepting both pathways, the five flavonoids may offer a more comprehensive anti-inflammatory effect than single-pathway inhibitors. This approach potentially addresses the multi-layered nature of UC inflammation.

Beyond pathway-level insights, our study identified five core genes that are critical to UC pathogenesis: NPTX2, SIGLEC12, SPINK4, SOX12, and ZNF575. These genes serve as molecular targets for the flavonoid compounds. Critically, four of the five core genes have been functionally validated in UC or colitis models. NPTX2 promotes Th17-mediated inflammation in experimental colitis^[12]. SIGLEC12 executes necroptotic cell death, contributing to epithelial damage and inflammatory propagation^[13]. SPINK4 exerts mucosal protection by promoting goblet cell differentiation via EGFR-Wnt/ β -catenin signaling and has been validated as a therapeutic target synergizing with anti-TNF- α agents^[14]. SOX12 suppresses intestinal inflammation by driving Foxp3 expression and peripheral Treg differentiation^[15]. These findings establish the biological plausibility of targeting these genes for UC therapy.

Linking these gene-level functions to the pathway-level findings, single-gene GSEA demonstrated that all five core genes are functionally embedded within cytokine/JAK-STAT, TLR/NF- κ B, TNF/NF- κ B, and chemokine-related inflammatory networks. For instance, NPTX2, SIGLEC12, and ZNF575 were enriched in IL-2/IL-6 family cytokine signaling and type I interferon responses—pathways that amplify Th17 differentiation and TNF- α production. Conversely, SOX12 and SPINK4 were linked to chemokine receptor-PI3K-RAC signaling and immune receptor activation, which modulate immune cell trafficking and NF- κ B activation downstream of TNF- α . These enrichment profiles collectively establish that the five core genes are not merely UC-associated biomarkers. Rather, they are mechanistically connected to the very inflammatory pathways (IL-17 and TNF- α) identified at the systems level.

Molecular docking analysis further supported this mechanistic link. It demonstrated that all five *C. spicatus* flavonoids exhibit favorable binding affinities to the five core target proteins (−8.0 to −4.8 kcal/mol). This structural evidence suggests that the flavonoids can physically interact with these inflammation-related proteins. Such interactions potentially modulate their expression, activity, or downstream signaling. This multi-target engagement at the molecular level complements the pathway-level modulation of IL-17 and TNF- α . Ultimately, it provides a plausible mechanistic basis for the anti-UC effects of *C. spicatus* polyphenols.

In summary, our study revealed a dual-level therapeutic mechanism for the five polymethoxylated flavonoids from *C. spicatus* in UC. At the pathway level, they modulate the IL-17 and TNF- α signaling axes that drive UC inflammation. At the molecular level, they target five core inflammation-related genes embedded within these pathways. This multi-target, multi-pathway mode of action aligns with the holistic philosophy of traditional herbal medicine and offers a promising strategy for addressing the complexity of UC pathogenesis. However, several limitations warrant consideration. Our findings derive primarily from computational analyses; experimental validation remains essential. The predicted binding interactions require biochemical confirmation, and the functional consequences of flavonoid-target engagement on inflammatory pathways must be assessed in cellular and animal models of colitis. Moreover, the pharmacokinetic properties and *in vivo* efficacy of these compounds have yet to be established. Future work integrating *in vitro* mechanistic studies and preclinical testing will be critical to translate these predictions into therapeutic applications.

5 Conclusion

In this study, five polymethoxylated flavonoids (compounds **1-5**) were isolated from *C. spicatus*. These compounds were evaluated for their anti-UC mechanisms through an integrative computational approach. Our findings revealed that these flavonoids

exerted therapeutic effects via dual mechanisms. At the systems level, they modulated the IL-17 and TNF- α inflammatory pathways. At the molecular level, they targeted five core inflammation-related genes. Specifically, these genes included NPTX2, SIGLEC12, SPINK4, SOX12, and ZNF575. Subsequently, molecular docking confirmed favorable compound-protein interactions. Collectively, these results provided mechanistic insights into the traditional use of *C. spicatus* for inflammatory disorders. Ultimately, they highlighted its potential as a multi-target therapeutic agent for UC.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supporting information

This work is mainly supported by Excellent Youth program of Shenyang Pharmaceutical University (Admission No. YQ202118), and Strategic Industrialization Project of Liaoning Provincial Department of Education (LJ212410163010).

Data availability

Data will be made available on request.

References

- [1] Calvez V, Puca P, Di Vincenzo F, et al. Novel Insights into the Pathogenesis of Inflammatory Bowel Diseases. *Biomedicines*, 2025, 13 (2): 305.
- [2] Katsanos KH, Papadakis KA. Inflammatory Bowel Disease: Updates on Molecular Targets for Biologics. *Gut Liver*, 2017, 11 (4): 455-63.
- [3] Paula Sanchez Zanotti M, Cataldi de Alcântara C, Junko Inoue C, et al. Involvement of IL17A and IL17RA variants in interleukin-17A levels and disease activity in ulcerative colitis. *Cytokine*, 2024, 182: 156716.
- [4] Soldera J. Navigating treatment resistance: Janus kinase inhibitors for ulcerative colitis. *World J Clin Cases*, 2024, 12 (24): 5468-72.
- [5] Wetwittayakhleng P, Lakatos PL. Advanced combination therapy: is it the best way to break the therapeutic ceiling? *Therap Adv Gastroenterol*, 2024, 17: 17562848241272995.
- [6] Jiao X, Jin Q, Zhu P, et al. *Clerodendranthus spicatus*: a comprehensive review of the chemical constituents, pharmacology, quality control and clinical applications. *Front Pharmacol*, 2025, 16: 1452797.
- [7] Li KW, Raza F, Jiang LD, et al. *Clerodendranthus Spicatus*: A review of its active compounds, mechanisms of action, and clinical studies in urinary diseases. *Fitoterapia*, 2024, 177: 106082.
- [8] Ma GX, Zhang XP, Li PF, et al. Four new phenolic acid with unusual bicycle [2.2.2] octane moiety from *Clerodendranthus spicatus* and their anti-inflammatory activity. *Fitoterapia*, 2015, 105: 61-65.
- [9] Ren D, Dong Z, Wang M, et al. Novel insight into the gene etiology of ulcerative colitis gained from transcriptome association study and single-cell sequence analysis. *Medicine (Baltimore)*, 2025, 104 (51): e46678.
- [10] Yan JW, Nie M, Zhang H, et al. Strengths, weaknesses, opportunities, and threats analysis of combination therapy for inflammatory bowel disease. *World J Gastroenterol*, 2025, 31 (9): 100607.
- [11] Lundekvam JA, Høivik ML, Anisdahl K, et al. Tumour necrosis factor inhibitors in Ulcerative colitis: real-world data on Therapeutic drug monitoring and evaluation of current treatment targets (STRIDE II). *Ann Med*, 2025, 57 (1): 2424447.
- [12] Magisetty J, Gadiraju B, Kondreddy V. Genomic analysis in the colon tissues of omega-3 fatty acid-treated rats identifies novel gene signatures implicated in ulcerative colitis. *Int J Biol Macromol*, 2024, 258 (Pt 1): 128867.
- [13] Noh H, Hashem Z, Boms E, et al. SIGLEC12 mediates plasma membrane rupture during necroptotic cell death. *Nature*, 2026, 649 (8096): 460-66.
- [14] Wang Y, Han J, Yang G, et al. Therapeutic potential of the secreted Kazal-type serine protease inhibitor SPINK4 in colitis. *Nat Commun*, 2024, 15 (1): 5874.
- [15] Tanaka S, Suto A, Iwamoto T, et al. Sox12 promotes T reg differentiation in the periphery during colitis. *J Exp Med*, 2018, 215 (10): 2509-19.