



OPEN

Identification of subclusters and prognostic genes based on GLS-associated molecular signature in ulcerative colitis

Yang Xie^{1,2}, Jun Li^{1,2}, Qing Tao^{1,2}, Yonghui Wu¹, Zide Liu¹, Youxiang Chen^{1✉} & Chunyan Zeng^{1✉}

Ulcerative colitis (UC) is a chronic and recurrent inflammatory disease that affects the colon and rectum. The response to treatment varies among individuals with UC. Therefore, the aim of this study was to identify and explore potential biomarkers for different subtypes of UC and examine their association with immune cell infiltration. We obtained UC RNA sequencing data from the GEO database, which included the training set GSE92415 and the validation set GSE87473 and GSE72514. UC patients were classified based on GLS and its associated genes using consensus clustering analysis. We identified differentially expressed genes (DEGs) in different UC subtypes through a differential expression analysis of the training cohort. Machine learning algorithms, including Weighted Gene Co-Expression Network Analysis (WGCNA), Least Absolute Shrinkage and Selection Operator (LASSO), and Support Vector Machine Recursive Feature Elimination (SVM-RFE), were utilized to identify marker genes for UC. The CIBERSORT algorithm was used to determine the abundance of various immune cells in UC and their correlation with UC signature genes. Finally, we validated the expression of GLS through in vivo and ex vivo experiments. The expression of GLS was found to be elevated in patients with UC compared to normal patients. GLS and its related genes were able to classify UC patients into two subtypes, C1 and C2. The C1 subtype, as compared to the C2 subtype, showed a higher Mayo score and poorer treatment response. A total of 18 DEGs were identified in both subtypes, including 7 up-regulated and 11 down-regulated genes. Four UC signature genes (CWH43, HEPACAM2, IL24, and PCK1) were identified and their diagnostic value was validated in a separate cohort (AUC > 0.85). Furthermore, we found that UC signature biomarkers were linked to the immune cell infiltration. CWH43, HEPACAM2, IL24, and PCK1 may serve as potential biomarkers for diagnosing different subtypes of UC, which could contribute to the development of targeted molecular therapy and immunotherapy for UC.

Keywords Ulcerative colitis, Machine-learning algorithms, Diagnostic markers, Immune cell infiltration

Ulcerative colitis (UC) is a chronic inflammatory bowel disease characterized by inflammation and ulceration of the colon and rectum¹. It affects millions of people worldwide, with a higher prevalence in developed countries². Current treatment options for UC mainly include medications such as amino salicylates, corticosteroids, immunomodulators, and biologics². These therapeutic agents aim to control inflammation, induce and maintain remission, and prevent relapse. However, despite the availability of these treatments, a considerable proportion of UC patients still experience inadequate disease control, adverse effects, or loss of response over time³.

Glutamine is an important component of amino acid metabolism and is involved in the synthesis and degradation of amino acids in the body, maintains the body's nitrogen balance, participates in energy production, regulates the immune response, and fights oxidative stress, among other important roles^{4–6}. Several previous studies have reported that glutamine is essential for tumor cell function, including the synthesis of metabolites that maintain mitochondrial metabolism, the generation of antioxidants to scavenge reactive oxygen species, the synthesis of non-essential amino acids (NEAAs), purines, pyrimidines, and fatty acids required for cellular

¹Department of Gastroenterology, Digestive Disease Hospital, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, China. ²These authors contributed equally: Yang Xie, Jun Li and Qing Tao. ✉email: Chenyx102@126.com; zengcy896@ncu.edu.cn

replication, and the activation of cell signaling⁴. Targeting glutamine metabolism can inhibit the growth of a variety of tumor cells^{7,8}, such as prostate cancer⁹, pancreatic cancer¹⁰, ovarian cancer¹¹, breast cancer^{12,13}, and lung cancer¹⁴. In addition, glutamine is involved in various metabolic pathways in cancer stem cells, such as the redox pathway, ATP generation, biosynthesis of non-essential amino acids and nucleotides, as well as ammonia production and epigenetic modification¹⁵. However, glutamine has been less frequently reported in UC. Glutamine was found to protect the intestinal mucosa by promoting mucosal cell growth and repair^{16,17}. Further studies by Jeong et al. found that glutamine ameliorated DSS-induced colonic inflammation in mice by inducing the MAPK signaling pathway¹⁸. Similarly, a study by Giris et al. found that the intestinal mucosal structure was preserved in the glutamine-treated group, and that glutamine attenuated the damage to the colon caused by TNBS-induced colitis¹⁹. Interestingly, Bernd Sido et al. found that glutamine supplementation prevented inflammatory tissue damage in experimental ileitis. However, immune enhancement of colonic inflammatory cells by glutamine increased oxidative tissue damage²⁰. These studies suggest that whether glutamine has a therapeutic effect in colitis remains controversial.

Glutaminase (GLS) is the first key enzyme that catalyzes the conversion of glutamine to glutamate, which plays a crucial role in cellular metabolism and energy production. GLS expression has been found to be elevated and pro-cancer in many tumors, such as breast cancer²¹, colorectal cancer²² and prostate cancer²³. However, there are no studies related to the role and specific mechanisms of glutaminase in ulcerative colitis.

The aim of this study was to understand the role of GLS in UC, to determine the significance of GLS-related genes for typing in UC, and to identify signature genes in different subtypes of UC. In this study, we used machine learning algorithms to investigate and validate the characterized genes in patients with different subtypes of UC. In addition, we quantified immune cell infiltration in different subtypes of UC using the CIBERSORT algorithm, revealing the correlation between UC signature gene expression and immune cell infiltration. Gene set enrichment analysis (GSEA) was used to identify potentially relevant signaling pathways in different subtypes of UC. Finally, we verified the expression of GLS in acute DSS mouse model tissues and LPS-induced enteritis NCM460 cell model by western blotting (WB).

Materials and methods

GEO datasets selection and GLS-related genes

We conducted a systematic search for ulcerative colitis in the GEO database, focusing on colon samples from patients with ulcerative colitis and datasets containing relevant clinical information. After excluding animal studies, RNA-seq datasets, and whole blood samples, we identified several studies. The GSE92415 dataset included expression profiles from tissue samples of 162 UC patients and 21 healthy controls, while the GSE87473 dataset included expression profiles from tissue samples of 106 UC patients and 21 normal controls. Dataset GSE72514 comprises transcriptomic data obtained from colonic tissue samples collected from 97 UC patients. GSE92415 was designated as the discovery cohort, GSE87473 and GSE72514 was designated as the validation cohort. We identified GLS-related genes from the STRING database (<https://string-db.org>) based on protein interactions and signaling pathways.

Consensus clustering analysis

Consensus cluster analysis is an unsupervised machine learning algorithm used to identify groups of samples that remain clustered together²⁴. The R package "ConensusClusterPlus" was utilized to classify the UC samples in GSE92415 into different subtypes. The clustering process was repeated 10 times to ensure the confidence of the classification. The optimal number of clusters (k) was selected by examining the consistency matrix for each k-value and using the cumulative density function (CDF) of the consistency index. Following clustering, the different subtypes were visualized using principal component analysis and cluster analysis.

Differently expressed genes (DEGs) determination

DEGs between cluster1 and cluster2 samples were identified using the Limma package in Rstudio. Criteria for DEGs included $|\text{Log}_2\text{FC}| > 1$, $p < 0.05$, and a false discovery rate (FDR) < 0.05 . FDR quantifies the proportion of false discoveries among a set of hypothesis tests deemed significant.

Weighted gene co-expression network analysis (WGCNA) and functional enrichment analysis

WGCNA is a systems biology method used to identify co-expression modules, which are groups of genes jointly regulated under diverse conditions²⁵. In our study, we utilized the WGCNA package in R to construct a co-expression network for GSE92415. Gene expression profiles were used to calculate correlations between gene pairs, resulting in a neighbor-joining matrix. Hierarchical clustering was then used to group highly correlated genes into co-expression modules. Modules containing more than 50 genes were identified using a hierarchical clustering tree method. These modules were subsequently correlated with external phenotypic data to identify relevant ones. We also identified core genes within each module that may play pivotal roles in specific biological processes or diseases. The gene ontology (GO) database²⁶, developed by the Gene Ontology Consortium, serves as a standardized semantic vocabulary for characterizing and annotating the functions of genes and proteins across various species, with ongoing updates reflecting advances in research. GO annotations are categorized into three main domains: molecular functions (MF), biological processes (BP), and cellular components (CC). Kyoto Encyclopedia of Genes and Genomes (KEGG)^{27–29} is an extensive database that amalgamates genomic, chemical, and systemic functional data. It specializes in housing gene pathway information across diverse species. In this study, enrichment analyses, including GO and KEGG, were conducted on the modular genes using the clusterProfiler package in R. Significantly enriched pathways were determined by $p < 0.05$.

GSEA functional enrichment analysis

The Gene Set Enrichment Analysis (GSEA) method relies on the concept of utilizing a pre-defined set of genes, typically derived from functional annotations or previous experimental findings, to rank genes based on their expression disparities across two sample types³⁰. Subsequently, it evaluates whether this pre-defined gene set exhibits enrichment at either end of the ranked list. In this study, we utilized the clusterProfiler³¹ package to conduct GSEA analysis on the gene expression profile of dataset GSE92415. Specifically, we employed "c2.cp.kegg.v7.4.symbols.gmt" and "c5.go.bp.v7.4.symbols.gmt" as reference gene sets³², with significance defined at a threshold of $p < 0.05$.

Characteristic biomarker identification by machine learning and validation

The Support Vector Machine Recursive Feature Elimination (SVM-RFE) algorithm, developed by Guyon et al. in 2002, is a machine learning approach used to identify tumor-specific genes. This method iteratively removes the least significant feature from a sorted list until the remaining features meet specified criteria³³. SVM is a supervised machine learning method commonly employed for classification and regression tasks. To prevent overfitting, the Recursive Feature Elimination (RFE) algorithm is applied to select the optimal genes from the feature pool. Therefore, SVM-RFE technique is employed to identify the subset of genes with the highest discriminatory capability. Introduced by Tibshirani in 1996, the LASSO algorithm establishes a linear model linking input features and phenotypes by minimizing the sum of least squares loss and the L1 penalty term. LASSO, as a dimensionality reduction method, excels in handling high-dimensional data compared to regression analysis. The SVM-RFE and LASSO algorithms were independently implemented on the GEO training cohort using the R packages "e1071" and "glmnet". The common characteristic genes identified by both SVM-RFE and LASSO were extracted and presented in a Venn diagram using the "venn" package. Subsequently, the expression levels of these characteristic genes underwent validation in an independent validation cohort. To assess the classification performance of key genes between cluster1 and cluster2 samples, ROC curves and AUCs were computed using the "pROC" package in R. Statistical analyses were conducted using R version 4.3.0 and Prism software (GraphPad Prism 8.0, USA).

Evaluation of immune cell infiltration

The CIBERSORT computational approach (<http://cibersort.stanford.edu/>) relies on a deconvolution algorithm for gene expression, enabling the assessment of genetic variation within a sample relative to the entire genome. The CIBERSORT algorithm, developed by Newman et al. in 2015, is a computational tool for deconvoluting expression matrices of human immune cell subtypes³⁴. Based on linear support vector regression principles, it allows for the estimation of immune cell abundances within a given expression matrix³⁴. In our study, we used the expression matrix derived from UC samples in the GSE92415 dataset as input for the CIBERSORT algorithm, enabling the calculation of immune cell abundances. The "vioplot" R package was employed to depict the distributions of the 22 immune cells in Cluster1 and Cluster2. Heatmaps illustrating the quantitative correlations among different immune cells were generated using the "corrplot" package. Additionally, the "ggplot2" R package facilitated the exploration of the relationship between the expression of diagnostic markers and the proportions of immune cells. We conducted statistical comparisons of immune cell infiltration levels across UC subtypes using Wilcoxon rank sum tests and visually represented the results through violin plots. To elucidate the relationship between characteristic gene expression and immune cell infiltration levels, we applied Spearman correlation analysis, and the results are depicted in lollipop plots.

Mouse model of ulcerative colitis

A mouse model of chronic colitis was constructed using DSS, following previously described methods³⁵. Normal control mice were given distilled water, while the experimental group received DSS. After 7 days of DSS treatment, all mice were anesthetized with 1% sodium pentobarbital (3 ml/kg) and euthanized by cervical dislocation. Subsequently, mice were dissected and pertinent tissues, such as the colon, were harvested for subsequent biological, immunological, or molecular experiments. All animal experiments were carried out according to the Guidelines for the Care and Use of Laboratory Animals.

Cell culture

The normal human intestinal epithelial mucosa cell line (NCM460) was obtained from the Chinese Academy of Sciences (Shanghai, China). NCM460 cells were cultured in Gibco's RPMI 1640 medium (Grand Island, NY, USA) supplemented with 10% fetal bovine serum (FBS) from Gibco, and 1% penicillin–streptomycin from Solarbio (Beijing, China). The cell cultures were maintained in a controlled environment, and incubated at 37 °C with 5% CO₂.

Construction of the cellular inflammation model

To establish a model of inflammation in human intestinal epithelial cells, we acquired lipopolysaccharide (LPS) from Selleck.cn (China), consistent with a previous study³⁶. The concentration of LPS utilized in this experiment was 2 µg/ml.

Histological analysis of colitis

The colon tissues were immersed in a 4% paraformaldehyde solution for 36 h and then embedded in paraffin. Subsequently, 5 µm paraffin sections were prepared and subjected to hematoxylin and eosin staining. The extent of inflammation in the colonic tissues was evaluated using a light microscope at both 40× and 100× magnifications.

RNA extraction and RT-qPCR analysis

Samples were used for total RNA extraction using the RNA extraction kit. cDNA was synthesized with a reverse transcription kit (YEASEN, 11141ES10). For quantitative PCR, the cDNA, Hieff UNICON[™] Universal Blue Q-RCR SYBR Green Master Mix (YEASEN), and primers were combined. GAPDH was used as an internal reference gene, and the relative expression levels of the genes were calculated using the $2^{-\Delta\Delta C_t}$ method. The primers used in this study were as follows: IL-1 β , forward: 5'-ATGATGGCTTATTACAGTGGCAA-3', reverse: 5'-GTCCGGA GATTCGTAGCTGGA-3'; IL-6, forward: 5'-ACTCACCTCTTCAGAACGAATTG-3', reverse: 5'-CCATCTTTG GAAGGTTTCAGGTTG-3'; TNF- α , forward: 5'-CCTCTCTCTAATCAGCCCTCTG-3', reverse: 5'-GAGGAC CTGGGAGTAGATGAG-3'; GAPDH, forward: 5'-ACCCAGAAGACTGTGGATGG-3', reverse: 5'-TCAGCT CAGGGATGACCTTG-3'; GLS, forward: 5'-AGGGTCTGTTACCTAGCTTGG-3', reverse: 5'-ACGTTCCGA ATCCTGTAGATTT-3'.

Western blot analysis

Inflammatory and normal colorectal tissue samples were added to RIPA lysis buffer with PMSF and homogenized on ice using a tissue grinder. The supernatant proteins were then extracted by centrifugation after 30 min of lysis and quantified using a BCA protein assay kit (Beyotime, Shanghai, China). Subsequently, 20 μ g of proteins were separated using SDS-polyacrylamide gel electrophoresis and transferred to a polyvinylidene fluoride membrane. Following incubation with 5% skimmed milk for 1 h at room temperature, the membranes were exposed to primary antibodies (anti-GLS, anti-COX2, anti-IL-1 β , anti-IL-10, anti-TNF- α , and anti- β -actin; proteintech, Wuhan, China) at 4 °C overnight (12–16 h). Later, the membranes were incubated with HRP-conjugated secondary antibodies and detected using an ECL chemiluminescence kit (Thermo Fisher). Protein quantification was then performed by normalizing to β -actin using ImageJ software.

Statistical analysis

In this study, we performed all statistical analyses using R software version 4.3.2 and GraphPad Prism 8.0. We used the Wilcoxon rank-sum test to assess for significant differences between the two groups, with a p-value < 0.05 considered statistically significant.

Ethics approval

All animal experiments were conducted in accordance with the ARRIVE guidelines. It was approved by the Ethics Committee of the First Affiliated Hospital of Nanchang University (Ethical number: (2022) CDYFYLK (06-025)). All procedures were performed in accordance with relevant guidelines and regulations.

Results

The flowchart of the study design is presented in Fig. 1.

Identification of GLS expression and its related genes in UC

We identified significantly higher expression of GLS1 in UC patients compared to the healthy population in the GSE92415 dataset (Fig. 2A). The proteins interacting with GLS1 were obtained through protein interaction analysis and are shown in Fig. 2B. Consensus clustering analysis based on GLS1 interaction-related genes was performed to classify patients with UC in the GEO training cohort into two subtypes (Fig. 2C–E). The PCA and t-SNE plot demonstrated complete distinction of 162 UC patients into the two subtypes (Fig. 2F–G). Analysis of the typing and clinical data revealed a higher Mayo score for Cluster 1 and no significant difference in age between Cluster 1 and Cluster 2 (Table 1, Fig. 2H,I). Moreover, the number of non-responding patients was higher in the Cluster 1 group after 6 weeks of placebo or golimumab treatment (Fig. 2J, $P = 0.0076$). The difference in GLS1 expression in normal and UC patients in the GSE87473 dataset was further validated (Fig. 2K). The typing of UC patients using GLS1-related genes resulted in an equal division into two types (Fig. 2L–N). The PCA and t-SNE plot revealed complete division of 106 UC patients into two subtypes (Fig. 2O,P). The UC patients in the Cluster 1 group presented a wider range of intestinal lesions, while those in the Cluster 2 group had a more limited range of intestinal lesions (Fig. 2Q). Additionally, upon validation in dataset GSE72514, the GLS1-associated genes successfully stratified UC patients into two distinct subtypes (Fig. 2R–T). PCA and t-SNE plots demonstrated clear segregation of 97 UC patients into two distinct groups (Fig. 2U,V).

Identification of differential expression genes between C1 and C2 subtypes

A total of 18 differentially expressed genes (DEGs) were identified between the C1 and C2 subtypes, meeting the criteria $|\log_2(\text{fold change})| \geq 2$ and $\text{FDR} < 0.05$, including 7 upregulated genes and 11 downregulated genes. A volcano plot and heatmap of the DEGs and their expressions were generated and presented (Fig. 3A,B). The DEGs were then extracted for enrichment analyses using the DAVID database. Figure 3C depicts the thirty most highly enriched terms for gene ontology (GO) analysis, including biological pathways (BP), cytological components (CC), and molecular functions (MF). The top five enriched BP terms were cell chemotaxis, defense response to bacterium, one-carbon compound transport, defense response to fungus, and antimicrobial humoral response. The three most significant enriched CC terms included basolateral plasma membrane, basal plasma membrane, and basal part of the cell. The eight most enriched MF terms were G protein-coupled receptor binding, receptor ligand activity, RAGE receptor binding, water channel activity, water transmembrane transporter activity, amide transmembrane transporter activity, CCR chemokine receptor binding, and chemokine receptor binding. Figure 3D depicts the thirty most highly enriched terms for Disease Ontology (DO) analysis. The most significant enriched DO terms included intestinal disease, along with diseases such as ulcerative colitis and colitis.

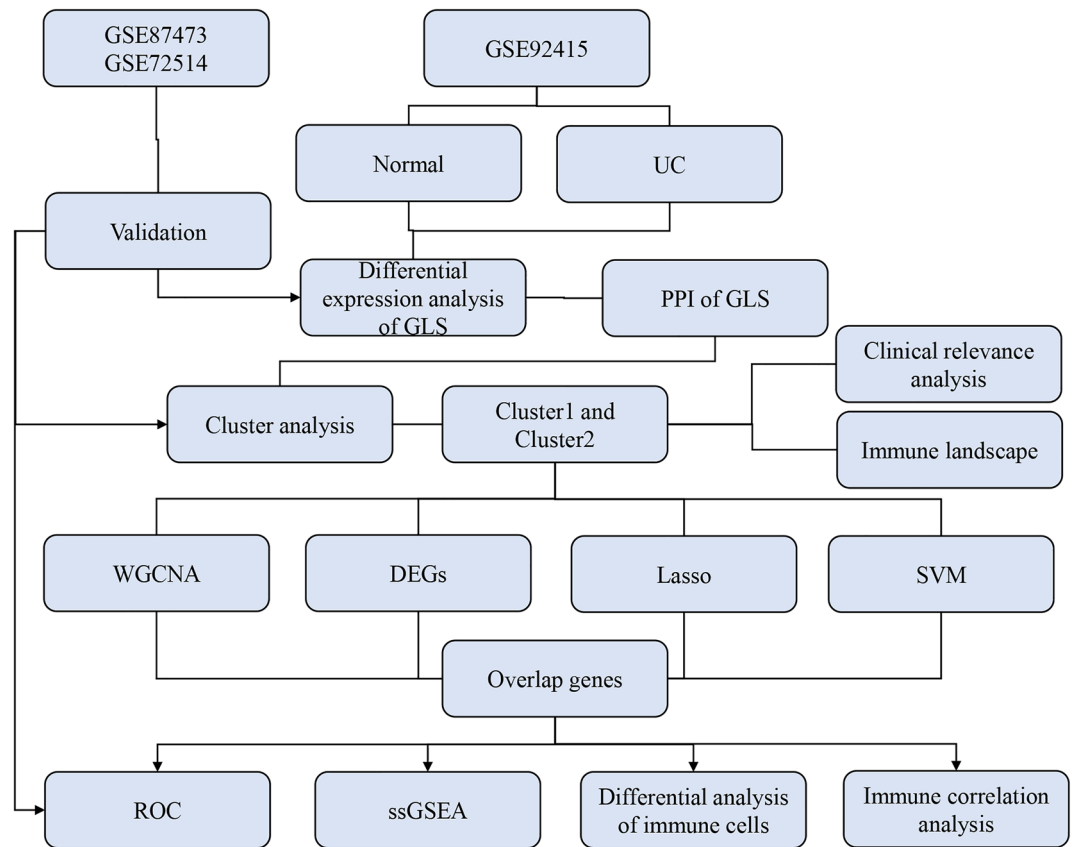


Figure 1. Flowchart of study design.

To assess the differences in signaling pathways between subtypes of UC patients, a gene set enrichment analysis (GSEA) was performed. The results showed that the chemokine signaling pathway, complement and coagulation cascades, cytokine-cytokine receptor interaction, hematopoietic cell lines, and primary immunodeficiencies were activated in all cluster 1 subtypes (Fig. 3E). Meanwhile, butanoate metabolism, citrate cycle TCA cycle, drug metabolism cytochrome p450, metabolism of xenobiotics by cytochrome p450, and retinol metabolism were activated in all cluster 2 subtypes (Fig. 3F).

Characteristic biomarker identification in the UC training cohort

In the GEO training cohort, we identified the most relevant module in the subtype as ME_{grey} through WGCNA analysis (Fig. 4A,B). A total of 225 feature genes were included in the analysis (Fig. 4E). Additionally, the LASSO regression algorithm identified 5 feature genes (Fig. 4C,E), while the SVM-RFE algorithm identified 4 feature genes (Fig. 4D,E). The Venn plot illustrated the overlapping feature genes in UC patients among the 3 algorithms (Fig. 4E; CWH43, HEPACAM2, IL24, PCK1).

Characteristic biomarker validation in the UC training cohort and validation cohort

Among these four signature genes, the expression levels of CWH43, HEPACAM2, and PCK1 were significantly higher in patients with UC subtype C2, and the expression level of IL24 was significantly increased in patients with UC subtype C1 (Fig. 5A–D, $p < 0.0001$). ROC curves were plotted to evaluate the diagnostic accuracy of the CWH43, HEPACAM2, IL24, and PCK1 signature genes in discriminating the two subtypes (Fig. 5E–P). The results in Fig. 5E–H showed that CWH43, HEPACAM2, IL24, and PCK1 could be effective biomarkers for diagnosing UC subtypes (CWH43: AUC = 0.898, 95% CI 0.848–0.944; HEPACAM2: AUC = 0.897, 95% CI 0.848–0.941; IL24: AUC = 0.914, 95% CI 0.870–0.952; PCK1: AUC = 0.907, 95% CI 0.858–0.949). The diagnostic efficacy of CWH43, HEPACAM2, IL24, and PCK1 was further validated in the validation cohort (Fig. 5I–L, CWH43: AUC = 0.901, 95% CI 0.853–0.944; HEPACAM2: AUC = 0.896, 95% CI 0.845–0.941; IL24: AUC = 0.915, 95% CI 0.871–0.952; PCK1: AUC = 0.907, 95% CI 0.857–0.948). Additionally, the diagnostic potential of CWH43, HEPACAM2, IL24, and PCK1 was validated in dataset GSE72514 (Fig. 5M–P, CWH43: AUC = 0.847, 95% CI 0.761–0.915; HEPACAM2: AUC = 0.878, 95% CI 0.808–0.937; IL24: AUC = 0.949, 95% CI 0.900–0.984; PCK1: AUC = 0.817, 95% CI 0.727–0.892).

Immune-infiltrating landscape of UC in training cohort

UC is a chronic inflammatory intestinal barrier disease, and mucosal immune dysregulation may be associated with the pathogenesis of UC³⁷. Therefore, to elucidate the differences in immune cells between C1 and C2

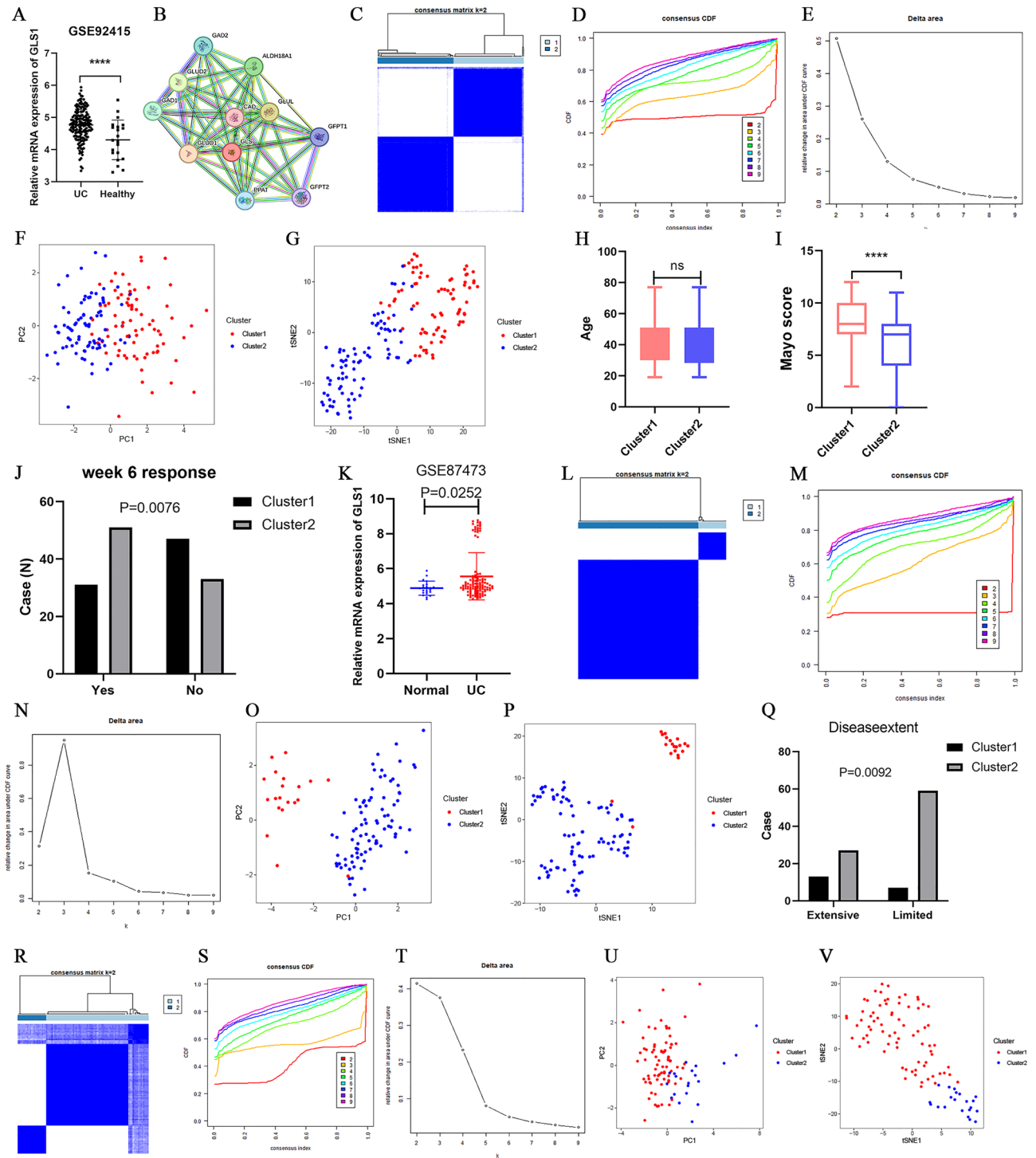


Figure 2. Consensus clustering analysis was performed for UC patients. The expression of GLS was examined in GSE92415 (A) and GSE87473 (K); (B) Schematic diagram of GLS protein interactions; Consensus clustering matrix for the two UC subtypes in GSE92415 (C), GSE87473 (L), and GSE72514 (R); The CDF curves from k=2 to 10 in GSE92415 (D), GSE87473 (M), and GSE72514 (S). The relative change in area under cumulative distribution function (CDF) curve from k=2–10 in GSE92415 (E), GSE87473 (N), and GSE72514 (T); PCA distribution plot for the UC subtype in GSE92415 (F), GSE87473 (O), and GSE72514 (U); t-SNE plot of Cluster1 and Cluster2 in GSE92415 (G), GSE87473 (P), and GSE72514 (V); (H) Relationship between subclusters and age in GSE92415; (I) Relationship between subclusters and Mayo score in GSE92415; (J) Relationship between subclusters and placebo or golimumab 6-week response in GSE92415; (Q) Relationship between subclusters and disease extent in GSE87473.

Characteristic	UC patients		p-value ^b
	Cluster1, N = 78 ^a	Cluster2, N = 84 ^a	
Age	42 ± 14	42 ± 14	0.851
Mayo score	8.00 ± 2.24	6.02 ± 2.82	< 0.001
Treatment			0.620
Golimumab	51 (65.4%)	58 (69.0%)	
Placebo	27 (34.6%)	26 (31.0%)	
Week 6 response			0.008
Yes	31 (39.7%)	51 (60.7%)	
No	47 (60.3%)	33 (39.3%)	

Table 1. Patient demographics and baseline characteristics in different UC subtypes. ^aMean ± SD; n (%). ^bWelch Two sample t-test; Pearson's Chi-squared test.

UC subtypes, we utilized the CIBERSORT algorithm to measure the infiltration of 22 types of immune cells and compared them between the C1 and C2 subtypes in the training set. The composition of the immune cell population in each sample of the C1 subtype and C2 subtype UC was visualized in histograms (Fig. 6A). A weak correlation was observed between the components of different immune cell infiltrations (Fig. 6B). The M2 type of macrophage was found to be slightly more closely related to other immune cells. It was weakly negatively correlated with Neutrophils ($r = -0.47$) and M0 macrophages ($r = -0.39$), and weakly positively correlated with T cells CD8 ($r = 0.4$) and Mast cells resting ($r = 0.43$). The difference in abundance of immune cell infiltration in the intestinal mucosa between the C1 and C2 subtypes in UC is depicted in Fig. 6C. Immune cells with higher levels of infiltration in the intestinal mucosa of patients with C1 UC compared to the C2 group included B cells naïve ($p = 0.017$), T cells CD4 naïve ($p = 0.019$), T cells CD4 memory activated ($p = 0.004$), T cells follicular helper ($p < 0.001$), M0 macrophages ($p < 0.001$), M1 macrophages ($p = 0.018$), activated dendritic cells ($p = 0.001$), and neutrophils ($p < 0.001$). However, the group of patients with C2 UC had higher levels of T cells CD8 ($p < 0.001$), T cells CD4 memory resting ($p < 0.001$), T cells regulatory ($p < 0.001$), Monocytes ($p = 0.019$), M1 macrophages ($p < 0.001$), Dendritic cells resting ($p < 0.001$), Mast cells resting ($p = 0.001$), and Eosinophils ($p < 0.001$).

UC characteristic biomarkers are associated with the tumor immune microenvironment

We investigated the correlation between immune cell infiltration in patients with ulcerative colitis (UC) and the expression levels of CWH43, HEPACAM2, IL24, and PCK1. The expression level of CWH43 was positively correlated with the infiltration levels of macrophage M2, CD8 T cells, resting mast cells, regulatory T cells, eosinophils, monocytes, resting dendritic cells, and memory resting CD4 T cells. Conversely, it demonstrated a negative correlation with naïve CD4 T cells, follicular helper T cells, activated dendritic cells, activated memory CD4 T cells, macrophages M0, macrophages M1, and neutrophils (Fig. 7A, $p < 0.05$). HEPACAM2 expression displayed a positive correlation with the infiltration levels of macrophages M2, memory resting CD4 T cells, regulatory T cells, CD8 T cells, eosinophils, resting mast cells, monocytes, and resting dendritic cells. Conversely, it showed a negative correlation with plasma cells, naïve CD4 T cells, activated dendritic cells, naïve B cells, macrophages M1, activated memory CD4 T cells, follicular helper T cells, macrophages M0, and neutrophils (Fig. 7B, $p < 0.05$). The expression level of IL24 was positively correlated with neutrophils, macrophages M0, follicular helper T cells, activated dendritic cells, naïve B cells, activated memory CD4 T cells, and macrophages M1. Conversely, it exhibited a negative correlation with memory resting CD4 T cells, monocytes, resting dendritic cells, eosinophils, resting mast cells, CD8 T cells, regulatory T cells, and macrophage M2 (Fig. 7C, $p < 0.05$). PCK1 expression level demonstrated a positive association with macrophages M2, CD8 T cells, regulatory T cells, eosinophils, memory resting CD4 T cells, resting dendritic cells, monocytes, resting mast cells, and gamma delta T cells. Conversely, it showed a negative correlation with activated dendritic cells, naïve CD4 T cells, follicular helper T cells, activated memory CD4 T cells, macrophages M1, macrophages M0, and neutrophils (Fig. 7D, $p < 0.05$).

Validation of GLS expression in a cellular model of LPS-induced enteritis and an acute mouse model of colitis

To validate the expression of GLS in UC, we conducted assessments on mouse enteritis tissues. We performed HE staining to confirm the successful modeling of UC in mice, as shown in Fig. 8A. Subsequently, we observed elevated expression of GLS in mouse intestinal inflammatory tissues, as depicted in Fig. 8B. Further validation was carried out in a cellular inflammation model. Figure 8C–E,G illustrate elevated expression of inflammatory factors in the NCM460 inflammatory cell model, confirming the successful establishment of the model. We then reassessed GLS expression, revealing elevated mRNA and protein levels in the NCM460 inflammatory cell model, as seen in Fig. 8F,G.

Discussion

Ulcerative colitis is a chronic inflammatory bowel disease with an unclear etiology³⁸. Treatment effectiveness and patient prognosis for UC vary from person to person according to guidelines. Some patients can maintain remission for an extended period, while others may experience periodic remissions and relapses³⁹. It would be beneficial to classify UC based on specific molecular characteristics in order to predict disease state and prognosis

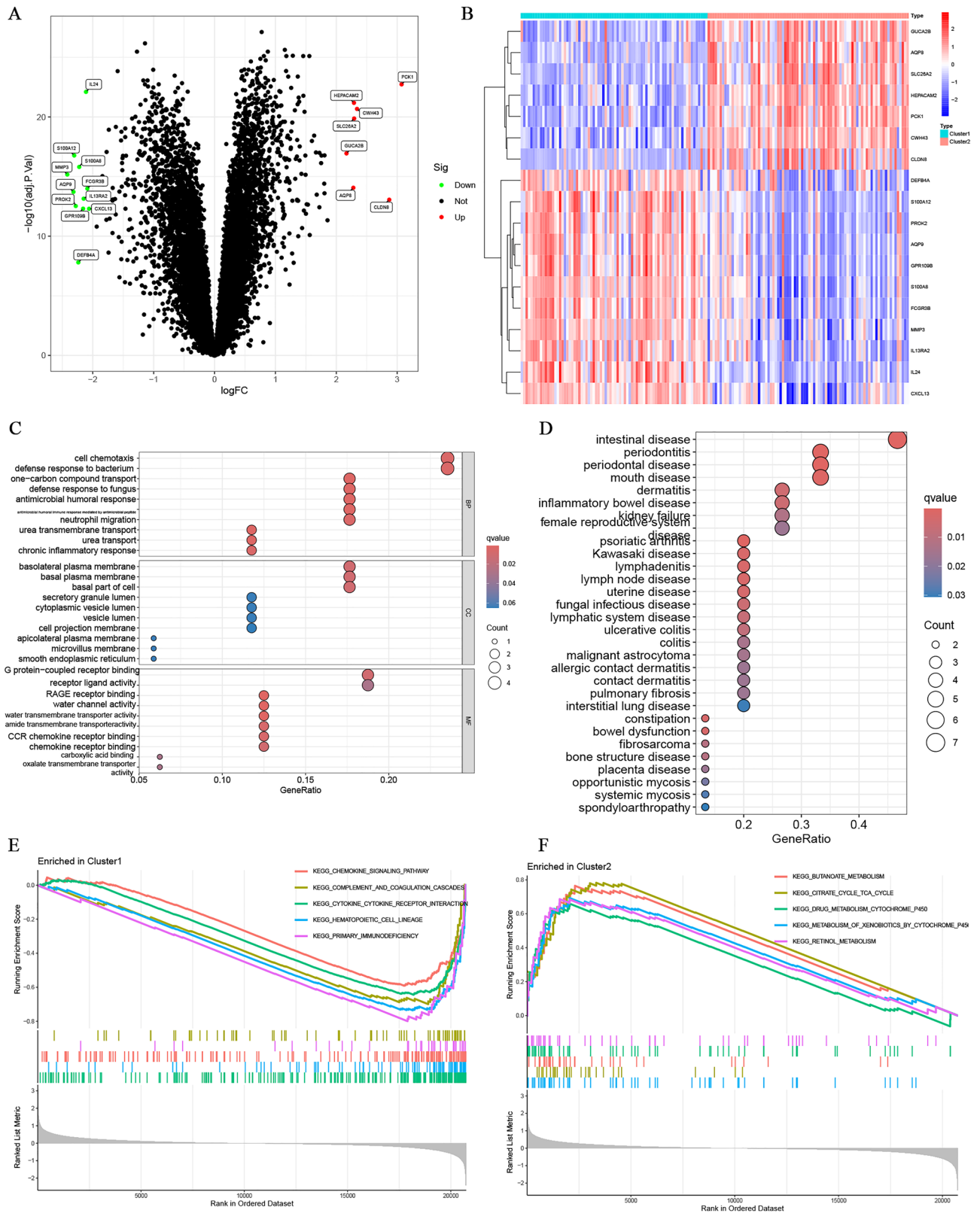


Figure 3. DEGs signature in UC patients. (A) Volcano plot showing differentially expressed genes between Cluster1 and Cluster2; (B) Heatmap of differentially expressed genes; (C) Gene Ontology (GO) analysis of DEGs; (D) Disease Ontology (DO) analysis of DEGs; GSEA between Cluster1 (E) and Cluster2 (F).

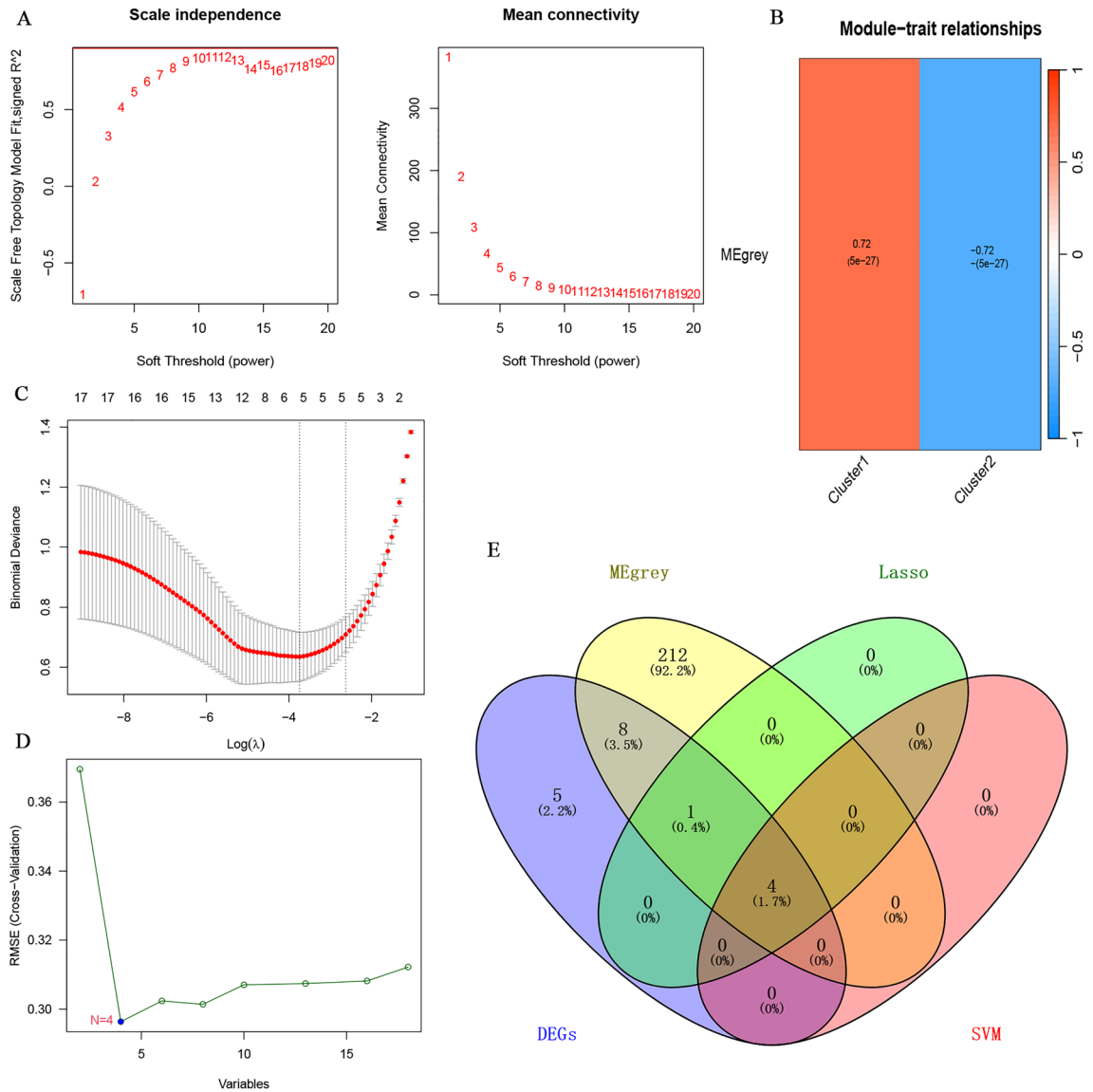


Figure 4. Identification of characteristic biomarkers in the UC training cohort. (A) Analysis of soft thresholding power in WGCNA; (B) Characteristic association of modules; (C) Screening of characteristic biomarkers for UC using the LASSO regression algorithm; (D) Identification of characteristic biomarkers for UC using the SVM-RFE machine-learning algorithm; (E) Intersection of genes among DEGs, WGCNA, LASSO regression algorithm, and SVM-RFE algorithm.

for better patient management. Our study aimed to classify UC patients based on GLS-related genes, identify differences in disease severity and treatment efficacy among different UC subtypes, and identify characteristic biomarkers for clinical guidance in different typologies.

In this study, we acquired the GSE92415 dataset from the GEO database as the training set, comprising transcriptome sequencing data from 162 UC samples and 21 normal colon tissue samples. Additionally, we obtained the GSE87473 dataset as a validation set, encompassing transcriptome sequencing data from 106 UC samples and 21 normal colon tissue samples. GSE72514 dataset, comprises transcriptomic data collected from 97 UC patients' colonic tissue samples, as another validation group. Our analysis revealed a significant up-regulation of GLS in UC samples across both datasets. We identified GLS interaction-related genes through the String database and conducted consistent clustering analysis on UC samples based on these genes, resulting in cluster1 and cluster2. Differential analyses between the two subtypes identified a total of 7 up-regulated and 11 down-regulated DEGs. Enrichment analyses indicated that these DEGs were primarily associated with cell chemotaxis, defense responses against bacteria, and were implicated in intestinal disorders and UC. Furthermore, GSEA analysis of C1 and C2 subtypes revealed distinct associations with crucial biological processes. C1 was linked to pathways such as chemokine signaling and cytokine-cytokine receptor interactions, while C2 showed associations with butyrate metabolism and the citric acid cycle TCA cycle. Chemokines play a pivotal role in various biological responses, including cell polarization, motility, and immune and inflammatory responses⁴⁰.

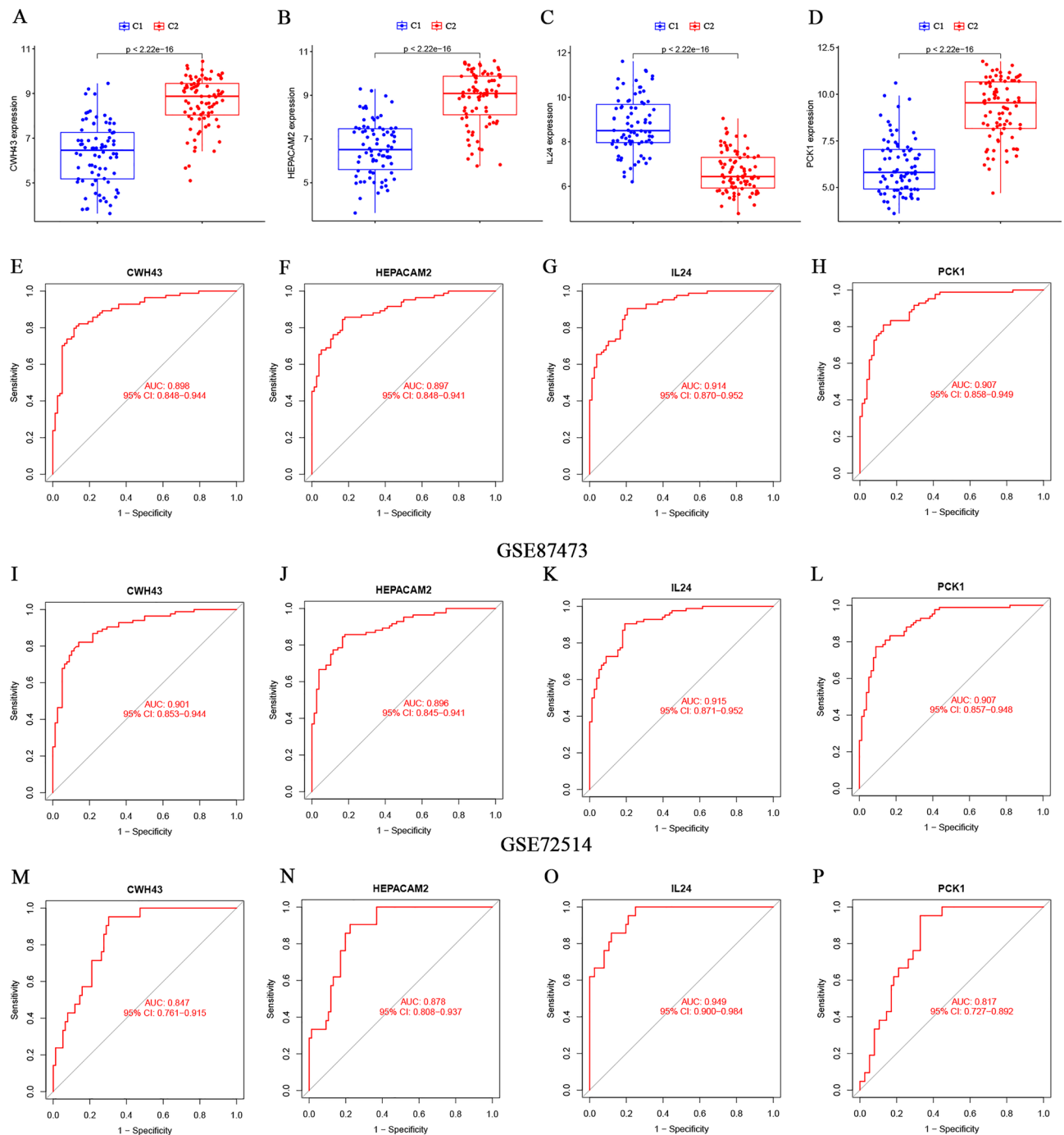


Figure 5. (A–D) The expression levels of the characteristic genes in the training cohort. ROC curves were used to evaluate the diagnostic accuracy of the characteristic genes in both the training cohort (E–H) and the validation cohort (GSE87473 I–L, GSE72514 M–P).

It has been found that chemokine receptors can be used as therapeutic targets for tumors⁴¹. In a study by Chang et al., chemokine ligand 14 (CXCL14) was found to promote metastasis in lung cancer as a prognostic biomarker through an ACKR2-dependent signaling pathway⁴². Notably, dysregulated expression of molecules involved in pro- and anti-inflammatory processes has been associated with UC pathogenesis. For instance, the chemokine CXCL8, a crucial pro-inflammatory factor in UC⁴³, influences the disease through multiple signaling pathways, including PI3K/Akt, MAPKs, and NF- κ B pathways^{44–46}. This dysregulation may contribute to the elevated disease activity and poorer prognosis observed in patients with the C1 subtype of UC. On the other hand, butyrate, a colonic metabolite of carbohydrates, emerged as a significant factor associated with the C2 subtype. Butyrate is considered a major energy source for the colonic mucosa and has been found to inhibit the growth of colorectal cancer⁴⁷. Studies also suggest that butyrate can alleviate enteritis induced by DSS or *Citrobacter* in mice⁴⁸, providing potential explanations for the better prognosis and treatment outcomes observed in patients with C2

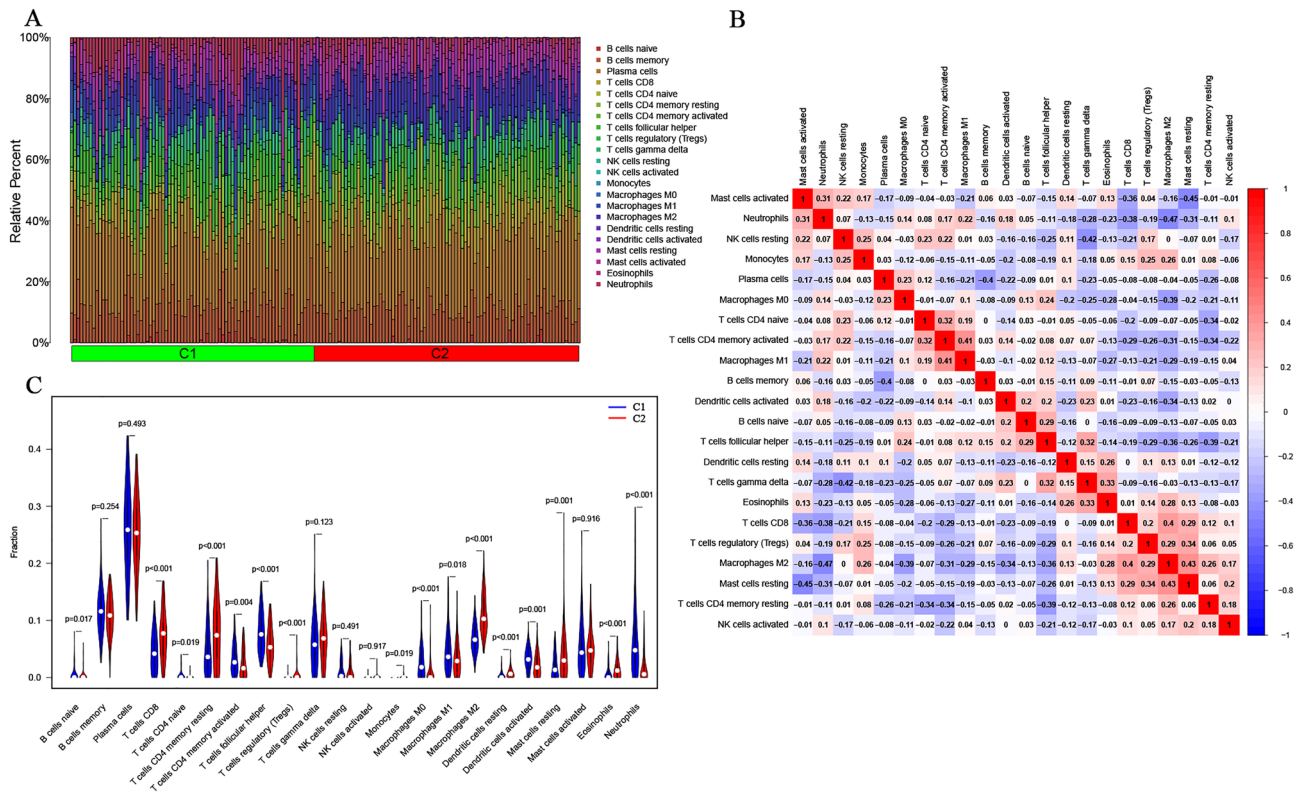


Figure 6. Proportion of immune cells in UC patients and their correlation analysis. (A) Percentage abundances of different immune cells in each sample of UC were determined by CIBERSORT. (B) Correlation matrix of different infiltrating immune cells in UC. The darker the blue, the stronger the negative correlation, and the darker the red, the stronger the positive correlation. (C) Comparison of immune cell infiltration in the intestinal mucosa between C1 subtype UC patients and C2 subtype UC patients. Blue represents the C1 group, and red represents the C2 group.

subtype UC. Therefore, in the future treatment of UC, for UC patients with type C1, whether we can consider the treatment with butyrate and chemokine inhibitors to reduce the patient's disease activity state remains to be further clinically verified.

We utilized WGCNA, SVM-RFE, and LASSO regression machine learning algorithms to identify signature genes for UC, including CWH43, HEPACAM2, IL24, and PCK1. The Cell Wall Biogenesis 43 C-Terminal Homolog (CWH43) promotes colorectal tumor cell growth by enhancing TTK-mediated cell cycle activity⁴⁹. Interestingly, patients with C1 subtype UC displayed higher Mayo scores and were less responsive to the 6-week treatment (Fig. 2I,J). CWH43 expression was lower in C1 subtype UC (Fig. 5A), suggesting a potential link between CWH43 and disease severity, warranting further investigation. HEPACAM2, a member of the HEPACAM family, encodes a protein related to the immunoglobulin superfamily, functioning in mitosis. HEPACAM has been reported to be down-regulated in breast cancer and induces senescence-like growth arrest by increasing the expression levels of senescence-associated proteins p21, p27 and p53⁵⁰. HEPACAM has been reported to be down-regulated in breast cancer and induces senescence-like growth arrest by increasing the expression levels of senescence-associated proteins p21, p27 and p53⁵⁰. It has also been implicated as an oncogene in various tumors, including bladder cancer⁵¹, prostate cancer^{52,53}, renal clear cell carcinoma⁵⁴. Interleukin 24 (IL24), an immunomodulatory cytokine, exhibits cancer-specific inhibitory effects such as angiogenesis inhibition, sensitivity to chemotherapy, and cancer-specific apoptosis⁵⁵. IL24 belongs to the IL-20 subfamily, participating in host defense against bacteria and fungi, tissue remodeling, and wound healing. Elevated IL24 expression has been reported in active UC⁵⁶. In GO functional enrichment analysis, cell chemotaxis, defense response against bacteria, defense response against fungi, and antimicrobial humoral response differed across UC subtypes, corresponding to the results of previous studies (Fig. 3C). Phosphoenolpyruvate carboxykinase1 (PCK1), the first rate-limiting enzyme in gluconeogenesis, plays a crucial role in glucose homeostasis and adipogenesis⁵⁷. PCK1 not only regulates glucose homeostasis but also regulates adipogenesis through activation of sterol regulatory element-binding proteins⁵⁸. Ye et al. found that deficiency of PCK1 leads to metabolic-associated fatty liver disease⁵⁹, and promotes CHK2 O-GlcNAcylation and hepatocellular carcinoma growth in glucose deficiency⁶⁰. In our study, we found that PCK1 expression was decreased in patients with C1 subtype UC, and further studies are needed to determine whether PCK1 is a potential cause of its poorer prognosis. Guidelines recommend fecal calprotectin, fecal lactoferrin, and CRP for activity monitoring and therapeutic decision-making in UC⁶¹. There are no other validated molecular biomarkers for UC typing and activity monitoring. Interestingly, our

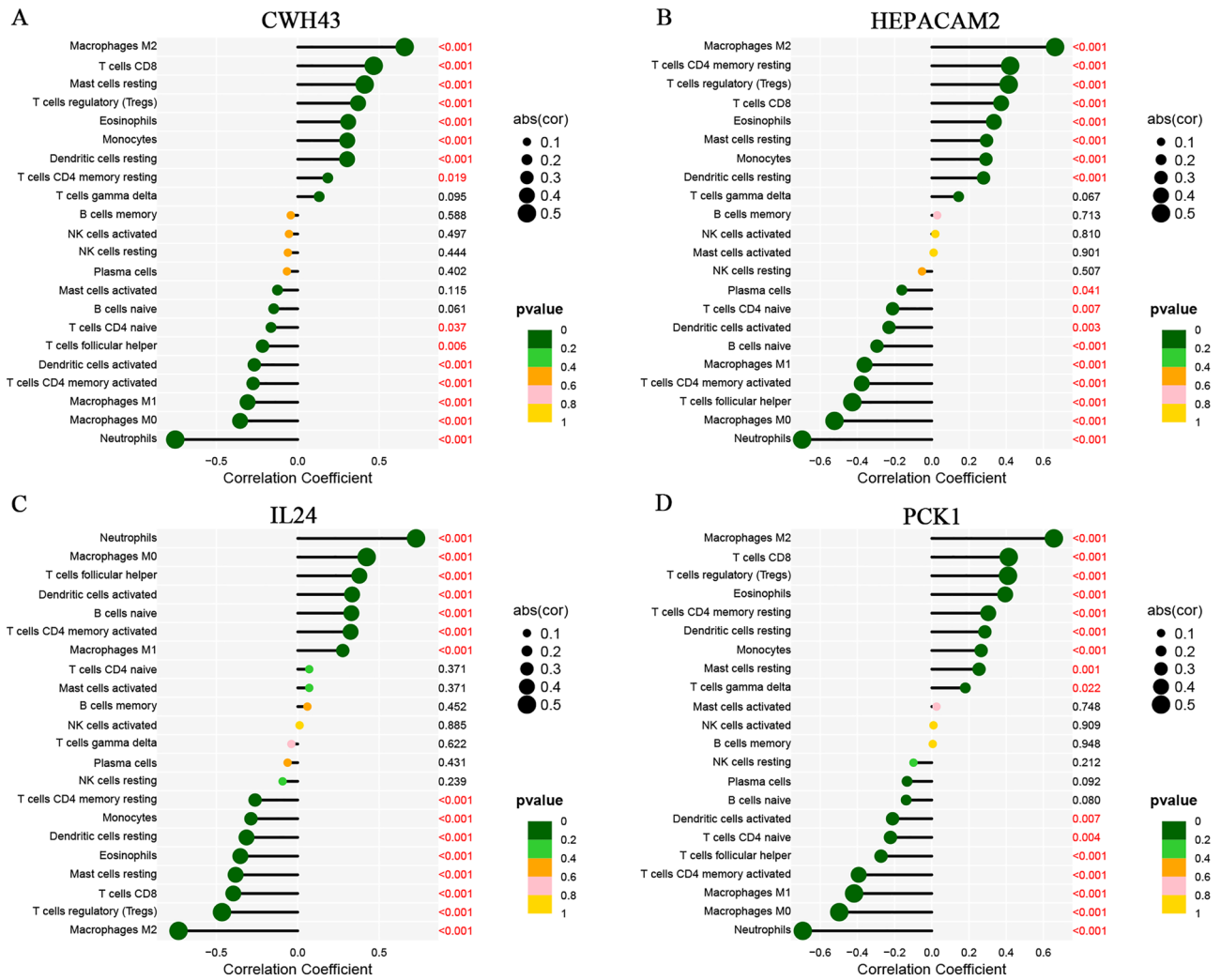


Figure 7. Relationship between the expression levels of characteristic biomarkers and infiltrating immune cells. (A) Correlation between the expression level of CWH43 and infiltrating immune cells. (B) Correlation between the expression level of HEPACAM2 and infiltrating immune cells. (C) Correlation between the expression level of IL24 and infiltrating immune cells. (D) Correlation between the expression level of PCK1 and infiltrating immune cells.

study found that the expression of CWH43, HEPACAM2, IL24, and PCK1 could better identify the 2 subtypes of UC (Fig. 5E–L) and established a link with UC for the first time.

We investigated the relationship between UC signature genes and immune cell infiltration using the CIBERSORT algorithm. Many studies have shown that the immune microenvironment is associated with the pathogenesis of UC⁶². Mitsialis et al. found significant differences in specific immune cell populations in mucosal and blood samples from UC patients and control patients⁶³. Significant changes in CD 8 T cells in UC have been observed by single-cell analysis⁶⁴. In our study, the proportion of seven immune cells was significantly higher in patients with C2 subtype UC than in patients with C1 subtype UC, including M2 macrophages, CD8 T cells, and regulatory T cells. It further suggests the role of CD 8 T cells in different UC subtypes. In addition, there was a positive regulatory relationship between the UC signature genes CWH43, HEPACAM2, and PCK1 and M2 macrophages, CD8 T cells, and regulatory T cells. And there was a negative regulatory relationship between IL24 and these immune cells. It is suggested that CWH43, HEPACAM2, IL24 and PCK1 may be involved in the development of different subtypes of UC through immune regulation. Of course, these speculations require further studies to verify the role of immune response through gene integration regulation in different UC subtypes.

We explored the relationship between UC signature genes and immune cell infiltration using the CIBERSORT algorithm. Numerous studies have demonstrated that the immune microenvironment is linked to the pathogenesis of UC⁵⁵. Mitsialis et al. identified significant differences in specific immune cell populations in mucosal and blood samples from UC patients and control patients⁵⁶. Single-cell analysis has also revealed significant changes in CD8 T cells in UC⁵⁷. In our study, we found that the proportion of seven immune cells was significantly higher in patients with C2 subtype UC compared to patients with C1 subtype UC, including M2 macrophages, CD8 T cells, and regulatory T cells. This suggests a potential role of CD8 T cells in different UC subtypes. Additionally, we identified a positive regulatory relationship between the UC signature genes CWH43,

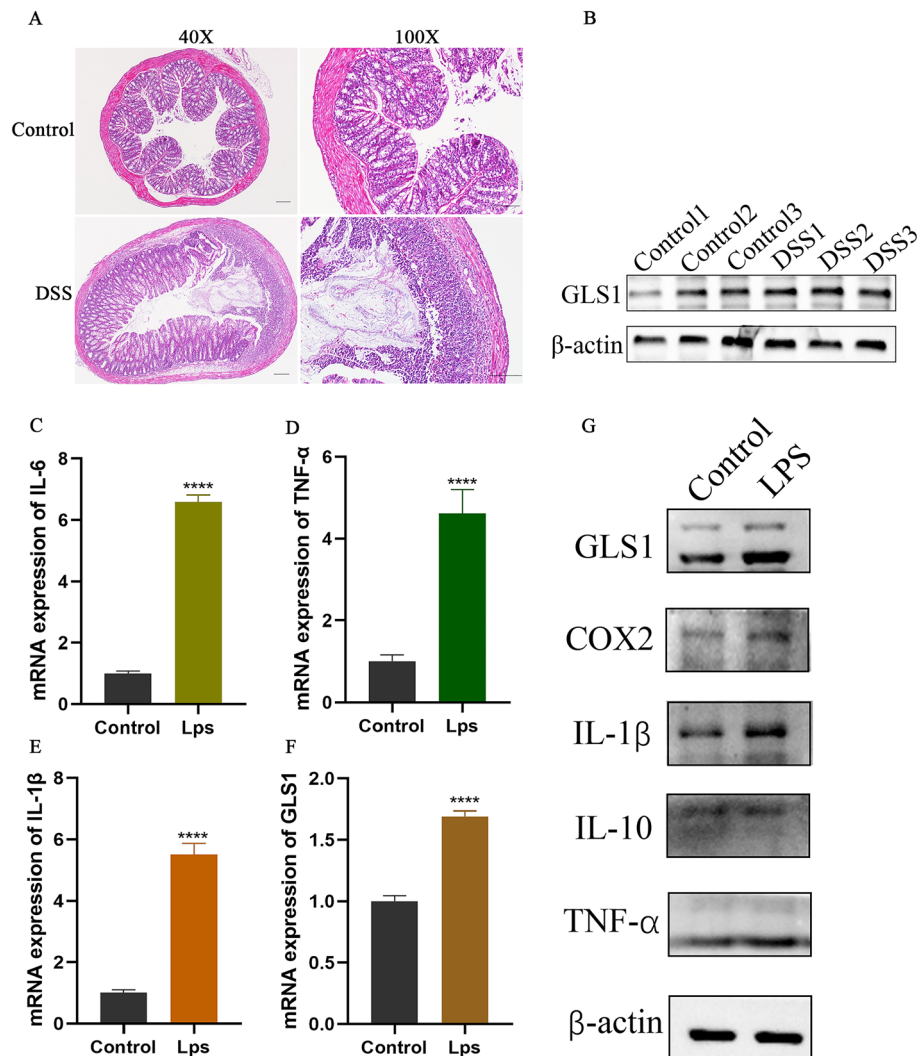


Figure 8. Cell and animal experiments validate GLS1 expression. (A) Hematoxylin and eosin staining of control and acute DSS mouse model intestines; (B) GLS protein expression in the intestines of control and ulcerative colitis mice; mRNA expression of inflammatory factors IL-6 (C), TNF- α (D), IL-1 β (E), and GLS1 (F) in control and cellular inflammation model; (G) Protein expression of inflammatory factors in control and cellular inflammation model.

HEPACAM2, and PCK1 and M2 macrophages, CD8 T cells, and regulatory T cells, while there was a negative regulatory relationship between IL24 and these immune cells. This implies that CWH43, HEPACAM2, IL24, and PCK1 may be involved in the development of different subtypes of UC through immune regulation. However, these speculations require further studies to verify the role of immune response through gene integration regulation in different UC subtypes.

Finally, we examined the expression of the GLS gene using RT-qPCR and WB. The results of the WB showed a significant increase in the expression of GLS in the UC mouse model of enteritis. Similarly, GLS levels were elevated in the enteritis cell model. However, our study has some limitations. Firstly, we were unable to assess the prognostic value of different UC signature genes due to the lack of corresponding clinical information and outcome data in the GEO dataset. Secondly, there was a lack of external experiments to verify the expression and role of UC signature genes in different subtypes and specific mechanisms. To address these limitations, a prospective cohort study could be conducted to assess the prognostic value of the signature genes in predicting disease progression, response to therapy, and patient prognosis through the long-term collection of clinical data and gene expression profiles of UC patients. In addition, *in vitro* and *in vivo* experiments using cell lines, organoids, and animal models or patient samples are used to validate the expression and functional roles of UC signature genes.

Conclusions

By integrating multiple bioinformatics tools and using GLS-related genes to characterize different subtypes, related pathways, and immune cell infiltration in UC tissues, we screened CWH43, HEPACAM2, IL24, and PCK1 as possible candidate diagnostic biomarkers for UC typing. These results will help to understand the pathogenesis of UC and provide new therapeutic tools.

Data availability

The data that endorse the results of this research are available upon reasonable request from the corresponding authors (CYZ and YXC).

Received: 17 January 2024; Accepted: 3 June 2024

Published online: 07 June 2024

References

- Ordás, I., Eckmann, L., Talamini, M., Baumgart, D. C. & Sandborn, W. J. Ulcerative colitis. *Lancet* **380**(9853), 1606–1619 (2012).
- Gros, B. & Kaplan, G. G. Ulcerative colitis in adults: A review. *JAMA* **330**(10), 951–965 (2023).
- Mirsepasi-Lauridsen, H. C. Therapy used to promote disease remission targeting gut dysbiosis, in UC patients with active disease. *J. Clin. Med.* **11**(24), 7472 (2022).
- Yang, L., Venneti, S. & Nagrath, D. Glutaminolysis: A hallmark of cancer metabolism. *Annu. Rev. Biomed. Eng.* **19**, 163–194 (2017).
- Cruzat, V., Macedo Rogero, M., Noel Keane, K., Curi, R. & Newsholme, P. Glutamine: Metabolism and immune function, supplementation and clinical translation. *Nutrients* **10**(11), 1564 (2018).
- de Oliveira, S. R. *et al.* L-Glutamine and physical exercise prevent intestinal inflammation and oxidative stress without improving gastric dysmotility in rats with ulcerative colitis. *Inflammation* **44**(2), 617–632 (2021).
- Martinez-Outschoorn, U. E., Peiris-Pagés, M., Pestell, R. G., Sotgia, F. & Lisanti, M. P. Cancer metabolism: A therapeutic perspective. *Nat. Rev. Clin. Oncol.* **14**(1), 11–31 (2017).
- Jin, J., Byun, J. K., Choi, Y. K. & Park, K. G. Targeting glutamine metabolism as a therapeutic strategy for cancer. *Exp. Mol. Med.* **55**(4), 706–715 (2023).
- Mukha, A., Kahya, U. & Dubrovskaya, A. Targeting glutamine metabolism and autophagy: The combination for prostate cancer radiosensitization. *Autophagy* **17**(11), 3879–3881 (2021).
- Xiao, Z. *et al.* Glutamine deprivation induces ferroptosis in pancreatic cancer cells. *Acta Biochim. Biophys. Sinica* **55**(8), 1288–1300 (2023).
- Yang, L. *et al.* Metabolic shifts toward glutamine regulate tumor growth, invasion and bioenergetics in ovarian cancer. *Mol. Syst. Biol.* **10**(5), 728 (2014).
- van Geldermalsen, M. *et al.* ASCT2/SLC1A5 controls glutamine uptake and tumour growth in triple-negative basal-like breast cancer. *Oncogene* **35**(24), 3201–3208 (2016).
- El Ansari, R. *et al.* Altered glutamine metabolism in breast cancer; subtype dependencies and alternative adaptations. *Histopathology* **72**(2), 183–190 (2018).
- Vanhove, K. *et al.* Glutamine addiction and therapeutic strategies in lung cancer. *Int. J. Mol. Sci.* **20**(2), 252 (2019).
- Pacifico, F., Leonardi, A. & Crescenzi, E. Glutamine metabolism in cancer stem cells: A complex liaison in the tumor microenvironment. *Int. J. Mol. Sci.* **24**(3), 2337 (2023).
- Reeds, P. J. & Burrin, D. G. Glutamine and the bowel. *J. Nutr.* **131**(9 Suppl), 2505S–2508S (2001) (**discussion 2523S–2504S**).
- Mao, J. *et al.* Glutamine deficiency links clindamycin-induced dysbiosis and intestinal barrier dysfunction in mice. *Br. J. Nutr.* **126**(3), 366–374 (2021).
- Jeong, S. Y., Im, Y. N., Youm, J. Y., Lee, H. K. & Im, S. Y. L-Glutamine attenuates DSS-induced colitis via induction of MAPK phosphatase-1. *Nutrients* **10**(3), 288 (2018).
- Giriş, M. *et al.* The effect of heme oxygenase-1 induction by glutamine on TNBS-induced colitis. The effect of glutamine on TNBS colitis. *Int. J. Colorectal Dis.* **22**(6), 591–599 (2007).
- Sido, B., Seel, C., Hochlehnert, A., Breikreutz, R. & Dröge, W. Low intestinal glutamine level and low glutaminase activity in Crohn's disease: A rationale for glutamine supplementation? *Dig. Dis. Sci.* **51**(12), 2170–2179 (2006).
- Zhang, D. *et al.* GLS as a diagnostic biomarker in breast cancer: In-silico, in-situ, and in-vitro insights. *Front. Oncol.* **13**, 1220038 (2023).
- Xiang, L. *et al.* Glutaminase 1 expression in colorectal cancer cells is induced by hypoxia and required for tumor growth, invasion, and metastatic colonization. *Cell Death Dis.* **10**(2), 40 (2019).
- Mukha, A. *et al.* GLS-driven glutamine catabolism contributes to prostate cancer radiosensitivity by regulating the redox state, stemness and ATG5-mediated autophagy. *Theranostics* **11**(16), 7844–7868 (2021).
- Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics (Oxford, England)* **26**(12), 1572–1573 (2010).
- Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
- Gene Ontology Consortium: Going forward. *Nucleic Acids Res.* **43**(Database issue), D1049–1056 (2015).
- Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**(1), 27–30 (2000).
- Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**(11), 1947–1951 (2019).
- Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**(D1), D587–d592 (2023).
- Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. & Mesirov, J. P. GSEA-P: A desktop application for gene set enrichment analysis. *Bioinformatics* **23**(23), 3251–3253 (2007).
- Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *Omic* **16**(5), 284–287 (2012).
- Zhou, M. *et al.* Recurrence-associated long non-coding RNA signature for determining the risk of recurrence in patients with colon cancer. *Mol. Ther. Nucleic Acids* **12**, 518–529 (2018).
- Huang, S. *et al.* Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genom. Proteom.* **15**(1), 41–51 (2018).
- Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**(5), 453–457 (2015).
- Okayasu, I. *et al.* A novel method in the induction of reliable experimental acute and chronic ulcerative colitis in mice. *Gastroenterology* **98**(3), 694–702 (1990).
- Bulek, K. *et al.* Epithelial-derived gasdermin D mediates nonlytic IL-1 β release during experimental colitis. *J. Clin. Investig.* **130**(8), 4218–4234 (2020).
- Kobayashi, T. *et al.* Ulcerative colitis. *Nat. Rev. Dis. Prim.* **6**(1), 74 (2020).
- Du, L. & Ha, C. Epidemiology and pathogenesis of ulcerative colitis. *Gastroenterol. Clin. N. Am.* **49**(4), 643–654 (2020).

39. Feuerstein, J. D. *et al.* AGA clinical practice guidelines on the management of moderate to severe ulcerative colitis. *Gastroenterology* **158**(5), 1450–1461 (2020).
40. Mellado, M., Rodríguez-Frade, J. M., Mañes, S. & Martínez, A. C. Chemokine signaling and functional responses: The role of receptor dimerization and TK pathway activation. *Annu. Rev. Immunol.* **19**, 397–421 (2001).
41. Torphy, R. J., Yee, E. J., Schulick, R. D. & Zhu, Y. Atypical chemokine receptors: Emerging therapeutic targets in cancer. *Trends Pharmacol. Sci.* **43**(12), 1085–1097 (2022).
42. Chang, T. M. *et al.* CXCL14 promotes metastasis of non-small cell lung cancer through ACKR2-dependent signaling pathway. *Int. J. Biol. Sci.* **19**(5), 1455–1470 (2023).
43. Zhu, Y. *et al.* CXCL8 chemokine in ulcerative colitis. *Biomed. Pharmacother. Biomed. Pharmacother.* **138**, 111427 (2021).
44. Ha, H., Debnath, B. & Neamati, N. Role of the CXCL8-CXCR1/2 axis in cancer and inflammatory diseases. *Theranostics* **7**(6), 1543–1588 (2017).
45. Knall, C., Worthen, G. S. & Johnson, G. L. Interleukin 8-stimulated phosphatidylinositol-3-kinase activity regulates the migration of human neutrophils independent of extracellular signal-regulated kinase and p38 mitogen-activated protein kinases. *Proc. Natl. Acad. Sci. U. S. A.* **94**(7), 3052–3057 (1997).
46. Hoffmann, E., Dittrich-Breiholz, O., Holtmann, H. & Kracht, M. Multiple control of interleukin-8 gene expression. *J. Leukoc. Biol.* **72**(5), 847–855 (2002).
47. Kang, X. *et al.* Roseburia intestinalis generated butyrate boosts anti-PD-1 efficacy in colorectal cancer by activating cytotoxic CD8(+) T cells. *Gut* **72**(11), 2112–2122 (2023).
48. Zhou, Z., Cao, J., Liu, X. & Li, M. Evidence for the butyrate metabolism as key pathway improving ulcerative colitis in both pediatric and adult patients. *Bioengineered* **12**(1), 8309–8324 (2021).
49. Lee, C. C. *et al.* CWH43 is a novel tumor suppressor gene with negative regulation of TTK in colorectal cancer. *Int. J. Mol. Sci.* **24**(20), 15262 (2023).
50. Moh, M. C., Zhang, T., Lee, L. H. & Shen, S. Expression of hepaCAM is downregulated in cancers and induces senescence-like growth arrest via a p53/p21-dependent pathway in human breast cancer cells. *Carcinogenesis* **29**(12), 2298–2305 (2008).
51. Du, H. F. *et al.* Expression of hepaCAM inhibits bladder cancer cell proliferation via a Wnt/ β -catenin-dependent pathway in vitro and in vivo. *Cancer Biol. Ther.* **16**(10), 1502–1513 (2015).
52. Deng, Q. *et al.* HepaCAM inhibits cell proliferation and invasion in prostate cancer by suppressing nuclear translocation of the androgen receptor via its cytoplasmic domain. *Mol. Med. Rep.* **19**(3), 2115–2124 (2019).
53. Chen, E. *et al.* Panobinostat reverses HepaCAM gene expression and suppresses proliferation by increasing histone acetylation in prostate cancer. *Gene* **808**, 145977 (2022).
54. Tan, B. *et al.* HepaCAM inhibits clear cell renal carcinoma 786-0 cell proliferation via blocking PKC ϵ translocation from cytoplasm to plasma membrane. *Mol. Cell. Biochem.* **391**(1–2), 95–102 (2014).
55. Persaud, L. *et al.* eIF2 α phosphorylation mediates IL24-induced apoptosis through inhibition of translation. *Mol. Cancer Res. MCR* **15**(8), 1117–1124 (2017).
56. Fonseca-Camarillo, G., Furuzawa-Carballeda, J., Granados, J. & Yamamoto-Furusho, J. K. Expression of interleukin (IL)-19 and IL-24 in inflammatory bowel disease patients: A cross-sectional study. *Clin. Exp. Immunol.* **177**(1), 64–75 (2014).
57. Tuo, L. *et al.* PCK1 negatively regulates cell cycle progression and hepatoma cell proliferation via the AMPK/p27(Kip1) axis. *J. Exp. Clin. Cancer Res. CR* **38**(1), 50 (2019).
58. Xu, D. *et al.* The gluconeogenic enzyme PCK1 phosphorylates INSIG1/2 for lipogenesis. *Nature* **580**(7804), 530–535 (2020).
59. Ye, Q. *et al.* Deficiency of gluconeogenic enzyme PCK1 promotes metabolic-associated fatty liver disease through PI3K/AKT/PDGFR axis activation in male mice. *Nat. Commun.* **14**(1), 1402 (2023).
60. Xiang, J., Chen, C., Liu, R., Gou, D., Chang, L., Deng, H., Gao, Q., Zhang, W., Tuo, L., Pan, X. *et al.* Gluconeogenic enzyme PCK1 deficiency promotes CHK2 O-GlcNAcylation and hepatocellular carcinoma growth upon glucose deprivation. *J. Clin. Investig.* **131**(8) (2021).
61. Singh, S. *et al.* AGA clinical practice guideline on the role of biomarkers for the management of ulcerative colitis. *Gastroenterology* **164**(3), 344–372 (2023).
62. de Souza, H. S. & Focchi, C. Immunopathogenesis of IBD: Current state of the art. *Nat. Rev. Gastroenterol. Hepatol.* **13**(1), 13–27 (2016).
63. Mitsialis, V. *et al.* Single-cell analyses of colon and blood reveal distinct immune cell signatures of ulcerative colitis and Crohn's Disease. *Gastroenterology* **159**(2), 591–608.e510 (2020).
64. Boland, B. S. *et al.* Heterogeneity and clonal relationships of adaptive immune cells in ulcerative colitis revealed by single-cell analyses. *Sci. Immunol.* **5**(50), 4432 (2020).

Author contributions

Y.X., J.L., and Q.T. carried out the experiments. Y.X., C.Y.Z., and Y.X.C. analyzed and interpreted the data, and wrote the manuscript. Y.H.W., Z.D.L., and other authors reviewed the manuscript. Y.X. and Q.T. revised critically of the manuscript. Y.X., C.Y.Z., and Y.X.C. designed and supervised the study.

Funding

The study was supported by grants of Jiangxi Clinical Research Center for Gastroenterology (Grant No. 20223BCG74011, PI: Youxiang Chen); This study was supported by grants from the National Natural Science Foundation of China (Grant No. 82060448 and 82360112, PI: Youxiang Chen) and 2023 clinical research program of the first affiliated hospital of Nanchang University (PI: Chunyan Zeng).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-63891-2>.

Correspondence and requests for materials should be addressed to Y.C. or C.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024