



OPEN

Exploring the butyrate metabolism-related shared genes in metabolic associated steatohepatitis and ulcerative colitis

Beiyong Deng^{1,2,4}, Yinghui Liu^{3,4}, Ying Chen^{1,2,4}, Pengzhan He^{1,2}, Jingjing Ma³, Zongbiao Tan¹, Jixiang Zhang¹ & Weiguo Dong¹✉

Metabolic-associated steatohepatitis (MASH) and ulcerative colitis (UC) exhibit a complex interconnection with immune dysfunction, dysbiosis of the gut microbiota, and activation of inflammatory pathways. This study aims to identify and validate critical butyrate metabolism-related shared genes between both UC and MASH. Clinical information and gene expression profiles were sourced from the Gene Expression Omnibus (GEO) database. Shared butyrate metabolism-related differentially expressed genes (sBM-DEGs) between UC and MASH were identified via various bioinformatics methods. Functional enrichment analysis was performed, and UC patients were categorized into subtypes using the consensus clustering algorithm based on sBM-DEGs. Key genes within sBM-DEGs were screened through Random Forest, Support Vector Machines-Recursive Feature Elimination, and Light Gradient Boosting. The diagnostic efficacy of these genes was evaluated using receiver operating characteristic (ROC) analysis on independent datasets. Additionally, the expression levels of characteristic genes were validated across multiple independent datasets and human specimens. Forty-nine shared DEGs between UC and MASH were identified, with enrichment analysis highlighting significant involvement in immune, inflammatory, and metabolic pathways. The intersection of butyrate metabolism-related genes with these DEGs produced 10 sBM-DEGs. These genes facilitated the identification of molecular subtypes of UC patients using an unsupervised clustering approach. ANXA5, CD44, and SLC16A1 were pinpointed as hub genes through machine learning algorithms and feature importance rankings. ROC analysis confirmed their diagnostic efficacy in UC and MASH across various datasets. Additionally, the expression levels of these three hub genes showed significant correlations with immune cells. These findings were validated across independent datasets and human specimens, corroborating the bioinformatics analysis results. Integrated bioinformatics identified three significant biomarkers, ANXA5, CD44, and SLC16A1, as DEGs linked to butyrate metabolism. These findings offer new insights into the role of butyrate metabolism in the pathogenesis of UC and MASH, suggesting its potential as a valuable diagnostic biomarker.

Keywords Metabolic associated fatty liver disease, Ulcerative colitis, Butyrate metabolism, Biomarkers, Machine learning

Inflammatory bowel disease (IBD) and metabolic-associated fatty liver disease (MAFLD) present a significant global burden, impacting millions of individuals worldwide and leading to substantial morbidity, mortality, and healthcare costs^{1,2}. In recent decades, MAFLD has become the most prevalent chronic liver disorder, with an estimated global adult prevalence of approximately 25%³. Simultaneously, IBD has remained a common chronic gastrointestinal condition, with both its incidence and prevalence continuing to increase^{4,5}.

¹Department of Gastroenterology, Renmin Hospital of Wuhan University, No 99 Zhangzhidong Road, Wuhan 430060, Hubei Province, China. ²Central Laboratory, Renmin Hospital of Wuhan University, Wuhan, China. ³Department of Geriatric, Renmin Hospital of Wuhan University, Wuhan, China. ⁴These authors contributed equally: Beiyong Deng, Yinghui Liu and Ying Chen. ✉email: dongweiguo@whu.edu.cn

Despite the elusive precise common pathogenesis and causal relationship between MAFLD and IBD, substantial evidence supports potential physiological and pathological interconnections between these conditions^{6–8}. Variations in results from diverse population data and multicenter studies notwithstanding, research consistently indicates a link between IBD and MASH, a more severe form of MAFLD⁹. Specifically, the incidence rates of MASH development in IBD patients and the incidence rates of IBD development in MASH patients are significantly higher than those in the general population¹⁰. This trend persists even when considering statistically significant variations in incidence rates or risk ratios for MASH development between Crohn's disease (CD) and ulcerative colitis (UC) patients^{11,12}. Additionally, evidence suggests that MASH mice induced by a high-fat diet are more susceptible to colitis development, resulting in notably higher mortality rates compared to the control group¹³. Comprehensive reviews have summarized shared risk factors between MAFLD and IBD, such as obesity, insulin resistance, and metabolic syndrome. Other contributing factors may include systemic inflammation, the use of IBD therapeutic drugs with potential hepatotoxicity, gut microbiota dysbiosis, bile acid metabolism, and intestinal barrier dysfunction^{9,14,15}. However, despite this existing knowledge, the underlying mechanisms linking IBD and MASH comorbidity remain unclear, posing a significant challenge in managing patients with this combined condition.

The role of butyrate in linking IBD and MAFLD has garnered significant attention due to its potential impact on dysregulated gut microbiota and its metabolites, which lead to inflammation and immune infiltration in both conditions¹⁶. Butyrate, a short-chain fatty acid, has shown beneficial effects in managing UC by promoting colonic motility, enhancing intestinal blood flow, improving intestinal mucosal barrier function, and modulating immune responses^{17–19}. Additionally, butyrate demonstrates anti-inflammatory effects in MASH by preserving gut barrier function, regulating gut microbiota, modulating lipid metabolism, and alleviating insulin resistance^{20,21}. These combined effects suggest that butyrate may play a crucial role in mitigating the common pathophysiology of UC and MASH.

Investigating the pathways of butyrate metabolism in both UC and MASH is therefore essential for understanding their common pathogenesis and clinical implications. This study identified ANXA5, CD44, and SLC16A1 as critical genes related to butyrate metabolism and explored their correlation with immune infiltration. The results were confirmed in the validation cohort and further validated through human samples. Notably, this is the first study to elucidate the common genetic characteristics and molecular mechanisms between UC and MASH, offering a new perspective on the intrinsic link between these two conditions.

Methods

GEO dataset selection

Colonic mucosa biopsy samples from UC patients and liver tissue samples from MASH patients were obtained from the GEO database (<http://www.ncbi.nlm.nih.gov/geo>). After removing outlier samples and genes with low expression and abundance, the "ComBat" algorithm from the "sva" R package was applied to minimize batch effects²². By integrating datasets GSE61260 and GSE63067, a consolidated cohort named "MASH MergeCo" was established. For a more comprehensive exploration of the mechanistic interplay between MASH and UC, the independent validation sets GSE213621 and GSE87466 were identified. Additionally, a comprehensive collection of 882 butyrate metabolism-related genes with relevance scores greater than 5 was compiled from GeneCards (<https://www.genecards.org>).

All liver biopsy specimens were staged according to the published NASH Clinical Research Network scoring system²³. For the analyses, fibrosis stages 1a, 1b, and 1c were combined and treated as stage 1 fibrosis.

Weighted gene co-expression network analysis (WGCNA)

The "WGCNA" R package facilitated WGCNA and the identification of gene modules associated with UC²⁴. This process involved assessing scale independence and average connectivity across various power values using the "pickSoftThreshold" function, ultimately selecting the optimal soft threshold β . Hierarchical clustering grouped genes with similar expression patterns into cohesive modules. Dynamic tree cutting segmented and recognized co-expression modules, followed by merging modules with similar expression profiles. Pearson correlation analysis was then performed to determine the relationship between modules and clinical attributes.

Identification and enrichment analyses of differentially expressed genes (DEGs)

DEGs were evaluated using Wilcoxon's rank-sum test, and potential shared genes were identified through Venn diagrams. An intersection analysis of butyrate metabolism-related genes with UC-related modules, UC-DEGs, and MASH-DEGs was performed, defining the resulting genes as shared butyrate metabolism-related DEGs (sBM-DEGs). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted to gain a comprehensive understanding of the biological mechanisms underpinning the identified DEGs²⁵. These analyses utilized the R packages "clusterProfiler" and "org.Hs.eg.db" with a set threshold of q -value 0.05 to ensure robust results²⁶.

Assessment of immune infiltration patterns

Enrichment levels of immune cell infiltration were computed using the ssGSEA method²⁷. The Student's t -test validated distinctions between the two groups, and results were graphically depicted using R's "ggboxplot" package. The "corrplot" package investigated the relationship between hub genes and immune cells, while the "pheatmap" package visualized the findings.

Identification of transcription factors (TFs)

Information from the STRING database (<https://string-db.org>) was used to create a human protein–protein interaction (PPI) network. The MCC algorithm was then applied using the cytoHubba plug-in in Cytoscape (version 3.9.1) to identify the top ten hub genes within the PPI network. The Transcriptional Regulatory Relationships Unraveled by Sentence-based Text Mining (TRRUST) v2 database (<https://www.grnpedia.org/trrust/>) provided regulatory insights into TFs and their target genes. A network demonstrating TF–target interactions was created using the discovered TF–target regulatory linkages.

Unsupervised clustering and functional enrichment analysis of patients with MASH

The "ConsensusClusterPlus" package was employed for unsupervised clustering analysis²⁸. The k-means algorithm was used with a maximum subtype number of k (k = 6), completing 1,000 iterations. The ideal subtype number was determined based on the analysis of the cumulative distribution function (CDF) curve, consensus matrix, and consistent cluster score.

The "GSVA" package was utilized to perform a GSVA enrichment analysis to compare the enriched pathways and functionalities of various butyrate metabolism-related subtypes²⁹. Two gene sets from the MSigDB database, "c2.cp.kegg.v7.4.symbols" and "c5.go.bp.v7.5.1.symbols," were used as input files. By comparing the GSVA scores between various subtypes, differential enrichment functions and pathways were identified using the "limma" R package. Functions and pathways with a |t| value of the GSVA score greater than two were considered notably enhanced.

Identification of signature genes by machine learning

Potential biomarkers were identified by further screening the sBM-DEGs using three machine learning algorithms: Light Gradient Boosting Model (LightGBM), Support Vector Machine-Recursive Feature Elimination (SVM-RFE), and Random Forest (RF). The "DALEX" package facilitated model interpretation, enabling the calculation of both local and global feature importance to enhance model understanding. The predictive accuracy of the feature genes identified by these machine-learning techniques was evaluated using the validation set, with results visually represented by ROC curves.

Immunohistochemistry (IHC)

Colon biopsies were obtained from patients with UC and healthy controls (HC), while liver biopsies were derived from patients with MASH and HC. Samples were sourced from the Department of Pathology, Renmin Hospital of Wuhan University. Paraffin-embedded slices underwent deparaffinization and rehydration using slide warmers and alcohol. Antigen retrieval was performed with citrate buffer for 60 min. To inhibit endogenous peroxidase activity, all sections were treated with 0.3% hydrogen peroxide for 10 min. A blocking step with 5% bovine serum albumin (BSA) was conducted before coating sections with primary antibodies, followed by overnight incubation at 4 °C. After washing, sections were incubated with a secondary antibody for 20 min at room temperature. Diaminobenzidine (DAB) staining was carried out using a DAB kit (Servicebio), and sections were counterstained with hematoxylin. Stained sections were examined using light microscopy. Primary antibodies targeting ANXA5 (Cat No. 11060-1-AP), CD44 (Cat No. 15675-1-AP), and SLC16A1 (Cat No. 20139-1-AP) were used.

The study adhered to the Declaration of Helsinki and received approval from the Ethics Committee of Renmin Hospital of Wuhan University (WDRY2023-K151 and WDRY 2022-K130). Informed consent was obtained from all patients, and the research was conducted in accordance with relevant guidelines and regulations.

Statistical analysis

All statistical analyses for this study were conducted using R (version 4.2.3) and GraphPad Prism (version 8.3.0). Statistical significance was defined as a *p*-value or adjusted *p*-value of less than 0.05, based on two-sided tests.

Result

Identification of DEGs between UC patients and HC

The UC microarray dataset GSE75214 was obtained from the GEO database. Using criteria of an adjusted *p*-value of 0.05 and an absolute mean difference greater than 1, a total of 1124 DEGs were identified, with 715 upregulated and 409 downregulated (Fig. 1A,B). All samples from GSE75214 underwent clustering and were included in the WGCNA. An optimal soft threshold power β of 24 was determined, yielding a scale-free R^2 of 0.85 (Fig. 1C). Dynamic tree cutting revealed 7 distinct modules after merging those with similar expression profiles (Fig. 1D,E). Among these, the turquoise and blue modules exhibited the most significant correlation with UC (Fig. 1E–G) (Table S1).

Identification of shared DEGs between MASH and HC

The principal component analysis plot of the MASH MergeCo successfully distinguished between the MASH patients and HC, confirming effective batch effect correction (Fig. 2A). Comparing HC to MASH patients revealed 159 DEGs, with Benjamini–Hochberg adjusted *p*-values of 0.05 and an absolute mean difference greater than 0.5, including 117 upregulated and 42 downregulated DEGs (Fig. 2B,C). Additionally, Venn diagram analysis identified 49 shared DEGs among MASH-DEGs, UC-DEGs, and the UC-related module (Fig. 2D) (Table S2).

Functional enrichment analysis and identified butyrate metabolism-related shared DEGs

Functional enrichment analysis was conducted to elucidate the putative biological functions of the shared DEGs. GO analysis revealed high enrichment in processes related to immune receptor activity, C–C chemokine receptor

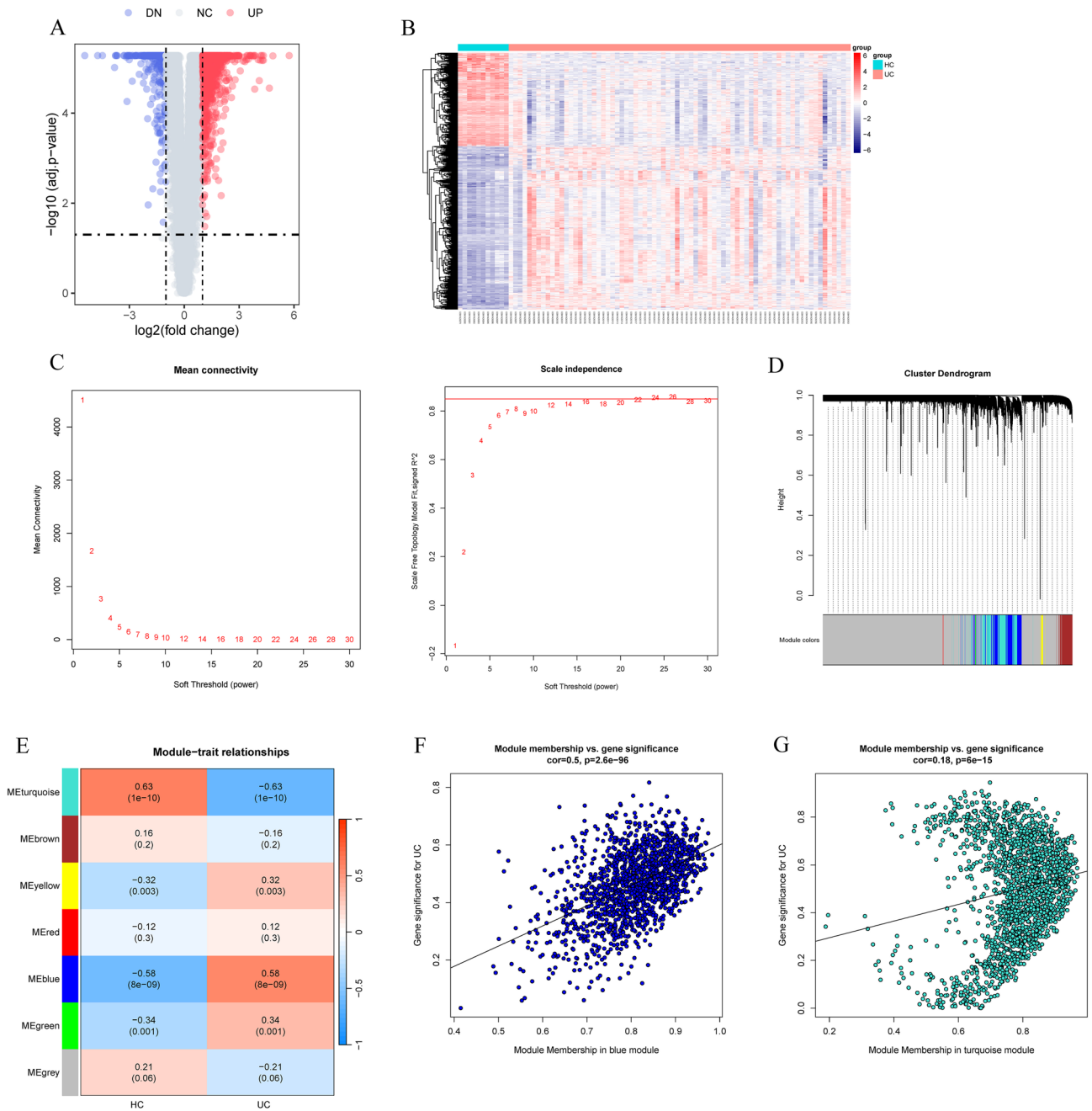


Figure 1. (A) Volcano plot illustrating DEGs in colon tissue between UC patients and HC. (B) Heatmap depicting gene expression profiles in UC patients and HC. (C) Analysis of scale-free index and mean connectivity for soft-threshold power determination. (D) Module clustering dendrogram based on dissimilarity measure (1-TOM). (E) Heatmap revealing a correlation between module eigengenes and UC. (F, G) Scatter plot of the GS for UC vs. the MM in the blue (F) and turquoise (G) modules. DEG, differentially expressed genes. UC, ulcerative colitis. HC, healthy controls. TOM, topological overlap matrix. GS, gene significance. MM, module membership.

activity, cell chemotaxis, and leukocyte chemotaxis (Fig. 3A). Additionally, KEGG analysis highlighted shared DEGs connected to inflammatory pathways, including the chemokine signaling pathway, IL-17 signaling pathway, intestinal immune network for IgA production, ECM-receptor interaction, PPAR signaling pathway, and TNF signaling pathway (Fig. 3B).

Gut microbiota alterations are thought to play a role in the disease process of IBD combined with MAFLD³⁰. IBD and MAFLD displayed similar changes in gut microbiota composition³¹. Changes in the composition of the intestinal flora in IBD patients lead to changes in the microbial metabolites, such as butyric acid, which can further aggravate damage to the intestinal barrier and thus may affect the liver through leakage through the damaged intestinal barrier, leading to MAFLD³¹. Given the pivotal role of the butyrate metabolic pathway in both MASH and UC pathogenesis, its mechanistic involvement in the co-pathogenesis of UC and MASH was

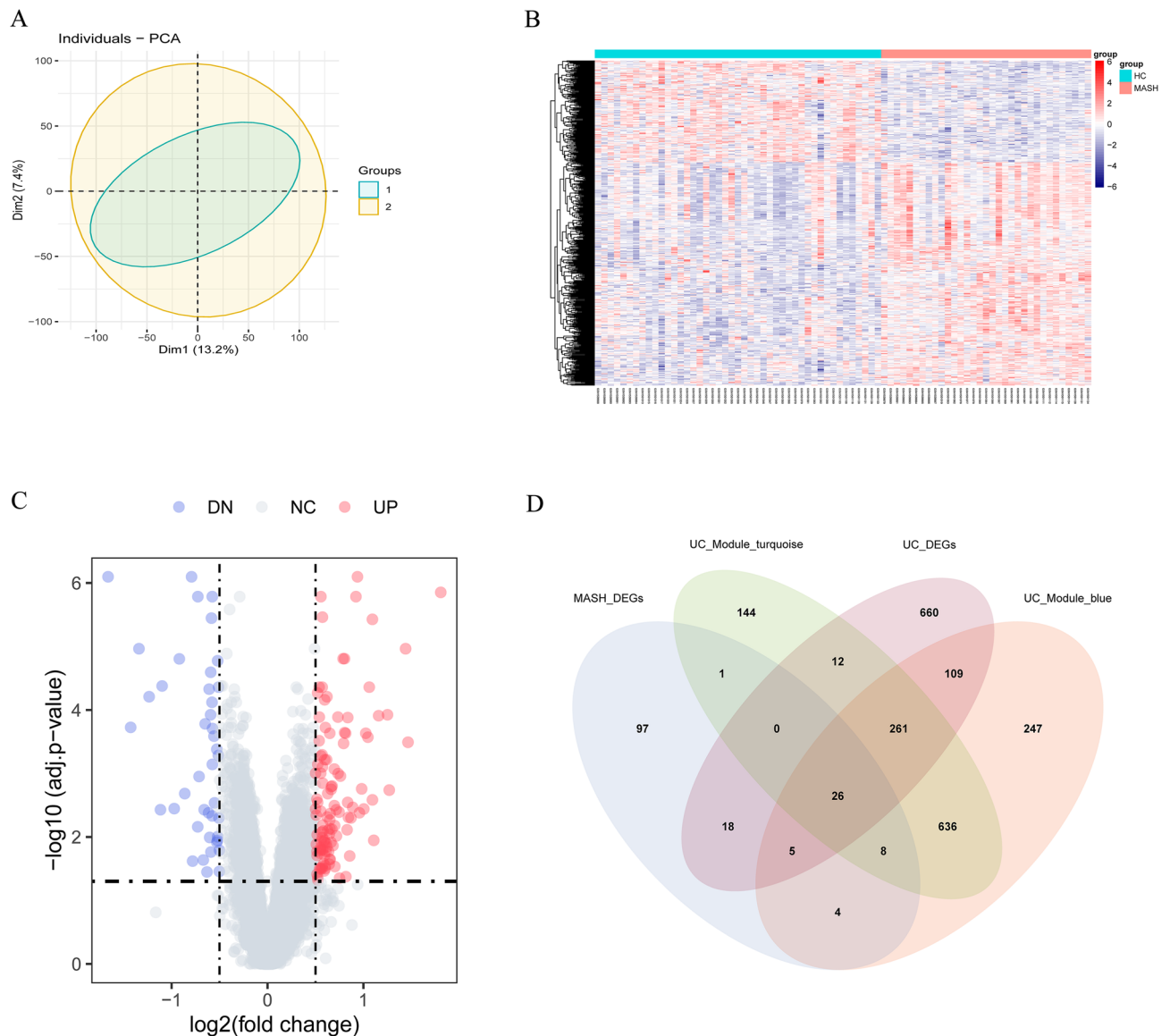


Figure 2. (A) PCA of two datasets after batch effect correction. (B) Heatmap depicting gene expression profiles in MASH patients and HC. (C) Volcano plot showcasing DEGs in liver tissue between MASH patients and HC. (D) Venn diagram indicating overlapping genes among MASH-DEGs, UC-DEGs, and UC-related modules.

investigated. By intersecting butyrate metabolism-related genes with the shared DEGs, a set of 10 sBM-DEGs was identified (Fig. 3C).

Identification of TFs

The preceding analysis revealed a mutual correlation among all 10 sBM-DEGs (Fig. 4A). Given that co-expressed genes often share common regulatory TFs, a TF enrichment analysis was conducted using the TRRUST database to identify key regulatory genes. This analysis identified SNAI2, ERG, IKKKB, HDAC1, ETS1, KLF5, USF1, CREB1, SPI1, and RELA as the top 10 significantly enriched TFs among the 10 sBM-DEGs. Notably, elevated colonic expression of SNAI2, ERG, ETS1, USF1, CREB1, and RELA was observed. In contrast, KLF5 expression was lower in UC patients compared to HC (Fig. 4B).

Development of UC molecular subtypes using unsupervised clustering

To illustrate the butyrate metabolism-related patterns of UC, an unsupervised cluster analysis was conducted on 74 UC patients using the "ConsensusClusterPlus" R package, based on the expression patterns of 10 sBM-DEGs. Setting the cluster number to 2 yielded subgroup consistency scores greater than 0.8, supported by stable results from the CDF plot, delta area plot, tracking plot, and consensus matrix (Fig. 5A–D). Consequently, the 74 UC patients were categorized into two distinct clusters: molecular subtype 1 (n = 69) and molecular subtype 2 (n = 5) (Fig. 5E).

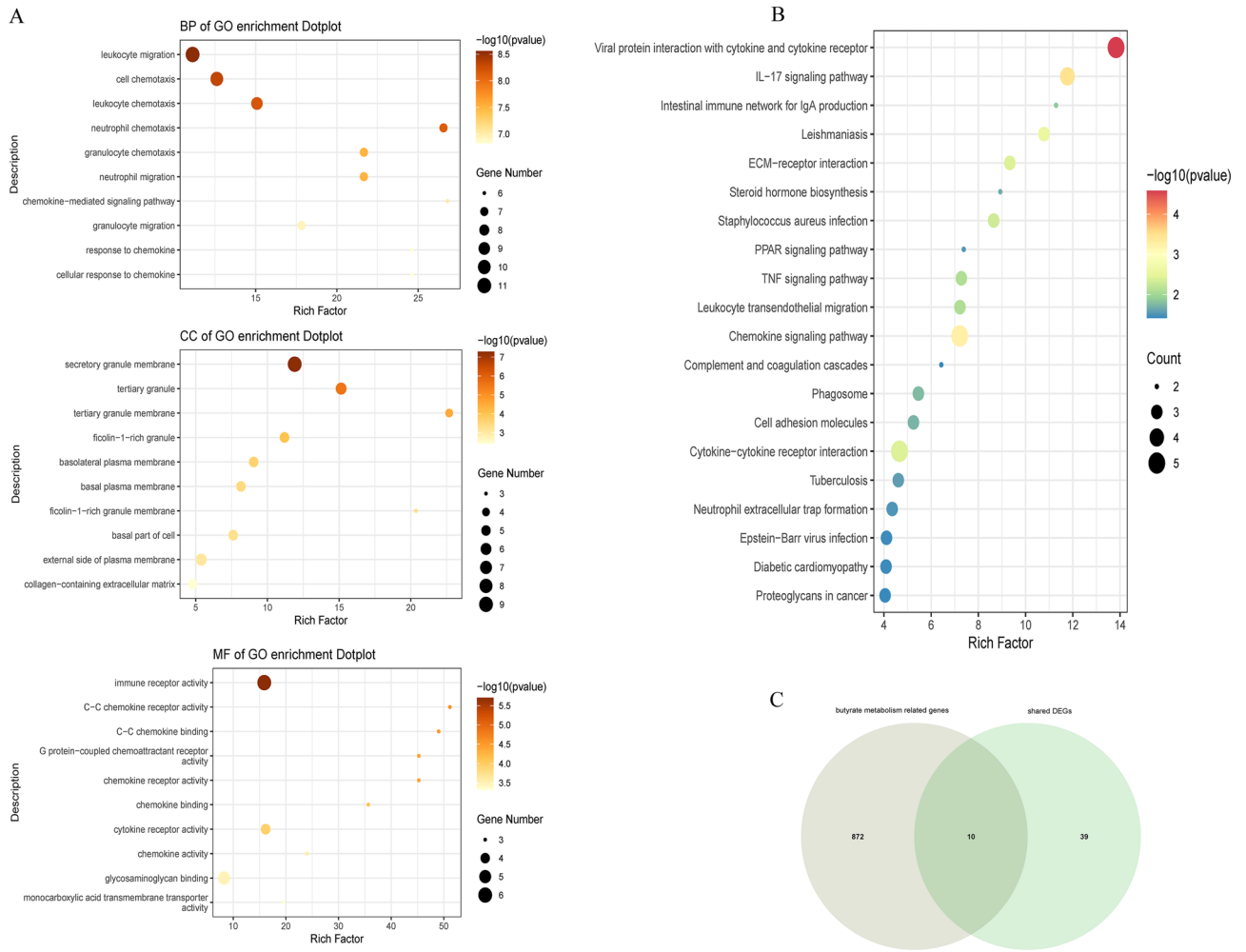


Figure 3. (A) GO enrichment analysis of shared DEGs. (B) KEGG pathway enrichment analysis of shared DEGs. (C) Venn diagram illustrating overlapping genes between butyrate metabolism-related genes and shared DEGs. GO, Gene Ontology. KEGG, Kyoto Encyclopedia of Genes and Genomes.

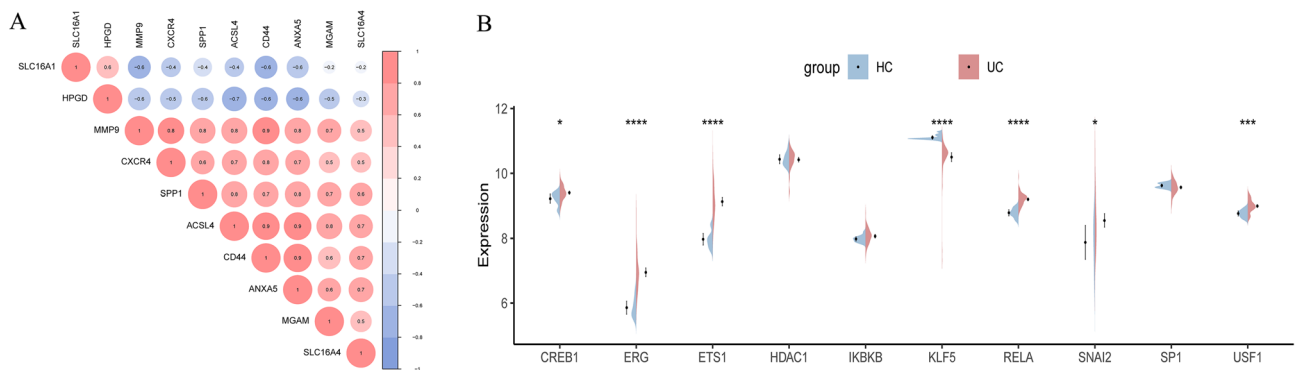


Figure 4. (A) Correlation plot of sBM-DEGs. (B) Split violin plot displaying transcription factor expression levels.

Identification of immune microenvironment and biological function characteristics in different subtypes

Differences in the expression of core regulators were evaluated to better understand the molecular characteristics between subtypes. Results indicated significant up-regulation of ACSL4, ANXA5, CD44, MGAM, MMP9, SLC16A4, and SPP1 in subtype 2, alongside notable down-regulation of HPGD (Fig. 6A,B). Additionally, substantial differences in immune cell infiltration quantities were observed between subtypes 1 and 2 (Fig. 6C).

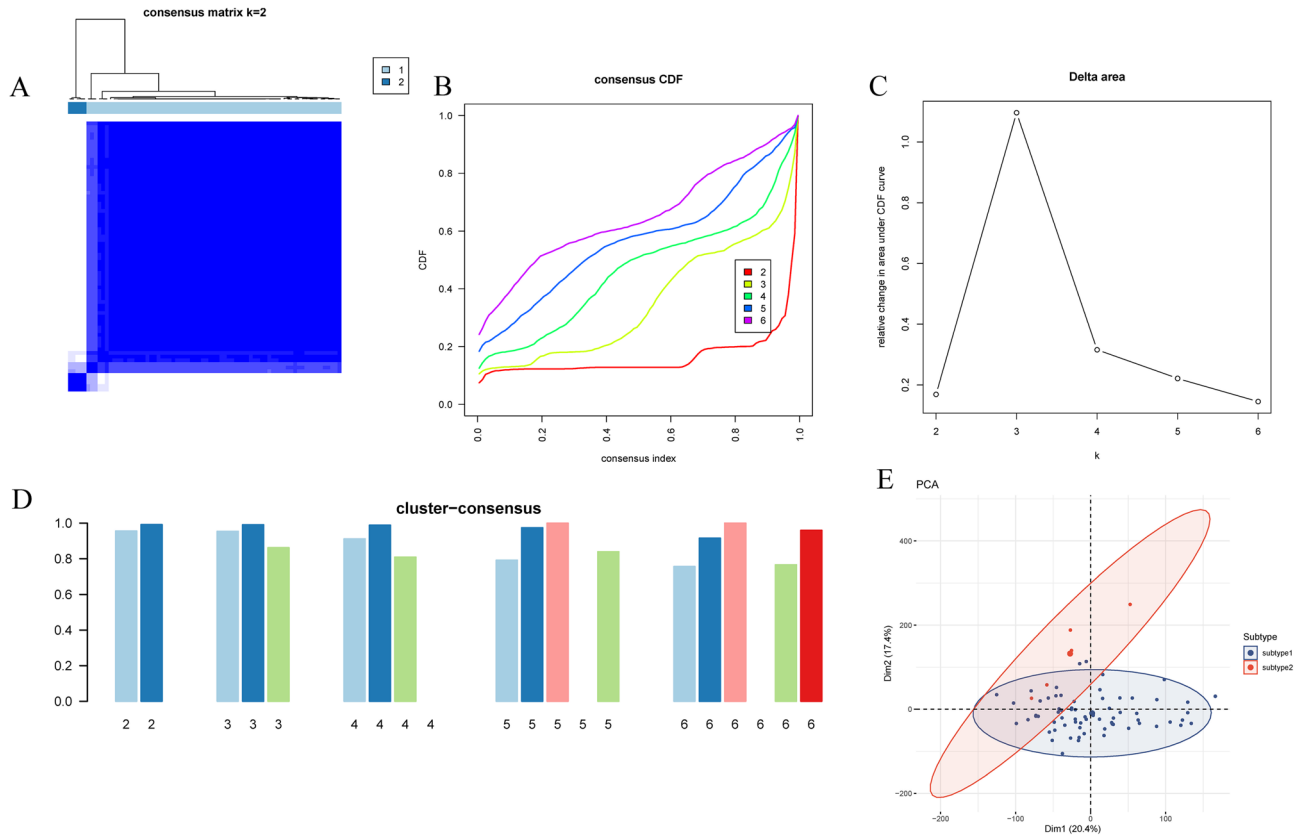


Figure 5. (A) Consensus clustering matrix for $k=2$. (B) Consensus CDF curves for $k=2$ to 6. (C) Relative alterations in CDF delta area curves. (D) Consensus score of each subtype for $k=2$ to 6. (E) PCA diagram separating subtype 1 (blue) and subtype 2 samples (red). CDF, cumulative distribution function. PCA, principal component analysis.

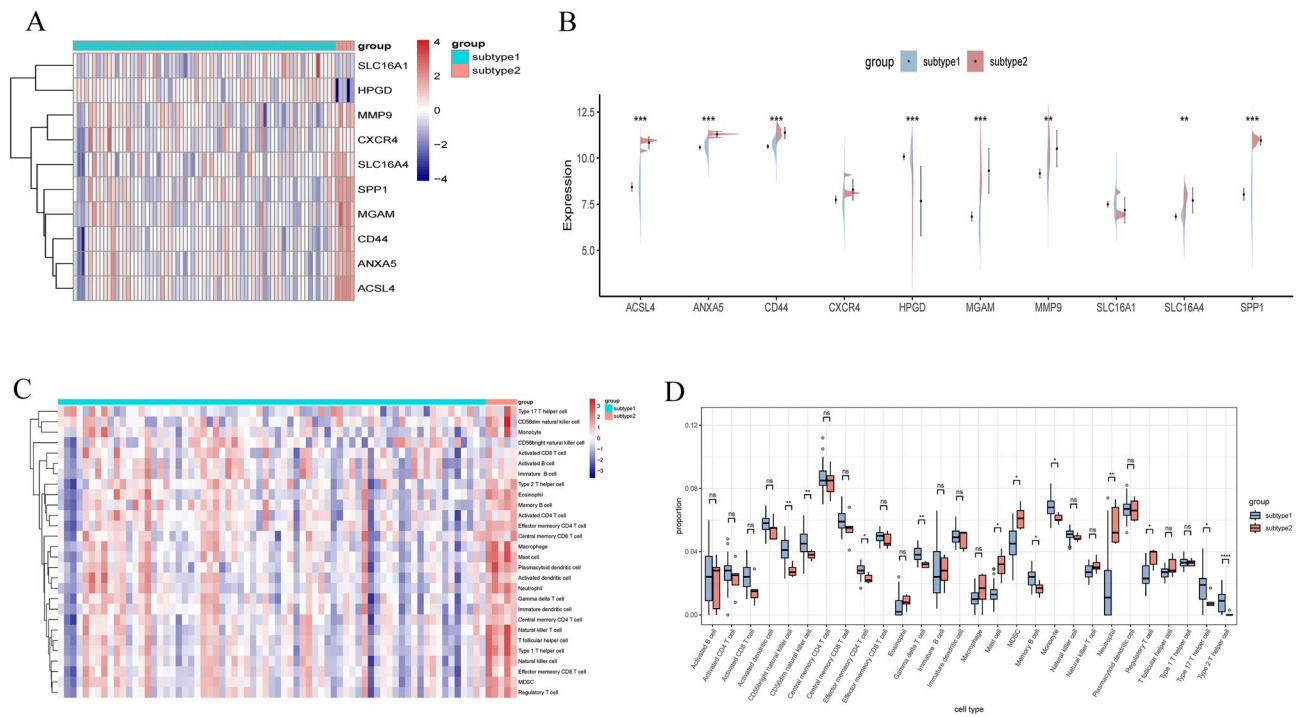


Figure 6. (A, B) Heatmap (A) and split violin plot (B) revealing the expression of 10 sBM-DEGs between subtypes. (C) Heatmap presenting diverse infiltrated immune cell abundances between subtypes. (D) Boxplot demonstrating immune cell differences between subtypes.

Specifically, subtype 1 exhibited increased amounts of CD56bright, CD56dim, effector memory CD4 T cells, $\gamma\delta$ T cells, memory B cells, monocytes, type 17 T helper cells, and type 2 T helper cells. In contrast, subtype 2 showed a higher abundance of activated mast cells, MDSCs, and neutrophils (Fig. 6D).

Significant differences in gene expression patterns between the two subtypes were observed (Fig. 7A,B). To assess variations in functional pathways linked to the distinct UC subtypes, GSEA was performed. The GSEA results revealed notable upregulation of processes such as very long chain fatty acid catabolic processes, pyruvate metabolic processes, late endosome to vacuole transport, peroxisome organization, and regulation of mitochondrial outer membrane permeabilization and inner membrane permeability in subtype 1 (Fig. 7C). Additionally, KEGG analysis of the GSEA results demonstrated upregulation of pathways including glycolysis gluconeogenesis, peroxisome, butanoate metabolism, ascorbate and aldarate metabolism, and oxidative phosphorylation in subtype 1 (Fig. 7D).

Identification of hub diagnostic gene by machine-learning algorithms

The selection of diagnostic genes was refined using three machine-learning techniques: LightGBM, SVM-RFE, and RF. By analyzing variable importance, characteristic genes were finalized based on the consensus of the LightGBM, SVM-RFE, and RF algorithms. ANXA5, CD44, and SLC16A1 were identified as the top three hub genes (Fig. 8).

Validation of hub diagnostic genes

The expression levels of ANXA5, CD44, and SLC16A1 were validated across multiple datasets: GSE75214, MASH MergeCo, and two external independent datasets (GSE87466 and GSE213621). In both UC and MASH patients, ANXA5 and CD44 were similarly upregulated, while SLC16A1 was downregulated (Figs. 9C,D and Fig. 10C,D).

ROC analysis highlighted the robust diagnostic potential of ANXA5, CD44, and SLC16A1, with AUC values of 0.994, 0.991, and 0.978 in GSE75214, and 0.773, 0.718, and 0.809 in MASH MergeCo, respectively (Fig. 9A,B). The diagnostic accuracy of these markers was further confirmed in the external independent datasets, demonstrating significant efficacy in GSE87466 (AUC: 0.932, 0.972, and 0.998) and GSE213621 (AUC: 0.711, 0.637, and 0.710) (Fig. 10A,B). These findings underscore the compelling diagnostic utility of these three feature biomarkers.

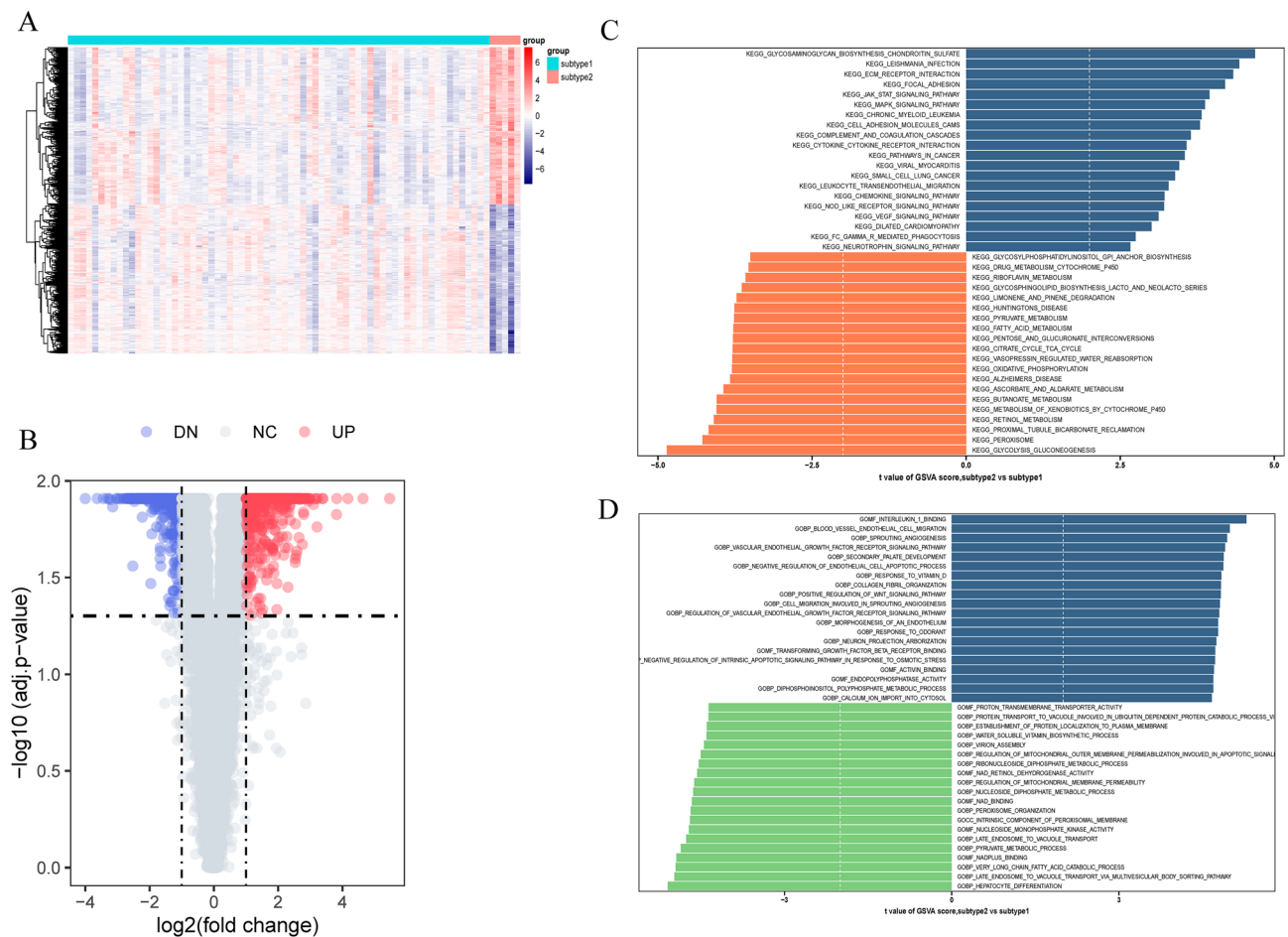


Figure 7. (A, B) Heatmap (A) and volcano plot (B) revealing the gene expression patterns between the two subtypes. (C, D) Differences in enriched biological functions (C) and hallmark pathways (D) between distinct immune microenvironment subtypes ranked by t values of GSEA scores. GSEA, gene-set variation analysis.

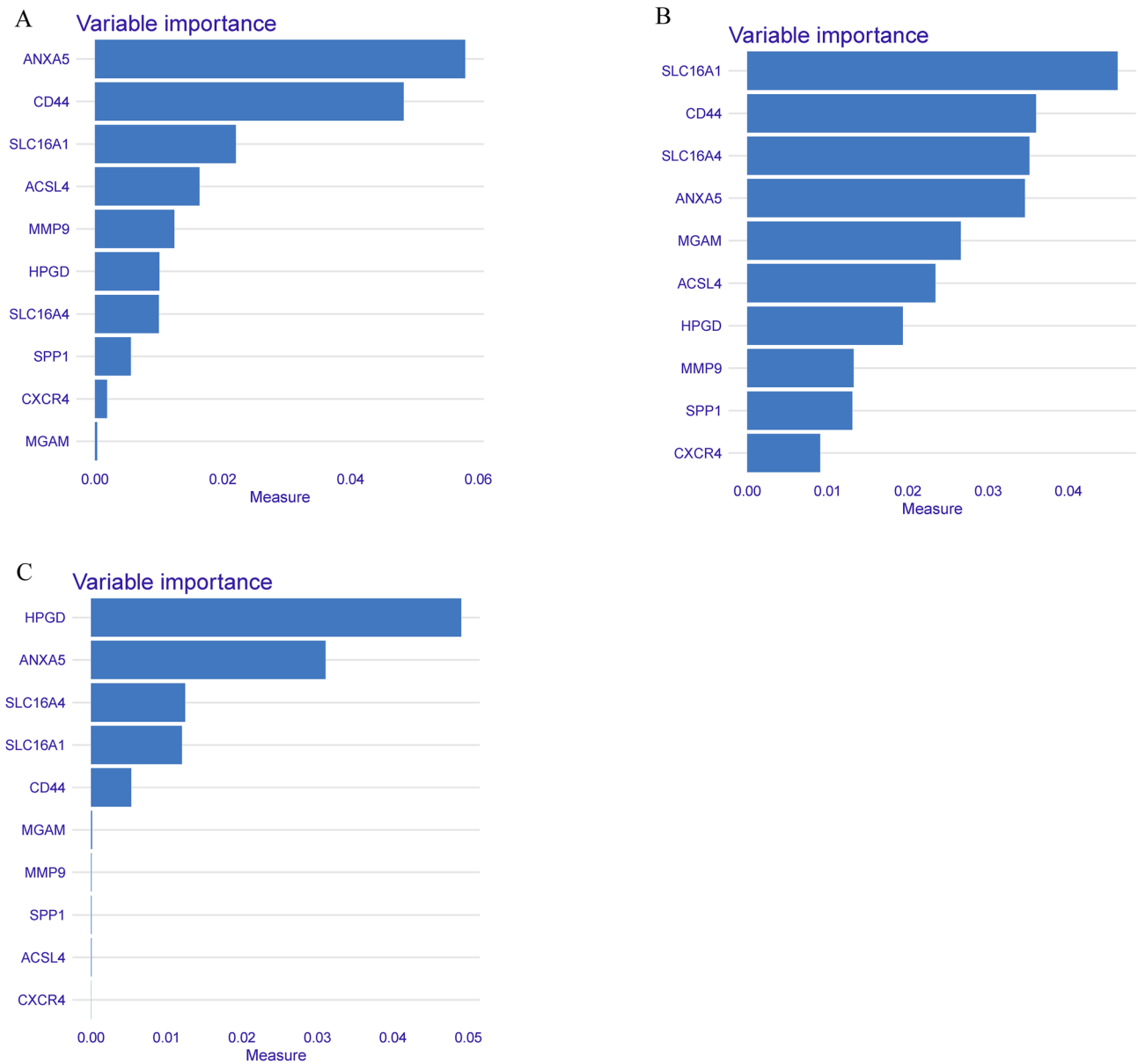


Figure 8. (A–C) Variable importance plot in machine learning: Random Forest (A), SVM-RFE (B), and LightGBM (C).

Substantial up-regulation of ANXA5 and CD44 was observed across the F0F1, F2, and F3F4 groups of MASH patients compared to HC (Fig. 10E). Additionally, CD44 exhibited significant expression differences solely between the F0F1 and F3F4 groups. SLC16A1 expression was significantly lower in the F0F1, F2, and F3F4 groups than in the control group, with significant differences between the F0F1 and F2 groups.

Validation of these genes in human UC and MASH samples confirmed these findings. As shown in Fig. 11, ANXA5 was significantly upregulated and SLC16A1 was downregulated in liver tissue of MASH patients and colonic tissue of UC patients. However, CD44 expression showed statistically significant difference only between UC patients and HC.

Immune cell infiltration analysis

The interaction between the three hub genes and the immunological microenvironment was thoroughly investigated. The distribution of 28 distinct immune cell types was assessed in UC (Fig. 12A) and MASH patients (Fig. 13A). UC patients showed significantly elevated proportions of various immune cell populations, including CD56bright natural killer cells, CD56dim natural killer cells, effector memory CD4 T cells, eosinophils, macrophages, memory B cells, natural killer T cells, T follicular helper cells, Type 1 T helper cells, Type 17 T helper cells, and Type 2 T helper cells, compared to HC (Fig. 12B). In MASH patients, elevated proportions of Type 17 T helper cells and Type 2 T helper cells were observed, while myeloid-derived suppressor cells (MDSCs) were reduced (Fig. 13B).

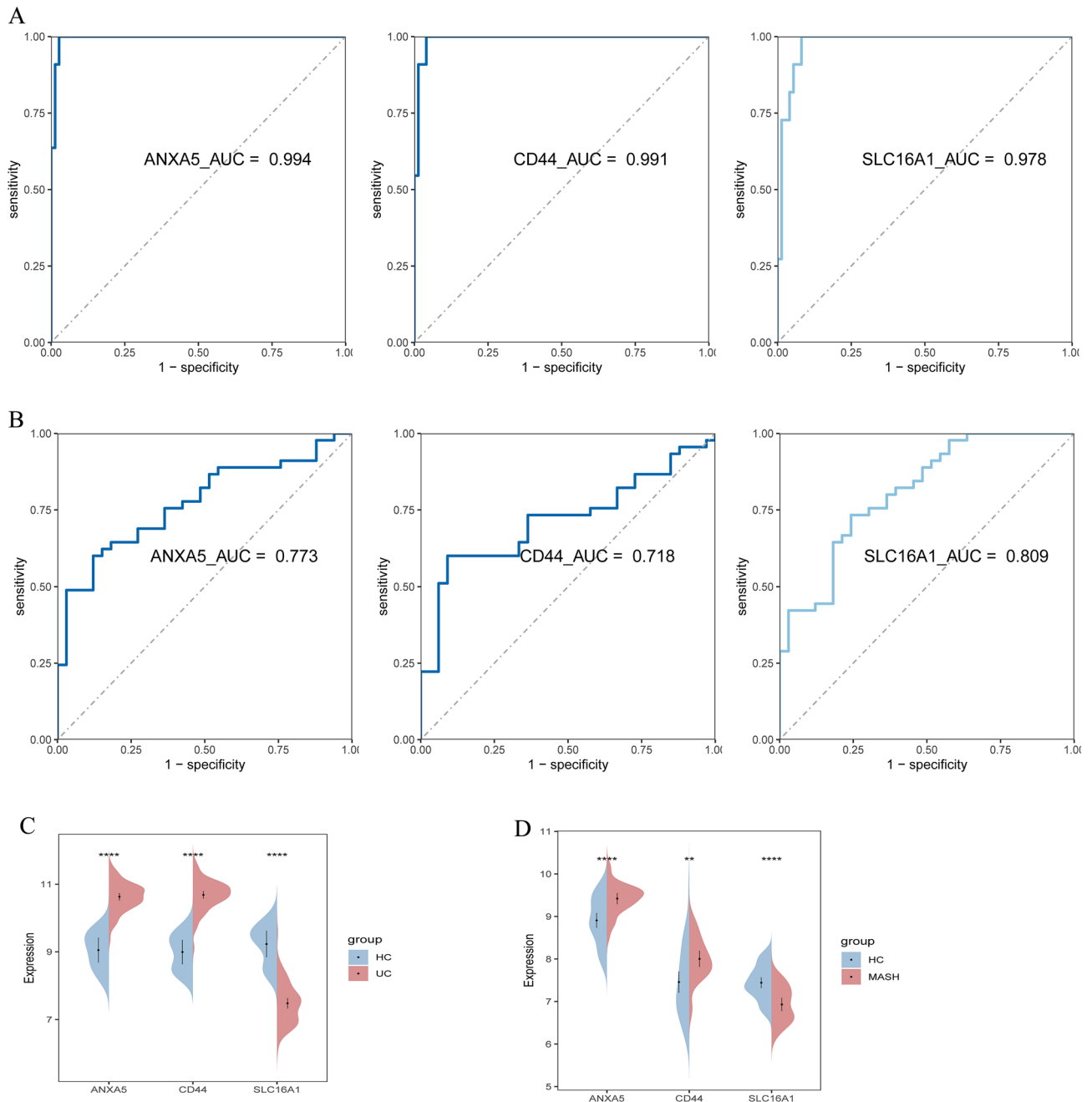


Figure 9. (A, B) ROC analysis revealing the diagnostic efficacy of 3 sBM-DEGs in the GSE75214 (A) and MASH MergeCo (B). (C, D) Expression levels of 3 sBM-DEGs in the GSE75214 (C) and MASH MergeCo (D).

Correlations between the hub genes and the content of immune cells were identified in UC (Fig. 12C) and MASH patients (Fig. 13C). In the context of UC, ANXA5 exhibited a negative correlation with both CD56bright and CD56dim natural killer cells (Fig. 12D). A positive association was observed between CD44 and MDSCs, along with a negative correlation between CD44 and CD56bright natural killer cells (Fig. 12E). SLC16A1 showed a positive association with central memory CD8 T cells and a negative relationship with MDSCs in UC (Fig. 12F). In MASH, ANXA5 and CD44 were positively associated with MDSCs and negatively associated with Type 17 T helper cells (Fig. 13D,E). Additionally, SLC16A1 demonstrated a positive relationship with CD56bright natural killer cells and a negative relationship with mast cells (Fig. 13F). These comprehensive findings illuminate the intricate interactions between the three sBM-DEGs and immune activity.

Discussion

Accumulating evidence highlighting the robust association between MAFLD and IBD, the specific molecules and mechanisms underlying this association remain elusive^{32,33}. Research revealed that decreasing levels of butyric acid could exacerbate damage to the intestinal barrier, affect the liver through leakage through the damaged intestinal barrier, leading to MAFLD³¹. And butyrate produced by *F. prausnitzii* downregulated the

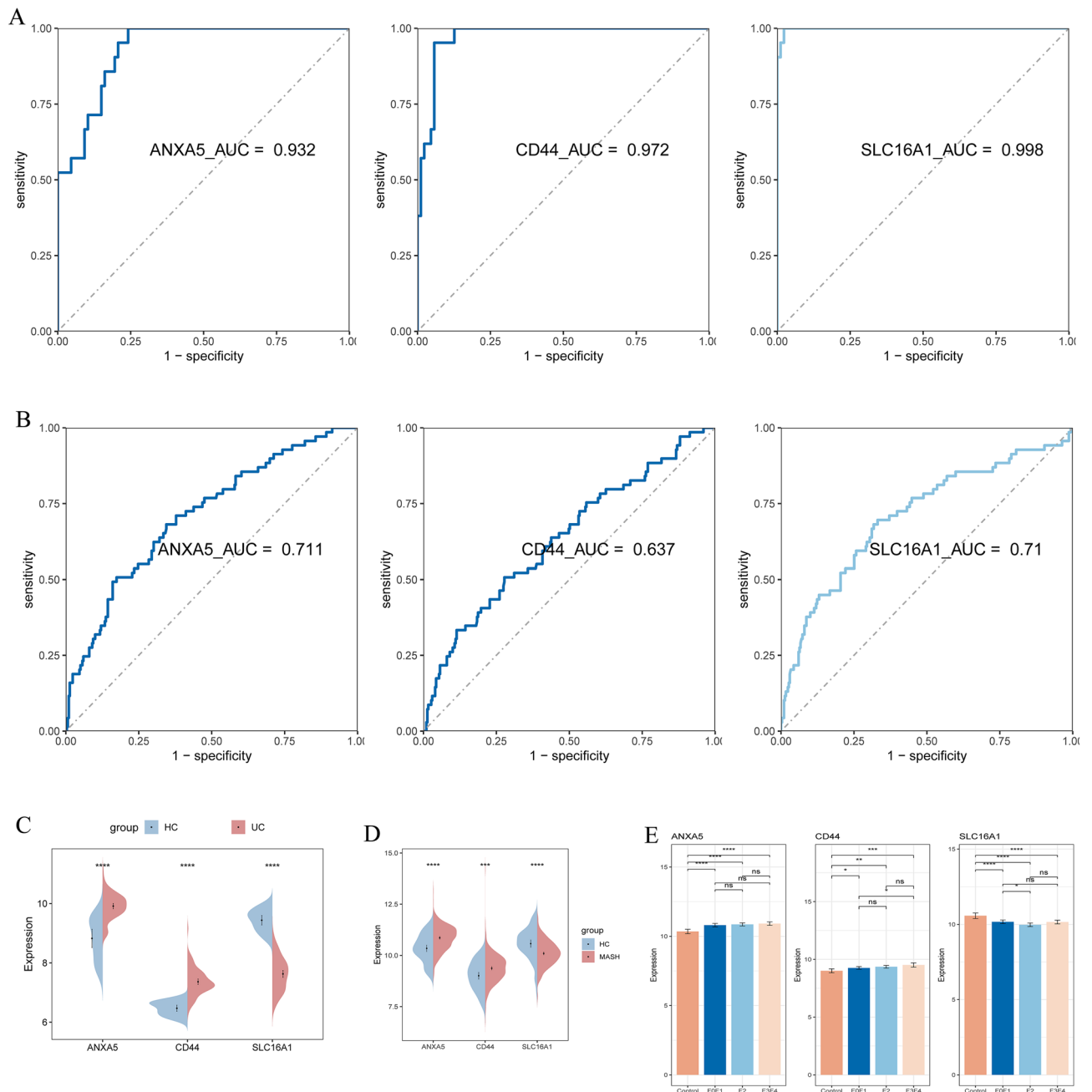


Figure 10. (A, B) ROC analysis showing the diagnostic efficacy of 3 SBM-DEGs in the GSE87466 (A) and GSE213621 (B). (C, D) Expression levels of 3 SBM-DEGs in the GSE87466 (C) and GSE213621 (D). (E) Histogram showing the expression levels of the 3 SBM-DEGs in MASH patients with different degrees of fibrosis.

pro-inflammatory IL-6/STAT3/IL-17 signal pathway via HDAC1 inhibition and transcriptional repression, while promoting expression of Foxp3 and therefore the maintenance of Treg cells³⁴. Indeed, butyrate was found to induce T cell-independent IgA secretion in the colon via activation of GPR41 and GPR109A, and inhibition of histone deacetylase to restore epithelial barrier function under inflammatory conditions. In addition, butyrate had recently emerged as activators of intracellular receptors that control immune responses, such as peroxisome proliferator-activated receptor- γ (PPAR γ)³⁵. Functional enrichment analysis showed that shared DEGs of MASH and UC patients connected to inflammatory pathways, such as IL-17 signaling pathway, intestinal immune network for IgA production, ECM-receptor interaction, and PPAR signaling pathway. In light of these findings, the mechanism of butyrate metabolism in the co-pathogenesis of MASH and UC was investigated, emphasizing the crucial clinical significance of early recognition and intervention. ANXA5, CD44, and SLC16A1 were identified as shared butyrate metabolism-related DEGs potentially implicated in the co-pathogenesis of MASH and UC.

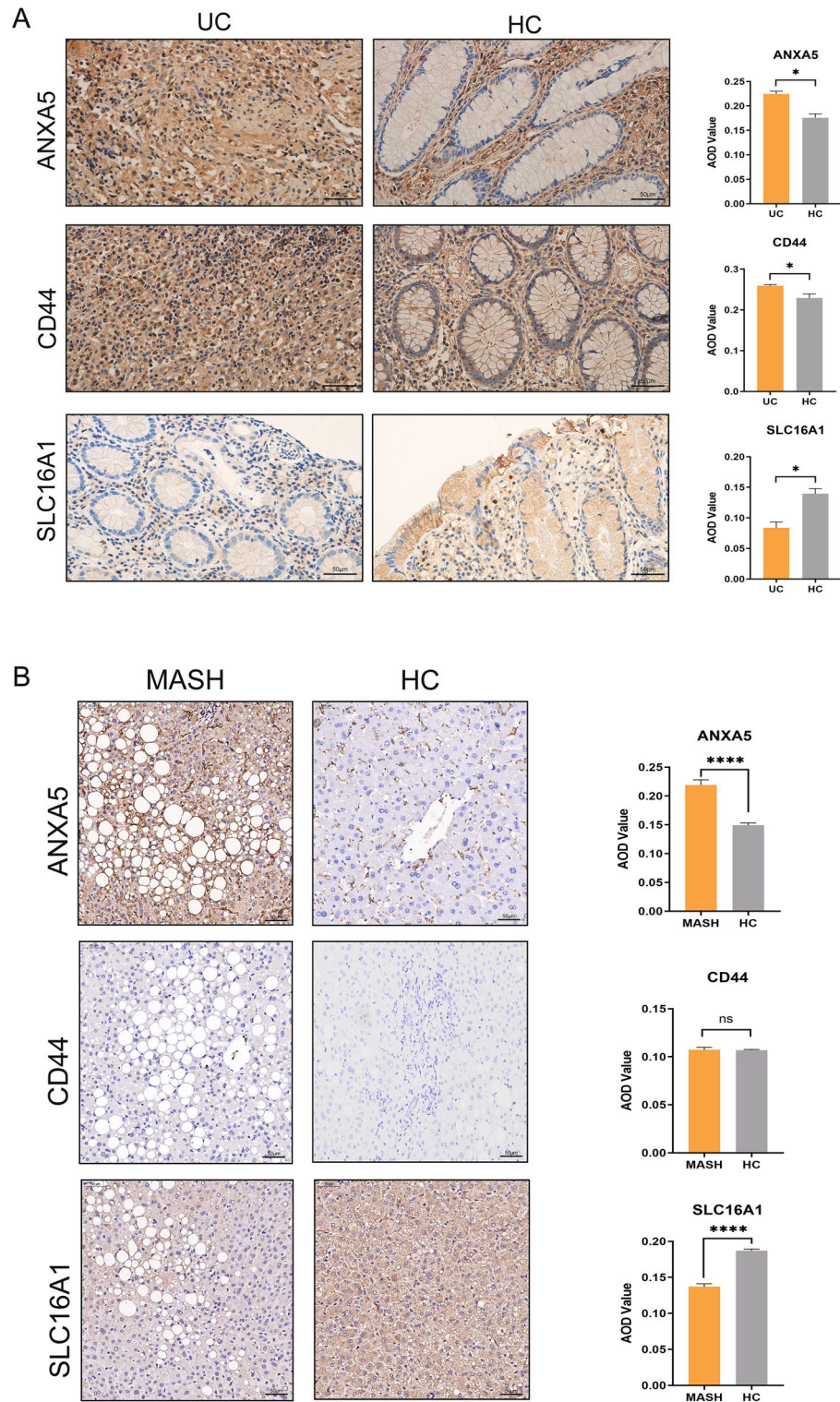


Figure 11. (A) Immunohistochemical detection and quantitative analysis of ANXA5, SLC16A1, and CD44 expression in UC patients and HC. Original magnification: 400 ×, scale bar = 50 μm. (B) Immunohistochemical detection and quantitative analysis of ANXA5, SLC16A1, and CD44 expression in MASH patients and HC. Original magnification: 200 ×, scale bar = 50 μm.

ANXA5 is a calcium-dependent phospholipid-binding protein that plays diverse roles in cellular processes. By attaching to phospholipids on the surface of apoptotic cells, ANXA5 limits their procoagulant activity and

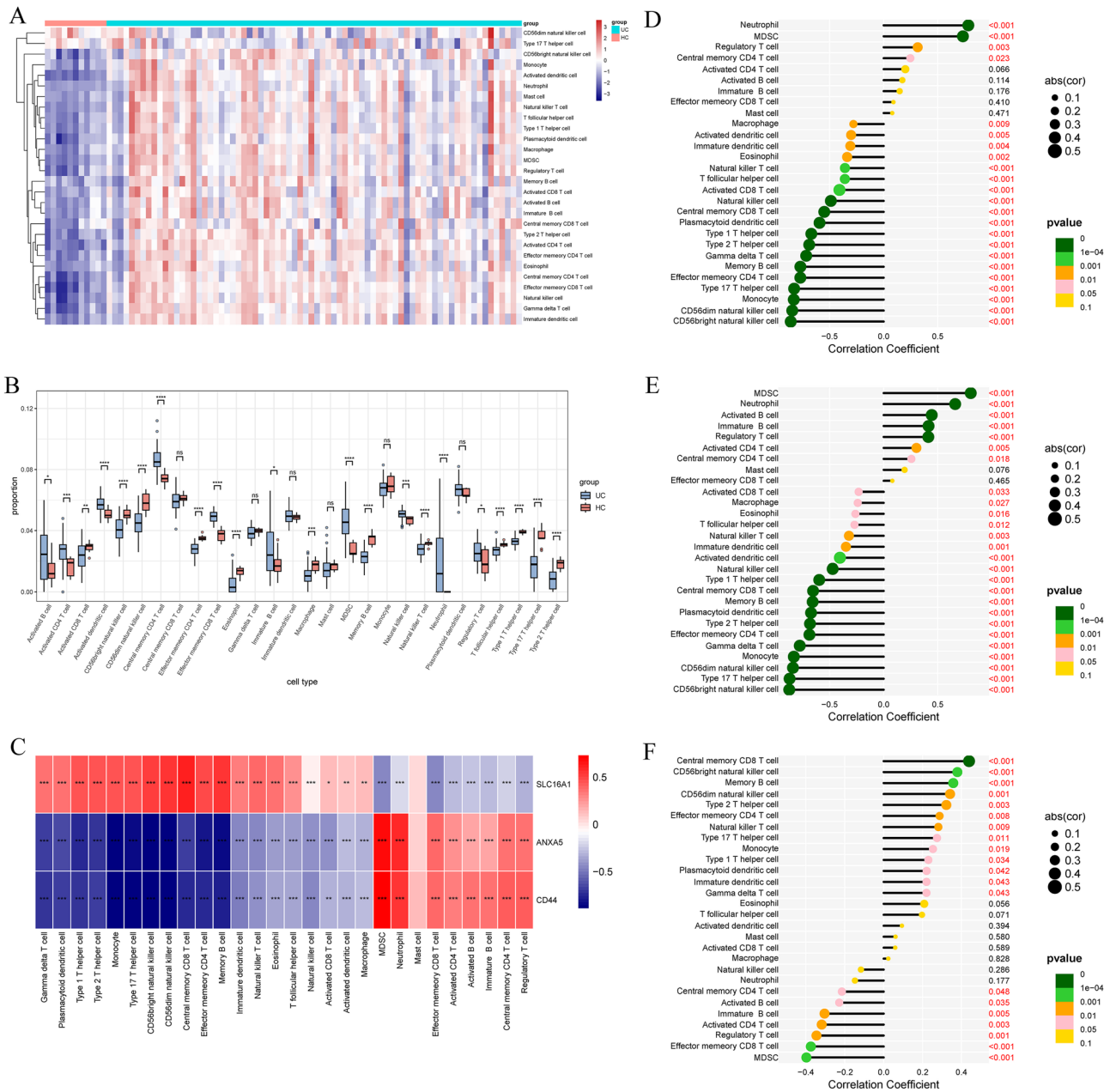


Figure 12. Immune characteristics between UC patients and HC. **(A)** Heatmap depicting the distribution of 28 immune cell infiltrations between UC patients and HC. **(B)** Boxplot revealing differences in infiltrated immune cells between UC patients and HC. **(C)** Heatmap delineating correlations in immune cell infiltrations. **(D–F)** Correlation analysis of ANXA5 **(D)**, CD44 **(E)**, and SLC16A1 **(F)** with immune infiltration.

regulates blood coagulation³⁶. Additionally, ANXA5 inhibits inflammation, regulates ion channels, and participates in cell signaling pathways³⁷. Research suggests that ANXA5 may modulate the inflammatory response in the liver and influence coagulation-related processes during MASH progression³⁸. Furthermore, ANXA5’s potential impact on intestinal inflammation and mucosal damage in UC has been explored. ANXA5 may regulate inflammatory processes in the gut, contributing to the chronic inflammation characteristic of UC^{39,40}.

The CD44 gene, named after its product, the cell surface glycoprotein CD44, plays a key role in cell adhesion, migration, and signal transduction⁴¹. Beyond its normal physiological functions, CD44 is implicated in various diseases. In patients with MASH, increased CD44 expression is linked to the recruitment of inflammatory cells to the liver, exacerbating liver damage and inflammation⁴². CD44’s interaction with hyaluronic acid in the liver’s extracellular matrix contributes to fibrosis development, a hallmark of MASH progression^{43–45}. Similarly, in UC, CD44 facilitates leukocyte migration and extravasation into inflamed intestinal tissue, intensifying chronic inflammation⁴⁶. CD44’s interaction with hyaluronic acid in the gut mucosa is also crucial for epithelial cell repair and tissue regeneration, processes often disrupted in UC patients⁴⁷. Additionally, CD44 isoforms may

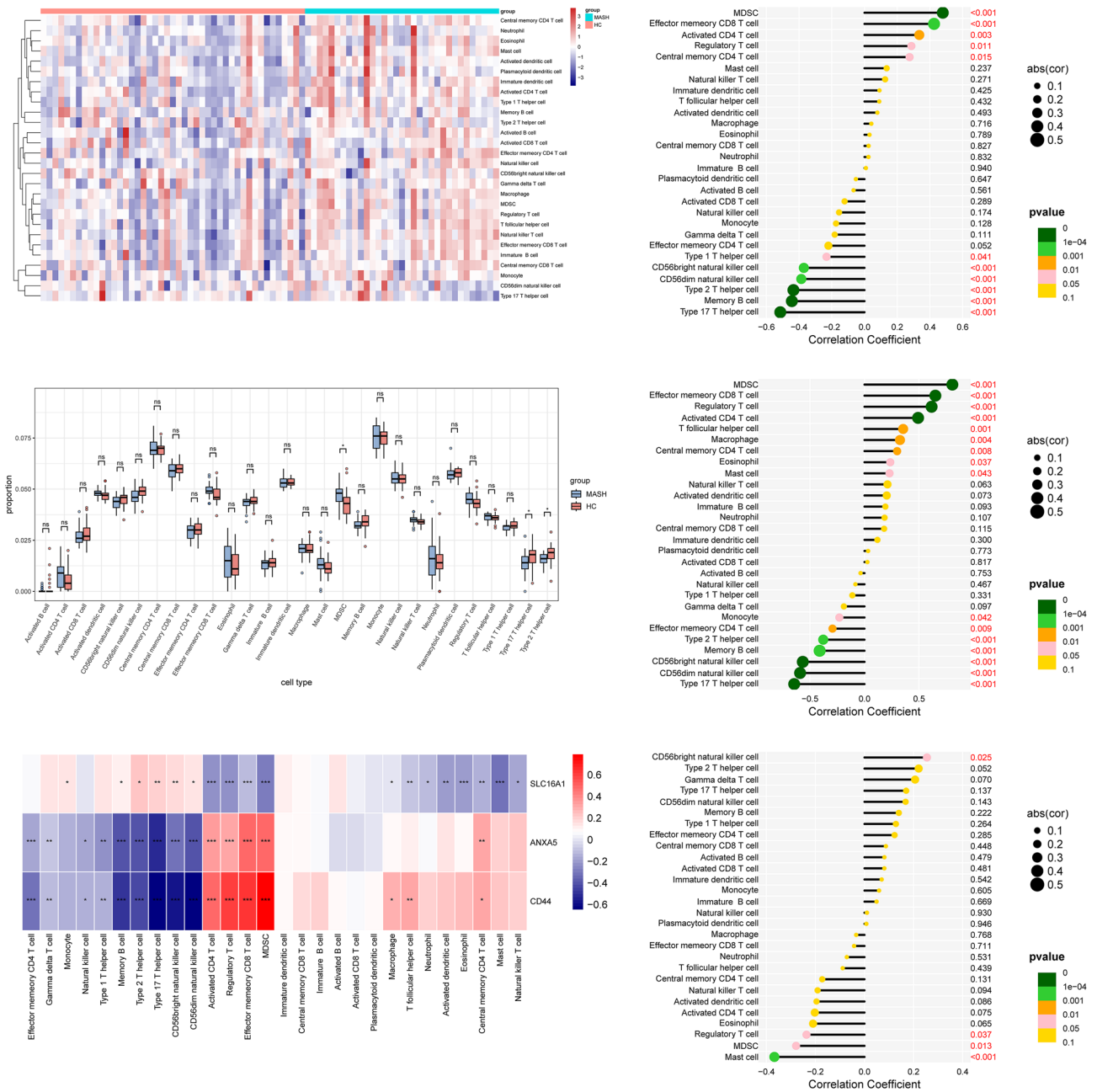


Figure 13. Immune characteristics between MASH patients and HC. **(A)** Heatmap depicting the distribution of 28 immune cell infiltrations between the MASH group and HC. **(B)** Boxplot highlighting differences in infiltrated immune cells between the MASH group and HC. **(C)** Heatmap illustrating correlations in immune cell infiltrations. **(D-F)** Correlation analysis of ANXA5 **(D)**, CD44 **(E)**, and SLC16A1 **(F)** with immune infiltration.

significantly contribute to both UC and MASH, with dysregulation of specific isoforms associated with disease severity and prognosis, potentially serving as vital biomarkers for disease activity and progression⁴⁸.

SLC16A1, also known as MCT1, is a membrane transporter protein crucial for the transmembrane transport of monocarboxylates such as lactate and pyruvate⁴⁹. It facilitates the uptake of these monocarboxylates generated during glycolysis into cells, where they can be used as an energy source or undergo further metabolism. Additionally, SLC16A1 exports lactate from highly glycolytic cells, preventing intracellular acidification and maintaining optimal pH levels. In the context of MASH, the transport of monocarboxylates such as lactate and pyruvate by SLC16A1 is vital for cellular energy production and waste removal⁵⁰. Dysregulation of SLC16A1 in the liver may lead to metabolic dysfunction and contribute to the progression of MASH⁵¹. Similarly, in UC, SLC16A1's role in regulating monocarboxylate transport may influence the metabolic and energy balance in inflamed intestinal tissue. Alterations in SLC16A1 expression or activity in UC could impact the availability of energy substrates for intestinal epithelial cells, thereby affecting tissue repair and inflammatory responses^{52–54}.

Accumulating evidence demonstrates that both innate and adaptive immune responses are crucial in inducing inflammation in UC and MASH⁹. Bioinformatics results indicate associations between UC, MASH, and various inflammatory and immune pathways. Butyrate significantly modulates adaptive and innate immune responses, regulating T cell differentiation and proliferation. Butyrate enhances Treg cell function, suppresses IL-17 levels, and reduces Th17 cells in the peripheral blood and colon tissues of TNBS-induced colitis rats compared to controls⁵⁵. It inhibits CD4⁺ and CD8⁺ T cell proliferation in a dose-dependent manner and induces apoptosis through the Fas-mediated pathway⁵⁶. Butyrate also promotes Treg cell differentiation by increasing histone H3 acetylation at the FOX3 gene locus⁵⁷. Additionally, it enhances Th1 differentiation by promoting IFN- γ levels and T-bet expression under healthy conditions while inhibiting Th1 differentiation via IL-10 production and T-bet expression during colonic inflammation⁵⁸. This study found that alterations in the abundance of infiltrating immune cells in UC are closely associated with butyrate metabolism-related genes, consistent with previous research. The absence of SLC16A1 impairs CD8⁺ T cell proliferation, shifting ATP production to mitochondrial oxidative phosphorylation. In SLC16A1^{fl/fl}T cell^{cre} mice fed a high-fat diet, a reduction in CD8⁺ T cells was observed⁵⁹.

This study successfully identified and validated three key butyrate metabolism-related genes expressed in UC and MASH patients. However, some limitations exist. First, the lack of sequencing data from UC-concomitant MASH patients in the GEO database limited the exploration of gene expression differences between UC and UC-concomitant MASH. Future studies should prioritize expanding the sample size and validating findings across multiple tissue types. Additionally, the mechanistic role of core butyrate metabolism-related genes in the relationship between UC and MASH remains unclear due to the absence of further molecular or animal experiments. Such studies are essential to deepen the understanding of this association.

Overall, this study advanced the understanding of the shared butyrate metabolism-related gene signature underpinning the interplay between UC and MASH. It unraveled potential mechanisms linking the two conditions, highlighting the roles of ANXA5, CD44, and SLC16A1 as pivotal DEGs in butyric acid metabolism between both UC and MASH.

Data availability

These data were derived from the following resources available in the public domain: Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61260>) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63067>) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE213621>) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87466>). Corresponding author can be directed for further inquiries.

Received: 20 March 2024; Accepted: 2 July 2024

Published online: 10 July 2024

References

- Kridin, K. & Ludwig, R. J. Isotretinoin and the risk of inflammatory bowel disease and irritable bowel syndrome: A large-scale global study. *J. Am. Acad. Dermatol.* **88**(4), 824–830. <https://doi.org/10.1016/j.jaad.2022.12.015> (2023).
- Paik, J. M. *et al.* Global burden of NAFLD and chronic liver disease among adolescents and young adults. *Hepatology* **75**(5), 1204–1217. <https://doi.org/10.1002/hep.32228> (2022).
- Younossi, Z. M. *et al.* Global epidemiology of nonalcoholic fatty liver disease—Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* **64**(1), 73–84. <https://doi.org/10.1002/hep.28431> (2016).
- Powell, E. E., Wong, V. W. S. & Rinella, M. Non-alcoholic fatty liver disease. *Lancet* **397**(10290), 2212–2224. [https://doi.org/10.1016/S0140-6736\(20\)32511-3](https://doi.org/10.1016/S0140-6736(20)32511-3) (2021).
- Ananthakrishnan, A. N., Kaplan, G. G. & Ng, S. C. Changing global epidemiology of inflammatory bowel diseases: Sustaining health care delivery into the 21st century. *Clin. Gastroenterol. Hepatol.* **18**(6), 1252–1260. <https://doi.org/10.1016/j.cgh.2020.01.028> (2020).
- Onwuzo, S. *et al.* Increased risk of non-alcoholic steatohepatitis in patients with inflammatory bowel disease: A population-based study. *Cureus* **15**(3), e35854 (2023).
- Bessissow, T. *et al.* Incidence and predictors of nonalcoholic fatty liver disease by serum biomarkers in patients with inflammatory bowel disease. *Inflamm. Bowel Dis.* **22**(8), 1937–1944. <https://doi.org/10.1097/MIB.0000000000000832> (2016).
- Zamani, M., Alizadeh-Tabari, S., Singh, S. & Loomba, R. Meta-analysis: Prevalence of, and risk factors for, non-alcoholic fatty liver disease in patients with inflammatory bowel disease. *Aliment Pharmacol. Ther.* **55**(8), 894–907. <https://doi.org/10.1111/apt.16879> (2022).
- Chao, C. Y. *et al.* Co-existence of non-alcoholic fatty liver disease and inflammatory bowel disease: A review article. *World J. Gastroenterol.* **22**(34), 7727–7734. <https://doi.org/10.3748/wjg.v22.i34.7727> (2016).
- Sourianarayanan, A. *et al.* Risk factors of non-alcoholic fatty liver disease in patients with inflammatory bowel disease. *J. Crohns. Colitis*. **7**(8), e279–e285. <https://doi.org/10.1016/j.crohns.2012.10.015> (2013).
- Chen, J. *et al.* Metabolic dysfunction-associated fatty liver disease and liver function markers are associated with Crohn's disease but not Ulcerative Colitis: A prospective cohort study. *Hepatol. Int.* **17**(1), 202–214. <https://doi.org/10.1007/s12072-022-10424-6> (2023).
- Martínez-Domínguez, S. J. *et al.* Liver fibrosis in non-alcoholic fatty liver disease and progression to hepatocellular carcinoma in patients with inflammatory bowel disease: A systematic review. *Cancers (Basel)* **15**(13), 3367. <https://doi.org/10.3390/cancers15133367> (2023).
- Wang, M. Y. *et al.* Premorbid steatohepatitis increases the seriousness of dextran sulfate sodium-induced ulcerative colitis in mice. *J. Clin. Transl. Hepatol.* **10**(5), 847–859. <https://doi.org/10.14218/JCTH.2021.00315> (2022).
- Abenavoli, L. *et al.* Gut microbiota in non-alcoholic fatty liver disease patients with inflammatory bowel diseases: A complex interplay. *Nutrients* <https://doi.org/10.3390/nu14245323> (2022).
- Fuchs, C. D. & Trauner, M. Role of bile acids and their receptors in gastrointestinal and hepatic pathophysiology. *Nat. Rev. Gastroenterol. Hepatol.* **19**(7), 432–450. <https://doi.org/10.1038/s41575-021-00566-7> (2022).
- Liu, W. *et al.* A bridge for short-chain fatty acids to affect inflammatory bowel disease, type 1 diabetes, and non-alcoholic fatty liver disease positively: By changing gut barrier. *Eur. J. Nutr.* **60**(5), 2317–2330. <https://doi.org/10.1007/s00394-020-02431-w> (2021).

17. Shen, Z. H. *et al.* Relationship between intestinal microbiota and ulcerative colitis: Mechanisms and clinical application of probiotics and fecal microbiota transplantation. *World J. Gastroenterol.* **24**(1), 5–14. <https://doi.org/10.3748/wjg.v24.i1.5> (2018).
18. Cui, L. *et al.* *Scutellaria baicalensis* Georgi polysaccharide ameliorates DSS-induced ulcerative colitis by improving intestinal barrier function and modulating gut microbiota. *Int. J. Biol. Macromol.* **166**, 1035–1045. <https://doi.org/10.1016/j.ijbiomac.2020.10.259> (2021).
19. Machiels, K. *et al.* A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut*. **63**(8), 1275–1283. <https://doi.org/10.1136/gutjnl-2013-304833> (2014).
20. Deng, M. *et al.* SCFAs alleviated steatosis and inflammation in mice with NASH induced by MCD. *J. Endocrinol.* **245**(3), 425–437. <https://doi.org/10.1530/JOE-20-0018> (2020).
21. Fraser, D. A. *et al.* A structurally engineered fatty acid, icosabutate, suppresses liver inflammation and fibrosis in NASH. *J. Hepatol.* **76**(4), 800–811. <https://doi.org/10.1016/j.jhep.2021.12.004> (2022).
22. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**(6), 882–883. <https://doi.org/10.1093/bioinformatics/bts034> (2012).
23. Kleiner, D. E. *et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41**(6), 1313–1321. <https://doi.org/10.1002/hep.20701> (2005).
24. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559. <https://doi.org/10.1186/1471-2105-9-559> (2008).
25. 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. <https://doi.org/10.1093/nar/gkac963> (2023).
26. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**(5), 284–287. <https://doi.org/10.1089/omi.2011.0118> (2012).
27. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**(43), 15545–15550. <https://doi.org/10.1073/pnas.0506580102> (2005).
28. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**(12), 1572–1573. <https://doi.org/10.1093/bioinformatics/btq170> (2010).
29. GSVA: Gene set variation analysis for microarray and RNA-seq data - PubMed. Accessed January 12, 2023. <https://pubmed.ncbi.nlm.nih.gov/23323831/>
30. Andoh, A. & Nishida, A. Alteration of the gut microbiome in inflammatory bowel disease. *Digestion* **104**(1), 16–23. <https://doi.org/10.1159/000525925> (2023).
31. Aron-Wisniewsky, J. *et al.* Gut microbiota and human NAFLD: Disentangling microbial signatures from metabolic disorders. *Nat. Rev. Gastroenterol. Hepatol.* **17**(5), 279–297. <https://doi.org/10.1038/s41575-020-0269-9> (2020).
32. Wang, W. *et al.* Shared biomarkers and immune cell infiltration signatures in ulcerative colitis and nonalcoholic steatohepatitis. *Sci. Rep.* **13**(1), 18497. <https://doi.org/10.1038/s41598-023-44853-6> (2023).
33. Maresca, R. *et al.* Inflammatory bowel diseases and non-alcoholic fatty liver disease: Piecing a complex puzzle together. *Int. J. Mol. Sci.* **25**(6), 3278. <https://doi.org/10.3390/ijms25063278> (2024).
34. Zhou, L. *et al.* *Faecalibacterium prausnitzii* produces butyrate to maintain Th17/Treg balance and to ameliorate colorectal colitis by inhibiting histone deacetylase 1. *Inflamm. Bowel Dis.* **24**(9), 1926–1940. <https://doi.org/10.1093/ibd/izy182> (2018).
35. Byndloss, M. X. *et al.* Microbiota-activated PPAR- γ signaling inhibits dysbiotic Enterobacteriaceae expansion. *Science* **357**(6351), 570–575. <https://doi.org/10.1126/science.aam9949> (2017).
36. Woodward, A., Faria, G. & Harrison, R. G. Annexin A5 as a targeting agent for cancer treatment. *Cancer Lett.* **547**, 215857 (2022).
37. Chen, W. *et al.* The ABCA1-efferocytosis axis: A new strategy to protect against atherosclerosis. *Clin. Chim. Acta.* **518**, 1–8. <https://doi.org/10.1016/j.cca.2021.02.025> (2021).
38. Xu, F. *et al.* Annexin A5 regulates hepatic macrophage polarization via directly targeting PKM2 and ameliorates NASH. *Redox Biol.* **36**, 101634 (2020).
39. Li, Y., Xu, M., Zhu, Z., Xu, F. & Chen, B. Transendothelial electrical resistance measurement by a microfluidic device for functional study of endothelial barriers in inflammatory bowel disease. *Front. Bioeng. Biotechnol.* **11**, 1236610 (2023).
40. Zhang, X. *et al.* Phosphatidylserine externalized on the colonic capillaries as a novel pharmacological target for IBD therapy. *Signal Transduct. Target Ther.* **6**(1), 235 (2021).
41. Weng, X., Maxwell-Warburton, S., Hasib, A., Ma, L. & Kang, L. The membrane receptor CD44: Novel insights into metabolism. *Trends Endocrinol. Metab.* **33**(5), 318–332. <https://doi.org/10.1016/j.tem.2022.02.002> (2022).
42. Patouraux, S. *et al.* CD44 is a key player in non-alcoholic steatohepatitis. *J. Hepatol.* **67**(2), 328–338. <https://doi.org/10.1016/j.jhep.2017.03.003> (2017).
43. Wiering, L. & Tacke, F. Treating inflammation to combat non-alcoholic fatty liver disease. *J. Endocrinol.* <https://doi.org/10.1530/JOE-22-0194> (2023).
44. Egan, C. E. *et al.* CCR2 and CD44 promote inflammatory cell recruitment during fatty liver formation in a lithogenic diet fed mouse model. *PLoS One* **8**(6), e65247. <https://doi.org/10.1371/journal.pone.0065247> (2013).
45. Jiang, X. *et al.* Lipid-injured hepatocytes release sOPN to improve macrophage migration via CD44 engagement and pFak-NF κ B signaling. *Cytokine* **142**, 155474 (2021).
46. Yang, H. *et al.* Human induced pluripotent stem cell-derived mesenchymal stem cells promote healing via TNF- α -stimulated gene-6 in inflammatory bowel disease models. *Cell Death Dis.* **10**(10), 718. <https://doi.org/10.1038/s41419-019-1957-7> (2019).
47. Song, X. *et al.* Efficient therapy of inflammatory bowel disease (IBD) with highly specific and durable targeted Ta(2) C modified with chondroitin sulfate (TACS). *Adv. Mater.* **35**, e2301585 (2023).
48. Franić, I. *et al.* Expression of CD44 in leukocyte subpopulations in patients with inflammatory bowel diseases. *Diagnostics (Basel)* **12**(8), 2014 (2022).
49. van Hasselt, P. M. *et al.* Monocarboxylate transporter 1 deficiency and ketone utilization. *N. Engl. J. Med.* **371**(20), 1900–1907. <https://doi.org/10.1056/NEJMoa1407778> (2014).
50. Hadjihambi, A. *et al.* Partial MCT1 inactivation protects against diet-induced non-alcoholic fatty liver disease and the associated brain dysfunction. *J. Hepatol.* **78**(1), 180–190. <https://doi.org/10.1016/j.jhep.2022.08.008> (2023).
51. Carneiro, L. *et al.* AMPK activation caused by reduced liver lactate metabolism protects against hepatic steatosis in MCT1 haploinsufficient mice. *Mol. Metab.* **6**(12), 1625–1633. <https://doi.org/10.1016/j.molmet.2017.10.005> (2017).
52. De Preter, V. *et al.* Impaired butyrate oxidation in ulcerative colitis is due to decreased butyrate uptake and a defect in the oxidation pathway. *Inflamm. Bowel Dis.* **18**(6), 1127–1136. <https://doi.org/10.1002/ibd.21894> (2012).
53. Stilling, R. M. *et al.* The neuropharmacology of butyrate: The bread and butter of the microbiota-gut-brain axis?. *Neurochem. Int.* **99**, 110–132. <https://doi.org/10.1016/j.neuint.2016.06.011> (2016).
54. Erdmann, P. *et al.* Dysregulation of mucosal membrane transporters and drug-metabolizing enzymes in ulcerative colitis. *J. Pharm. Sci.* **108**(2), 1035–1046. <https://doi.org/10.1016/j.xphs.2018.09.024> (2019).
55. Zhang, M. *et al.* Butyrate inhibits interleukin-17 and generates Tregs to ameliorate colorectal colitis in rats. *BMC Gastroenterol.* **16**(1), 84. <https://doi.org/10.1186/s12876-016-0500-x> (2016).
56. Zimmerman, M. A. *et al.* Butyrate suppresses colonic inflammation through HDAC1-dependent Fas upregulation and Fas-mediated apoptosis of T cells. *Am. J. Physiol. Gastrointest. Liver Physiol.* **302**(12), G1405–1415. <https://doi.org/10.1152/ajpgi.00543.2011> (2012).

57. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**(7480), 446–450. <https://doi.org/10.1038/nature12721> (2013).
58. Chen, L. *et al.* Microbiota metabolite butyrate differentially regulates Th1 and Th17 cells' differentiation and function in induction of colitis. *Inflamm. Bowel Dis.* **25**(9), 1450–1461. <https://doi.org/10.1093/ibd/izz046> (2019).
59. Macchi, C. *et al.* Monocarboxylate transporter 1 deficiency impacts CD8+ T lymphocytes proliferation and recruitment to adipose tissue during obesity. *iScience* **25**(6), 104435. <https://doi.org/10.1016/j.isci.2022.104435> (2022).

Acknowledgements

The authors thank the National Cancer Institute for providing the Gene Expression Omnibus (GEO) data set. In addition, the authors would like to thank the editors and the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

Author contributions

(I) Conception and design: B.D., Y.L., Y.C.; (II) Administrative support: Y.C., P.H., W.D.; (III) Provision of study materials or patients: P.H., Z.T., J.Z.; (IV) Collection and assembly of data: B.D., Y.L., J.M.; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All author.

Funding

The National Natural Science Foundation of China (No. 82370542) and Natural Science Foundation of Hubei Province (2022CFB703) funded this manuscript.

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/s41598-024-66574-0>.

Correspondence and requests for materials should be addressed to W.D.

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