ORIGINAL RESEARCH

Based on Weighted Gene Co-Expression Network Analysis Reveals the Hub Immune Infiltration-Related Genes Associated with Ulcerative Colitis

Zongbiao Tan [iD]*, Chuan Liu*, Pengzhan He*, Yanrui Wu, Jiao Li, Jixiang Zhang, Weiguo Dong [iD]

Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan, 430060, People's Republic of China

*These authors contributed equally to this work

Correspondence: Weiguo Dong, Department of Gastroenterology, Renmin Hospital of Wuhan University, No. 99 Zhangzhidong Road, Wuhan, Hubei Province, 430060, People's Republic of China, Tel +86 027-88041911, Email dongweiguo@whu.edu.cn

Purpose: Immune infiltration plays a pivotal role in the pathogenesis of mucosal damage in ulcerative colitis (UC). The objective of this study was to systematically analyze and identify genetic characteristics associated with immune infiltration in UC.

Patients and Methods: Gene expression data from three independent datasets obtained from the Gene Expression Omnibus (GEO) were utilized. By employing the ssGSEA and CIBERSORT algorithms, we estimated the extent of immune cell infiltration in UC samples. Subsequently, Weighted Correlation Network Analysis (WGCNA) was performed to identify gene modules exhibiting significant associations with immune infiltration, and further identification of hub genes associated with immune infiltration was accomplished using least absolute shrinkage and selection operator (LASSO) regression analysis. The relationship between the identified hub genes and clinical information was subsequently investigated.

Results: Our findings revealed significant activation of both innate and adaptive immune cells in UC. Notably, the expression levels of CD44, IL1B, LYN, and ITGA5 displayed strong correlations with immune cell infiltration within the mucosa of UC patients. Immunohistochemical analysis confirmed the significant upregulation of CD44, LYN, and ITGA5 in UC samples, and their expression levels were found to be significantly associated with common inflammatory markers, including the systemic immune inflammation indices, C-reactive protein, and erythrocyte sedimentation rate.

Conclusion: CD44, LYN, and ITGA5 are involved in the immune infiltration pathogenesis of UC and may be potential therapeutic targets for UC.

Keywords: Ulcerative colitis, Immune infiltration, WGCNA, CIBERSORT, LASSO

Introduction

Ulcerative colitis (UC) is one of the prominent forms of inflammatory bowel disease (IBD), which primarily affects the colorectal region, characterized by chronic abdominal pain, diarrhea, mucopurulent stools, and varying degrees of systemic manifestations.[1] Epidemiological studies indicate that UC has a higher prevalence in developed countries, such as Europe and the United States.[2] However, there has been a consistent increase in the incidence of UC in developing countries as a consequence of ongoing industrialization.[3] The recurrent and persistent nature of UC not only significantly impairs patients' quality of life but also imposes a substantial economic burden on society.[4]

The etiology of UC remains incompletely understood, with current understanding suggesting a multifactorial interplay involving environmental factors,[5] genetic predisposition,[6] infectious agents, and immune dysregulation.[7] Among the various contributing factors, immune dysregulation plays a key role in the pathogenesis of UC, and the pathogenesis of UC involves complex interactions between the adaptive immune system and the innate immune system. CD4[+] T cells

play a central role in regulating adaptive immunity in UC. Compared with Crohn's disease (CD), UC is more characterized by pronounced infiltration of the colonic mucosa with Type 2 T helper (Th2) cell-associated cytokines, such as IL-5 and IL-13, suggesting the involvement of atypical type 2 immune responses.[8,9] At the same time, other immune cell types, including Type17 T helper (Th17) cells, T regulatory cells, and B lymphocytes, further increase the complexity of the immune response. Furthermore, the aberrant activation of the adaptive immune system triggers an exaggerated response from the innate immune system. This results in substantial infiltration of neutrophils and macrophages into the intestinal mucosa, leading to the release of proteinaceous particles and reactive oxygen species. These elements inflict cellular and tissue damage, ultimately leading to characteristic mucosal manifestations in UC, such as cryptitis, mucosal erosion, and ulcer formation.[7,9,10] Therefore, the central therapeutic strategy for managing UC primarily focuses on attenuating the inflammatory mechanisms. At present, therapeutic interventions for UC encompass non-specific modalities such as 5-aminosalicylic acid, glucocorticoids, and immunomodulators, alongside targeted therapies including anti-tumor necrosis factor alpha (anti-TNF-α) and anti-IL-12/IL-23 agents, among other options.[11,12] Among these, anti-TNF-α is currently one of the most important targets in the clinical treatment of moderate-to-severe UC. Despite the increasing variety of biologics approved for UC, which can help UC patients induce and maintain long-term remission, there is still a high proportion of patients with primary or secondary nonresponse.[13] Consequently, there is an urgent need to enhance our understanding of the immune mechanisms underlying UC pathogenesis and search for new therapeutic targets. For single-cell analysis, traditional flow cytometry is expensive and time-consuming. In recent years, due to the popularity of high-throughput sequencing technologies, experts have developed several deconvolution methods to estimate the relative abundance of each cell type in a sample by using gene expression profile data,[14] such as cell-type identification by estimating relative subsets of RNA transcripts (CIBERSORT)[15] and Single sample Gene Set Enrichment Analysis (ssGSEA).[16] Meanwhile, to gain a deeper insight into the interactions and coordinated regulatory relationships between genes, a variety of comprehensive analysis techniques based on biological networks have been developed, such as weighted gene Co-expression network analysis (WGCNA). WGCNA relies on high-throughput gene expression data and fuses genes into co-expression modules by calculating co-expression relationships between genes, revealing the similarity and correlation of gene functions.[17] By identifying and analyzing these co-expression modules, WGCNA can identify collections of genes that play key roles in specific biological processes or diseases. Compared with other methods, WGCNA not only calculates the correlation coefficient between genes but also considers the weight of gene expression data, which can better deal with the uncertainty in gene expression data and avoid the loss of key information. Therefore, WGCNA is widely used in studies aimed at identifying potential biomarkers or therapeutic targets.[18,19]

Given the critical role of immunological factors in UC, the evaluation of hub genes involved in immune infiltration of UC will help us understand the pathogenesis of UC and provide new immunotherapeutic targets for UC. In this study, we explored the infiltration of immune cells in UC mucosal tissue. Based on the information on immune infiltration, the most relevant gene modules for immune cell infiltration were searched by WGCNA, and hub genes were identified using the least absolute shrinkage and selection operator (LASSO) regression analysis. It provides a reference for understanding the molecular mechanisms associated with immune infiltration of UC and the novel selection of therapeutic targets.

Materials and Methods
Data Collection
Using the search terms "inflammatory bowel disease and human" or "ulcerative colitis and human" as well as "inflammatory bowel disease and treatment" or "ulcerative colitis and anti-TNF-α treatment", we accessed and obtained the GSE92415,[20] GSE206285,[21] and GSE16879[22] datasets as series matrix files from the publicly available NCBI Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo). The GSE92415 dataset, based on the GPL13158 platform, comprised expression profiling data from 183 mucosal biopsy samples, including 162 samples from patients with UC and 21 control individuals.[20] This dataset also provided information on the patients' response to golimumab treatment and their disease activity scores (evaluated using the "Mayo score"). The GSE206285 dataset, annotated by the GPL13158 platform, consisted of expression profiling data from 568 mucosal biopsy samples.[21] Furthermore, the GSE16879 dataset, annotated by the GPL570

platform, included 48 samples from patients with UC and 6 control samples, along with information regarding patient response to infliximab treatment.[22] The GSE206285 and GSE16879 datasets were employed to validate the stability of the identified genes. A graphical representation of the study workflow is presented in Figure 1.

Analysis of Differentially Expressed Genes (DEGs) and Immune Infiltration

The GSE92415 dataset was subjected to standardization, gene annotation, and expression value calculations using the "limma" package in R. DEGs were identified based on predefined screening criteria: an absolute fold change ($|\log_2FC|$) ≥ 1 (equivalent to a fold change ≥ 2) and adjusted P-value < 0.05. Volcano plots illustrating the DEGs were generated using the "ggplot2" package. Two methods were used to evaluate the immune infiltration of the sample, including CIBERSORT and ssGSEA. The ssGSEA analysis is mainly performed using the GSVA software package and was performed with parameters as follows: method= ssgsea, kcdf = 'Gaussian', mx.diff=TRUE. Heatmaps and histograms depicting the immune cell infiltration patterns were generated using the "ggplot2" package for both CIBERSORT and
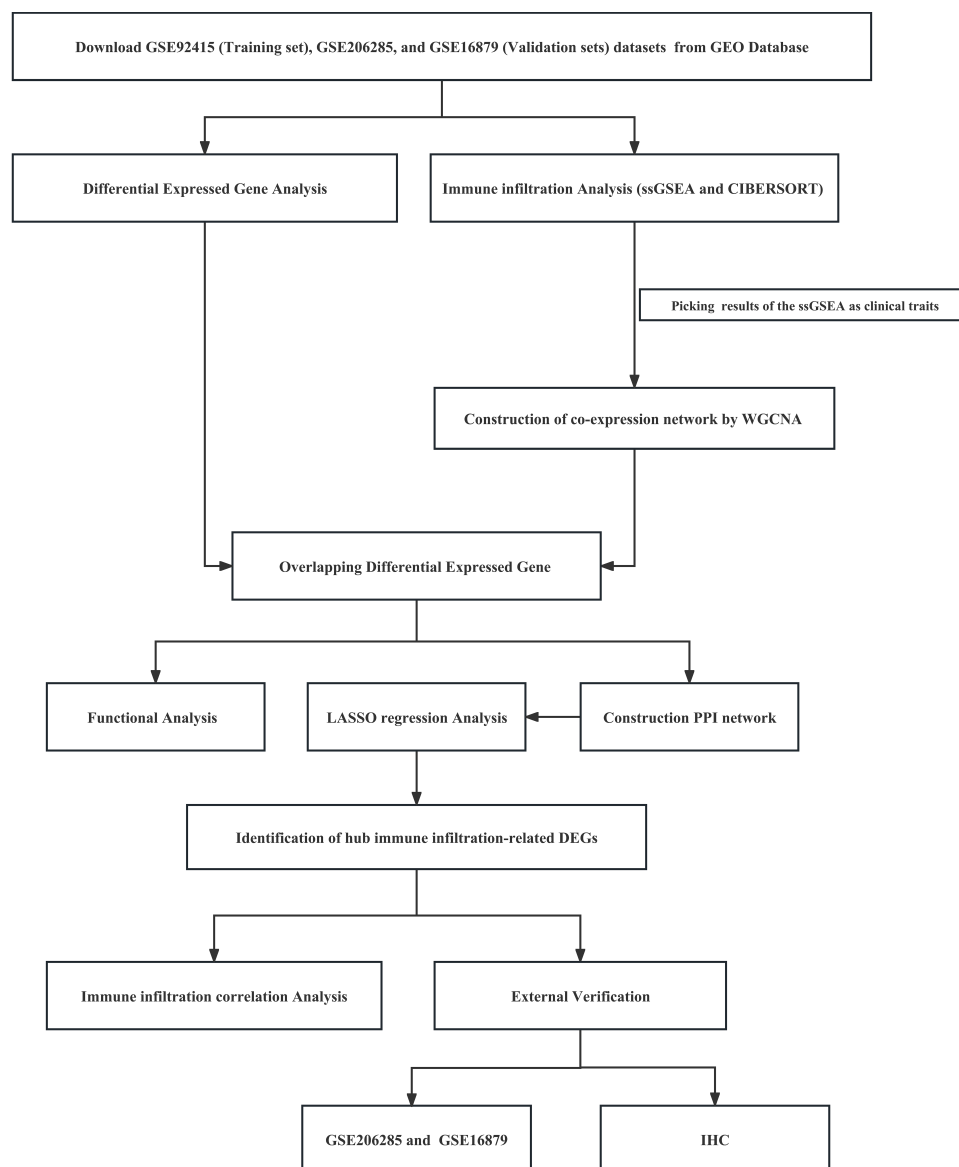

Figure 1 The workflow of this study.
Abbreviations: GEO, Gene Expression Omnibus; LASSO, Least absolute shrinkage and selection operator; PPI, Protein-protein interaction; DEGs, Differentially Expressed Genes; IHC: Immunohistochemical.

ssGSEA results. To compare the difference of immune cell infiltration between the control group and the UC group, the Wilcoxon rank-sum test was employed, and a *P*-value < 0.05 was considered statistically significant. The results were visualized using the "vioplot" package.

Construction of Co-Expression Networks

To explore the relationship between gene networks and immune cells, we employed WGCNA. Initially, the top 25% (5185) of genes with the highest median absolute deviation were selected as the gene set for the WGCNA clustering. The results of ssGSEA (proportions of various immune cells) were employed as clinical trait data for WGCNA analysis. Outliers were identified and removed through sample clustering. Subsequently, the adjacency matrix was transformed into a topological overlap matrix following the selection of an optimal soft threshold to construct the network. The gene dendrogram and module color were established based on gene similarity. The initial modules were further divided using dynamic tree cutting, and similar modules were merged. Finally, gene modules exhibiting the strongest associations with immune cells were identified as key gene modules for subsequent analysis. This was achieved by calculating Pearson correlation coefficients between module characteristics and sample traits. By setting the gene significance filter at 0.5 and the module significance filter at 0.8, the central genes within these modules were determined as key genes.

Functional Analysis of Immune Infiltration-Related DEGs

To identify DEGs specifically associated with immune infiltration, we employed the "Venn" package in R and intersected the DEGs with the central genes from the brown module. Subsequently, gene ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the immune infiltration-related DEGs were conducted using the DAVID online platform (https://david.ncifcrf.gov). DAVID (Database for Annotation, Visualization, and Integrated Discovery) is a widely used online analysis tool that integrates multiple biological data and analysis resources, aiming to provide comprehensive and systematic biological functional annotation information to users. Through the DAVID platform, users can access in-depth functional enrichment analysis, pathway analysis, interaction networks, and other comprehensive annotation results. These resources assist in gaining a deeper under-standing of the biological significance and interrelationships among genes or proteins.

Screening Hub Immune Infiltration-Related DEGs

Firstly, we performed protein-protein interaction (PPI) analysis of immune-related DEGs using STRING (http://string-db.org) with default parameters, and Cytoscape 3.9.1 software was used to construct and visualized. The betweenness centrality (BC) of genes was calculated by CytoNCA plug-in to rank key genes. In order to obtain the hub genes of immune infiltration, we used the least absolute shrinkage and selection operator (LASSO) regression analysis to select genes. The LASSO regression builds upon ordinary least squares by introducing a penalty term based on the sum of absolute values of regression coefficients. It could compress the coefficients of variables in the regression model, achieving variable selection at the cost of controlled estimation bias. We constructed a LASSO regression model using the "glmnet" package, specific parameter settings as follows: nlambda = 100, alpha = 1, and nfolds = 10. The final model was based on lambda.1se to select hub genes.

External Verification

This study enrolled a cohort of 22 patients who underwent comprehensive colonoscopy and pathological examination at the Digestive Endoscopy Center of the Renmin Hospital of Wuhan University (Wuhan, China) between January 2020 and December 2021. Among the participants, 12 patients were clinically diagnosed with UC based on their presenting symptoms, endoscopic phenotype, and pathological characteristics. The another 10 participants who had normal colonoscopy findings were included as the control group. Intestinal tissue samples were collected for immunohisto-chemical (IHC) analysis to further validate the difference in expression of hub immune infiltration-related genes between normal individuals and patirnts with UC. Additionally, blood routine parameters, C-reactive protein (CRP), and erythrocyte sedimentation rate (ESR) data during first hospitalization were collected. The systemic immune inflammation index (SII) was used to evaluate the inflammation status of participants, the specific calculation formula was based on the

formula proposed by Hu et al.[23] This study was approved by the Ethics Committee of Renmin Hospital of Wuhan University (WDRY2021-K025).

Immunohistochemical (IHC) Analysis

The specific IHC standard procedure refers to the previously reported procedure.[24] The primary antibodies used in the immunohistochemistry procedure were as follows: polyclonal rabbit anti-CD44 (Cat No. ab189524 Abcam), polyclonal rabbit anti-LYN (Cat No. 18135-1-AP PTG), and polyclonal rabbit anti-ITGA5 (Cat No. CY5979 Abways). Each sample had three visual fiields were chosen randomly at 400× magnification, and the integrated optical density (IOD) was semi-quantitatively measured using Image Pro Plus 6.0 software. The average IOD value from the three measurements was considered the semi-quantitative expression level of the target protein.

Statistical Analysis

Statistical analysis and visualization were performed using R version 4.1.3 and GraphPad Prism (Version 9.0). *T*-test was used for comparison between two groups, and one-way analysis of variance (ANOVA) was used for comparison between three or more groups. The correlation between the semi-quantitative expression level of the target protein (CD44, LYN, ITGA5) and CRP, ESR, and SII was performed using a Pearson test. All statistical tests were two-sided and P-values < 0.05 were considered as statistically significant.

Results

Results of Immune Infiltration Analysis

CIBERSORT was used to estimate the proportion of different immune cells infiltrating the sample, and the top five estimated percentages of infiltrating immune cells in the sample were Plasma cells, T cells follicular helper, Macrophages M2, Macrophages M1, and T cells CD8 (Figure 2A). Compared with control samples, patients with UC showed significantly increased levels of infiltration of immune cells (Figure 2B). The immune cells in the samples were then further classified using the ssGSEA. Again, the degree of infiltration of most immune cells increased significantly and was positively correlated with each other, especially Type T helper cells (Th cells), MDSC, and Activated dendritic cells (Figure 2C–E). However, CD56 dim natural killer cells were negatively correlated with other immune cells.

Co-Expression Network Construction

Firstly, a sample tree was constructed by sample clustering, and an outlier sample was found and removed (Figure 3A). Subsequently, the optimal soft threshold was determined to be 9 ($R^2 = 0.8$) to construct a scale-free network (Figure 3B). Network and recognition modules were further constructed, and 11 gene co-expression modules were generated (Figure 3C). Compared with other modules, the brown module is highly correlated with all types of immune cells (Figure 3D). Therefore, we selected the central genes of the brown module for further analysis to explore the role of hub genes in the immune infiltration mechanism of UC.

Screening of Immune Infiltration-Related DEGs and Functional Analysis

In the GSE92415 dataset, 873 DEGs were identified between the UC group and control group, including 556 up-regulated and 317 down-regulated DEGs (Figure 4A). The intersection of brown module central genes (293) and DEGs (873) resulted in 215 immune infiltration-related DEGs (Figure 4B). Results of GO showed that these genes are mainly involved in biological processes such as glycolipid binding, integrin binding, and toll-like receptor binding. Results of KEGG include inflammatory pathways such as the IL-17 signaling pathway and the NF-kB signaling pathway, as well as autoimmune diseases associated with rheumatoid arthritis (Figure 4C–E). We constructed a network of 215 immune infiltration-related DEGs by using STRING, and the hub genes were ranked according to the BC of genes (Figure 4F and G). Finally, four genes were screened by LASSO regression analysis as hub genes associated with immune infiltration in UC, namely CD44, IL1B, LYN, and ITGA5, respectively (Figure 4H and I).
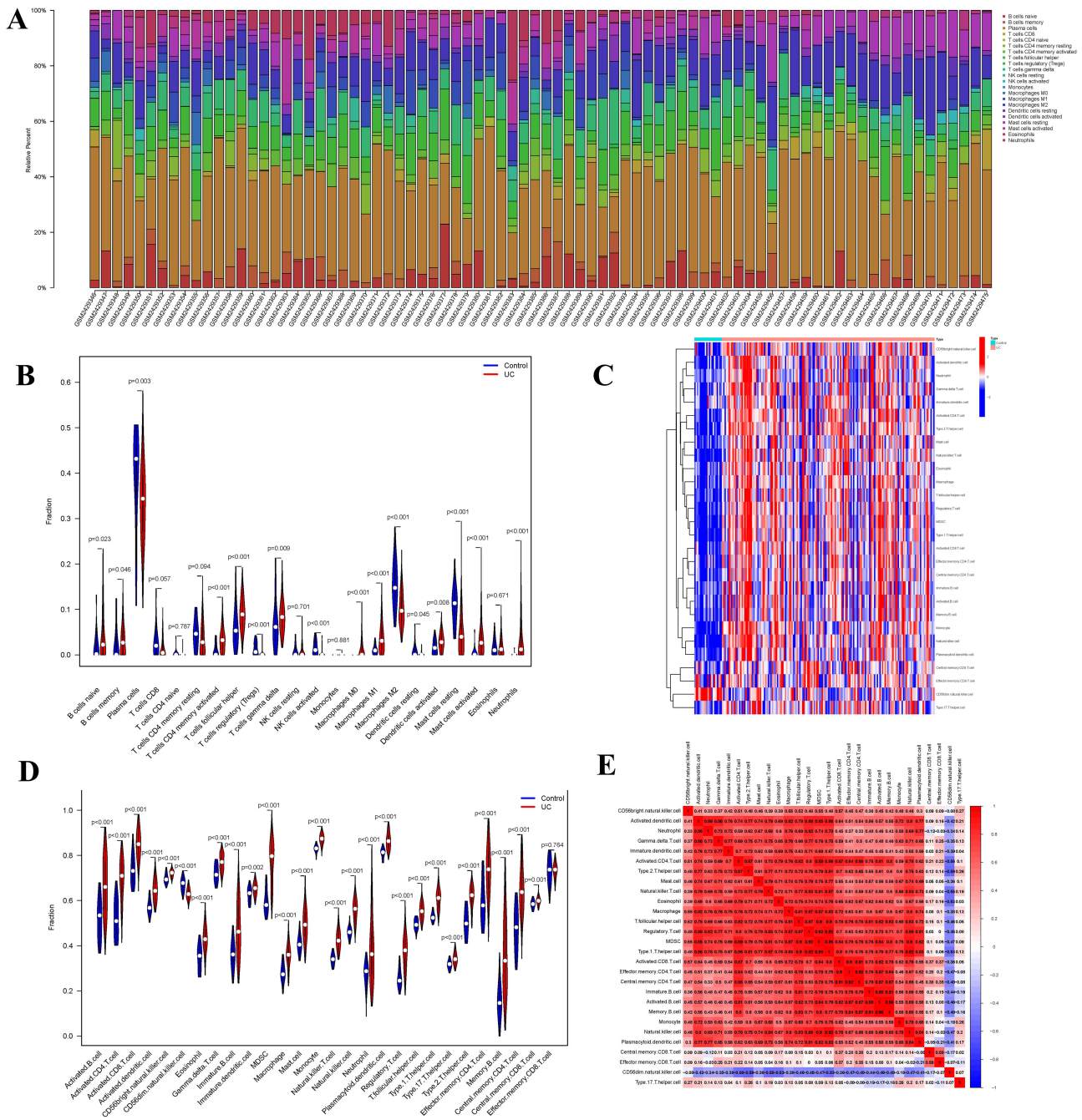
Figure 2 Evaluation of immune cell infiltration. (**A**) Proportion of immune cell infiltration (CIBERSORT) (**B**) Differences in immune cell infiltration between UC and control (CIBERSORT). (**C**) The heatmap of immune cell infiltration(ssGSEA). (**D**) Differences in immune cell infiltration between UC and control (ssGSEA). (**E**) Correlation between different immune cells, the value and color represent the correlation coefficient between immune cells (range −1 to 1). The closer to red the higher the positive correlation, the closer to purple the higher the negative correlation.

Correlation Analysis of Hub Genes with Immune Infiltration

Correlation analysis of CD44, IL1B, LYN and ITGA5 expression with the abundance of 28 types of immune cells showed that IL1B was most correlated with activated dendritic cells (r = 0.832, $P < 0.001$), CD44 was most correlated with natural killer cells (r = 0.854, $P <0.001$), LYN was most associated with activated dendritic cell (r = 0.902, $P <0.001$), and ITGA5 was most associated with Th1 cell (r = 0.859, $P < 0.001$) (Figure 5A–D). It should be pointed out that the expression of these genes was also significantly positively correlated with immune cells such as Th2 cells and Regulatory T cells, except for CD56dim natural killer cells (Figure 5A–D). In addition, there was a significant positive
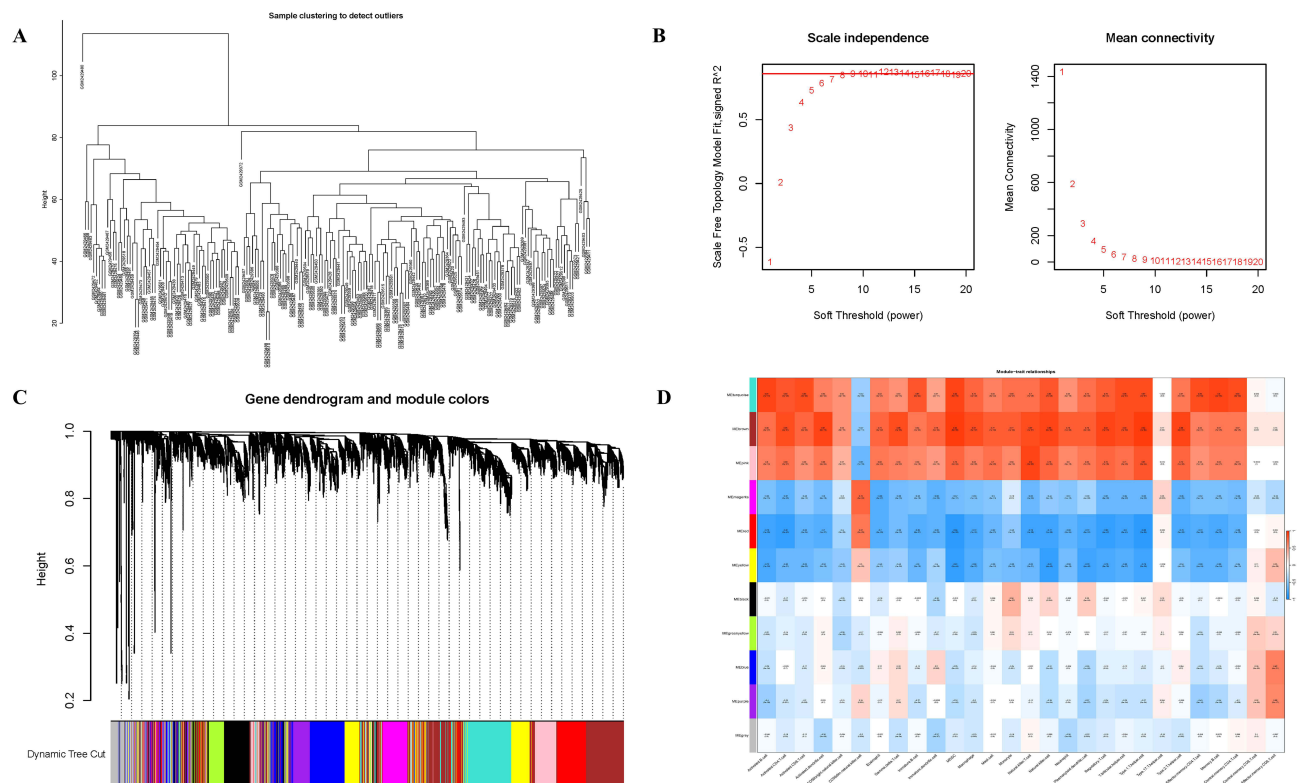
Figure 3 Results of Weighted gene co-expression network analysis. (**A**) Dendrogram of sample clustering. (**B**) Network topology analysis of various soft thresholds. (**C**) Clustering dendrogram of genes, different colors represent the obtained gene co-expression modules (A total of 11 modules). (**D**) Heatmap shows the correlation between gene co-expression modules and different immune cell (red for higher correlation, blue for lower correlation).

correlation between the expression of these four genes and the disease activity (Mayo score). The highest correlation was found in LYN (r = 0.44), followed by IL1B (0.39), CD44 (0.38), and ITGA5 (0.38) (all $P < 0.05$) (Figure 5E–H). To ensure the stability of the results obtained by the above method, we used the GSE206285 and GSE16879 datasets for verification. We assessed the immune infiltration of the samples using ssGSEA and then the WGCNA clustering module. The intersection of modules obtained from the three datasets obtained 38 common genes, including CD44, IL1B, etc (Supplement Figure 1A–C). Similarly, CD44, IL1B, LYN, and ITGA5 were highly correlated with all types of immune cells (Supplement Figure 1D–K).

The Value of Hub Genes in Clinical Treatment and Diagnosis

In the GSE92415 dataset, based on the responsiveness of UC patients to golimumab treatment at the 6th week, they were categorized as responders and non-responders. By comparing the gene expression levels of UC patients before and after treatment with control group, we observed significant upregulation of these four genes in UC patients (Figure 6A and B). Following golimumab treatment, in the responsive patients, disease activity significantly decreased, and the expression of CD44, LYN, and ITGA5 showed significant downregulation after treatment, except for IL1B (Figure 6A and C). In contrast, non-responsive patients showed no significant changes in disease activity or the expression of these four genes before and after treatment (Figure 6B and D). Additionally, we evaluated the diagnostic performance of individual genes in UC patients, and the results demonstrated that CD44, IL1B, LYN, and ITGA5 exhibited high diagnostic efficacy, with CD44 (AUC: 0.935, 95% CI: 0.892–0.971), IL1B (AUC: 0.961, 95% CI: 0.922–0.991), ITGA5 (AUC: 0.960, 95% CI: 0.917–0.991), and LYN (AUC: 0.942, 95% CI: 0.899–0.975) (Figure 6E–H). These results were confirmed in the validation dataset (Supplement Figure 2).
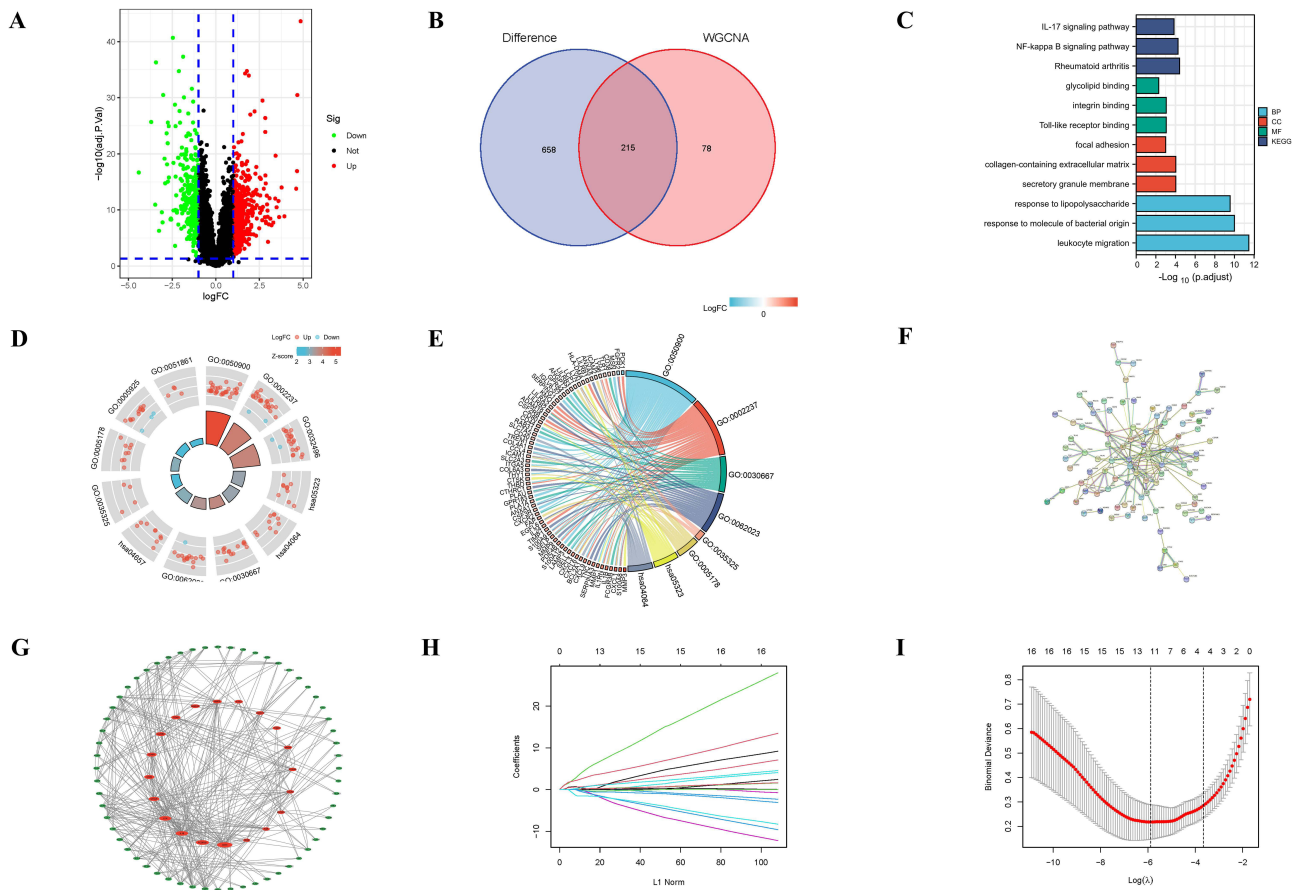
Figure 4 Screening and functional analysis of immune infiltration-related DEGs. (**A**)Volcanic map of DEGs (red for up-regulation and green for down-regulation). (**B**) Venn diagram of intersect genes (between DEGs (873) and the central genes of brown module (293). (**C**–**E**) Functional enrichment analysis of immune infiltration-related DEGs. (**F** and **G**) PPI networks of immune infiltration-related DEGs. (**H** and **I**) LASSO regression analysis (Each color represents a different gene in (**H**).
Abbreviations: DEGs, Differential Expressed Genes; BP, Biological Process; CC, Cellular Component; MF, Molecular Function; LASSO, least absolute shrinkage and selection operator.
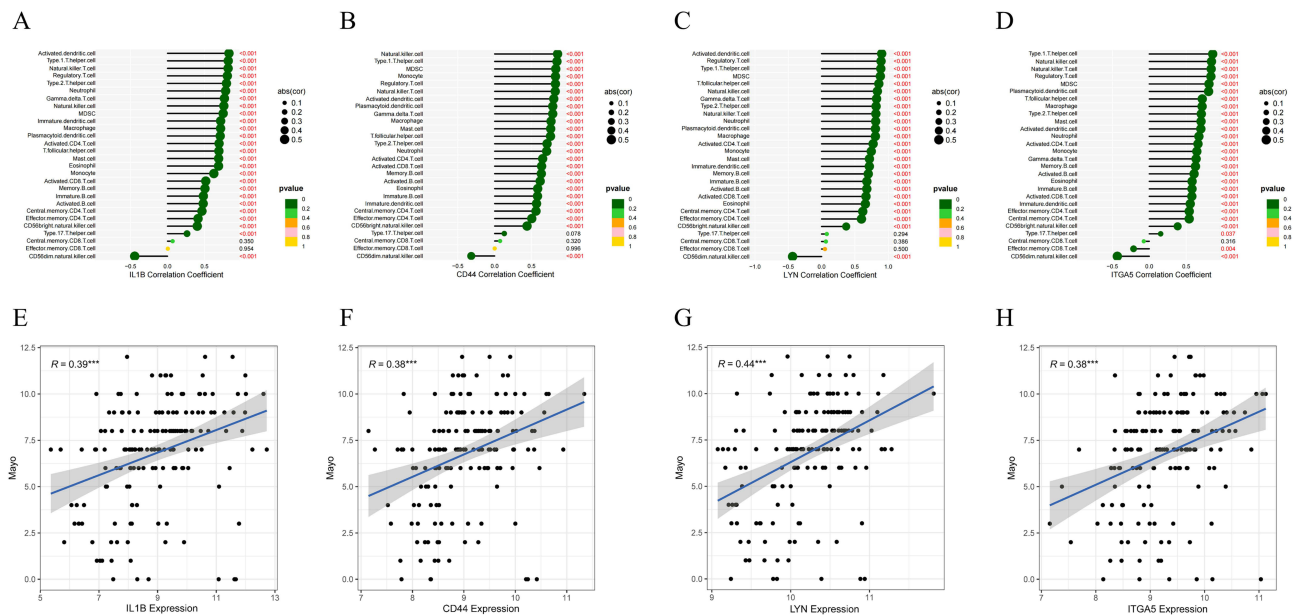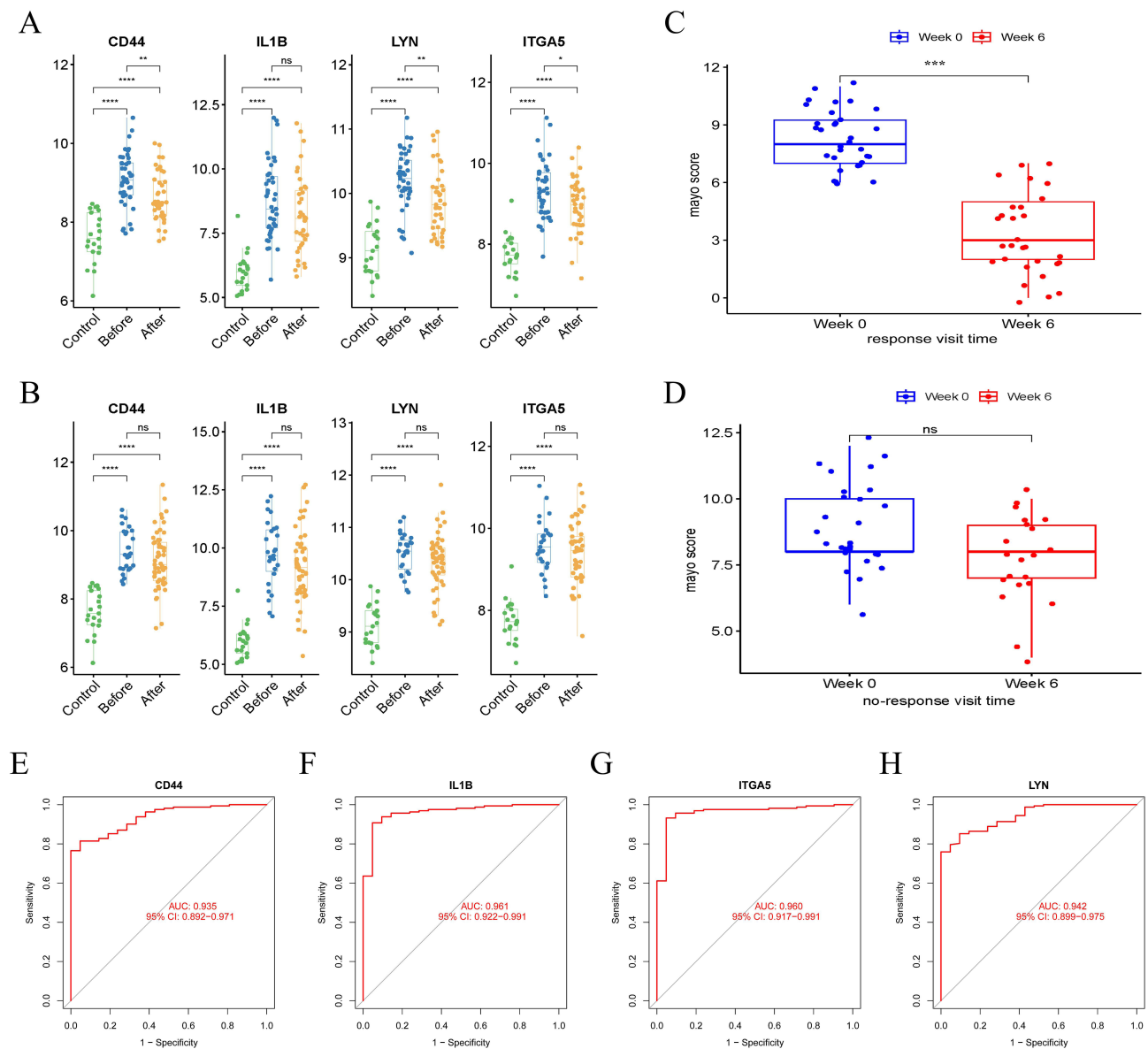

Figure 5 Correlation analysis of hub genes with immune infiltration. (**A**–**D**) Correlation between the expression of CD44, IL1B, LYN, and ITGA5 and immune cells. (The red value represents a significant correlation between gene expression and immune cells, ***$P < 0.001$). (**E**–**H**) Correlation analysis of CD44, IL1B, LYN, and ITGA5 expression with disease activity (Mayo score).

Figure 6 The value of hub genes in clinical treatment and diagnosis. (**A**) Gene expression levels in responsive UC patients before and after treatment and control group. (**B**) Gene expression levels in non-responsive UC patients before and after treatment and control group. (**C**) Disease activity in responsive UC patients before and after treatment. (**D**) Disease activity in non-responsive UC patients before and after treatment. (**E–H**) Diagnostic values of CD44, IL1B, LYN, and ITGA5 in the GSE92415. (ns: $P \geq 0.05$, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.001$).
Abbreviations: UC, ulcerative colitis; AUC: area under the curve.

The Expression of CD44, LYN, and ITGA5 in UC Tissue

Under endoscopy, we observed active bleeding and mucosal swelling in the colonic mucosa of patients with severe active UC (Figure 7A). After treatment, only small blood clots were observed on the mucosal surface, and the mucosa also recovered significantly (Figure 7B and C). Therefore, we validated our previous findings using specimens obtained from endoscopic biopsies at Renmin Hospital of Wuhan University. IHC results (Figure 7D–F) showed that the expression of CD44, LYN, and ITGA5 in UC's colon tissue was significantly increased compared to the corresponding normal colonic tissue (all $P < 0.05$) (Figure 7G–I), which was consistent with the previous bioinformatics results. Furthermore, we compared the levels of common inflammatory markers between the healthy control group and patients with UC and evaluated the associations between these genes and the common inflammatory markers. It was evident that the levels of inflammation markers were significantly elevated in patients with UC (Figure 8A–C). Correlation analysis showed that
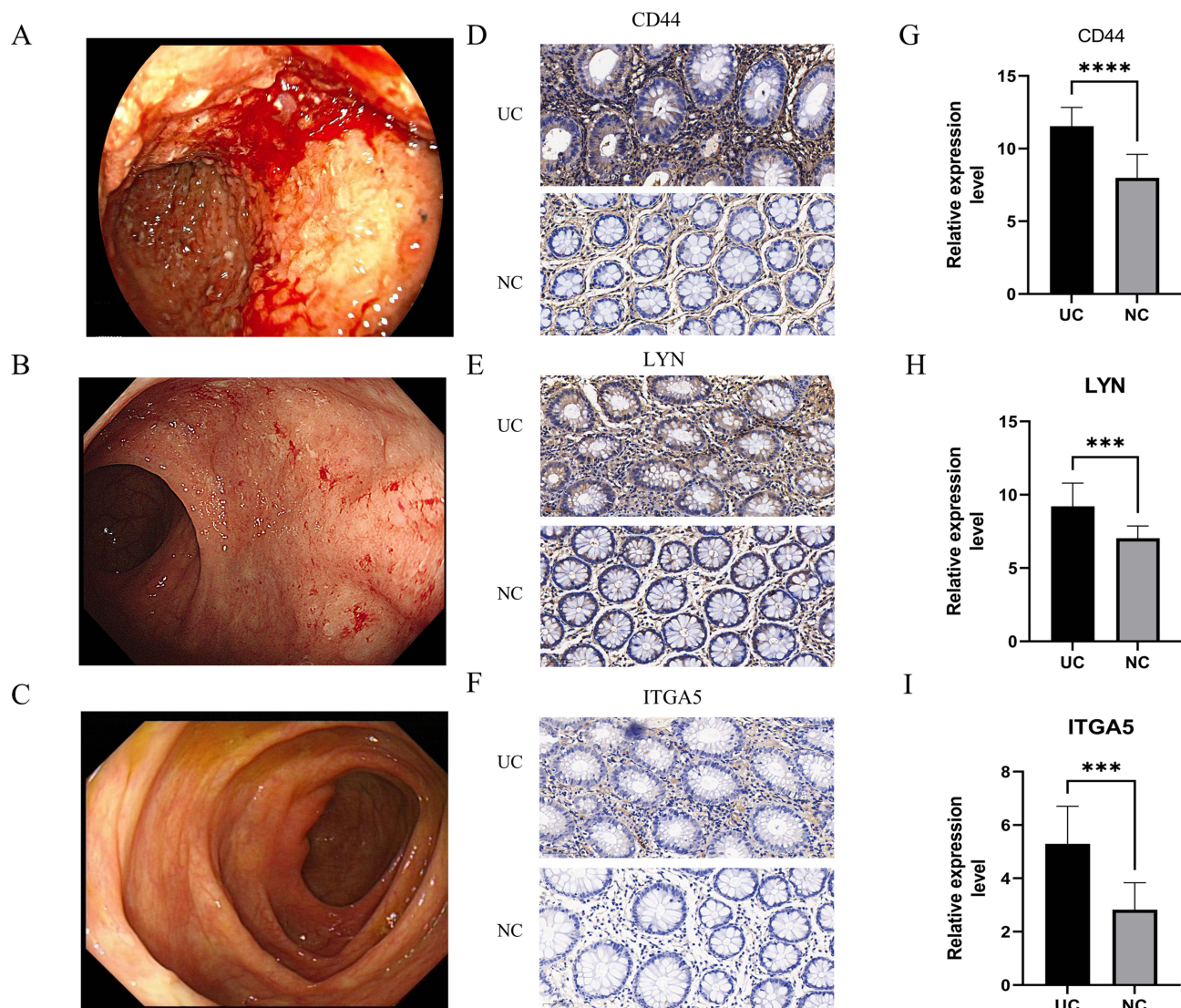
Figure 7 (**A**) Endoscopic image of severe UC. (**B**) Endoscopic image of mild UC. (**C**) Endoscopic image of UC in remission. (**D–F**) Expression of CD44, LYN and ITGA5 in UC tissues, (Magnification ×400). (**G–I**) Comparison of CD44, LYN and ITGA5 relative expression level between normal and UC intestinal tissues. (***P < 0.001, ****P < 0.001).
Abbreviations: IOD, Integrated optical density; NC, normal control; UC, ulcerative colitis.

CD44, LYN, and ITGA5 were significantly correlated with CRP, ESR, and SII (Figure 8D–L). Therefore, our results indicate the active involvement of CD44, LYN, and ITGA5 in the immune pathogenesis of UC.

Discussion

Immune factors are considered to be one of the most important factors contributing to UC. Based on intestinal mucosal barrier dysfunction, antigens stimulate the innate immune cells in the mucosa to produce various inflammatory factors, which in turn activate the body's acquired immunity, making the intestinal inflammation persist.[10] Therefore, identifying hub genes related to immune infiltration in the intestinal mucosa of patients with UC is not only of great significance for understanding the immune mechanism of UC patients but can also provide an important theoretical basis for new clinically targeted biological therapy for patients with UC.

In this study, we combined ssGSEA and CIBERSORT to explore the infiltration of various immune cells in the colonic mucosa of UC patients and the healthy individuals. As expected, dendritic cells, mast cells, monocytes, NK cells, etc., which constitute the innate immune system, and adaptive immune effector cells (including Th1 cell, Th2 cell, Th17
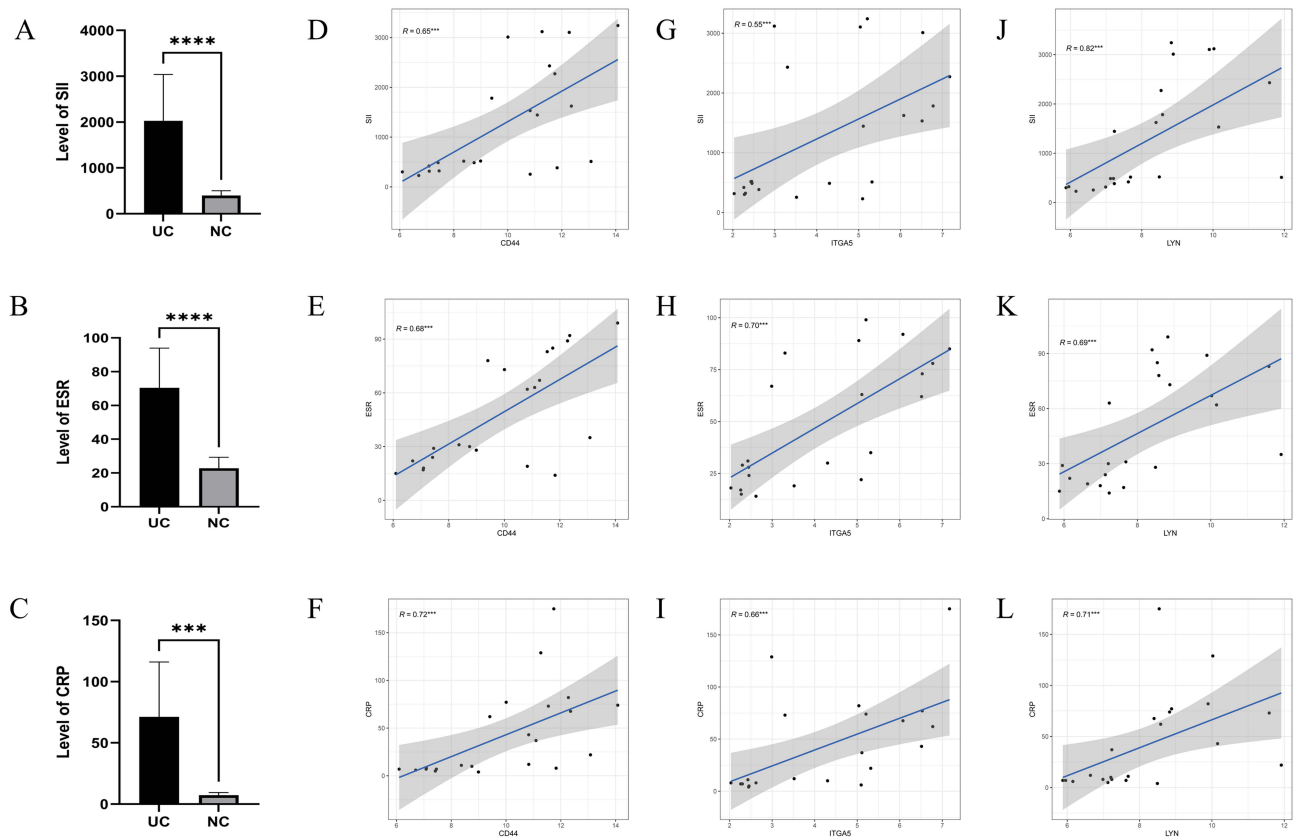
Figure 8 (**A**) Comparison of SII in UC with normal control. (**B**) Comparison of CRP in UC with normal control. (**C**) Comparison of ESR in UC with normal control. (**D–F**) Correlation analysis between CD44 and SII, CRP and ESR. (**G–I**) Correlation analysis between LYN and SII, CRP and ESR. (**J–L**) Correlation analysis between ITGA5 and SII, CRP and ESR. (***: $P < 0.001$, ****: $P < 0.001$).
Abbreviations: SII, systemic immune inflammation index; NC, normal control; UC, ulcerative colitis.

cells, etc.), infiltrate significantly more in the sample of patients with UC. Using this information (result of ssGSEA) as a phenotype, we obtained gene networks associated with immune infiltration through WGCNA. The brown modules were highly correlated with different types of immune cells. Therefore, these co-expressed genes may be the key factors involved in the immune infiltration of UC. The results of gene function annotation showed major enrichment to some inflammatory signaling pathways and autoimmune disease-related pathways, such as the NF-κB signaling pathway, IL-17 signaling pathway, and rheumatoid arthritis (RA). The activation of the NF-κB signaling pathway has been recognized as a pivotal mechanism underlying the mucosal injury observed in UC for a long time. NF-κB, a transcription factor, exerts a significant influence on the immune response by stimulating the production and secretion of pro-inflammatory cytokines, including tumor necrosis factor-alpha (TNF-α), interleukin-1 (IL-1), and interleukin-6 (IL-6), among others. Of particular importance is the role of TNF-α in amplifying the activation of NF-κB, establishing a positive feedback loop that perpetuates mucosal inflammation and contributes to the chronic nature of UC.[25,26] The IL-23/IL-17 axis has recently attracted considerable attention in the understanding of autoimmune diseases such as IBD and RA. Dysregulation of this axis has been implicated in the pathogenesis of various autoimmune diseases, including UC. It is thought that the IL-23/IL-17 axis promotes the activation of immune cells and subsequent production of pro-inflammatory cytokines, ultimately contributing to the progression and persistence of UC and other autoimmune diseases.[27]

Further through LASSO regression analysis, CD44, IL1B, LYN, and ITGA5 were finally screened, which were the key genes involved in UC immune infiltration. After treatment with golimumab, expression levels of CD44, LYN, and ITGA5 decreased significantly except for IL1B. Therefore, we collected a subset of clinical samples for CD44, LYN, and ITGA5 validation. The results showed that the expression level of these genes was significantly increased in UC and was

highly correlated with the level of inflammatory markers (CRP, ESR, and SII). CD44 is a cell surface glycoprotein that plays a crucial role in cell adhesion, migration, and inflammation.[28] CD44 is involved in leukocyte recruitment and migration to inflamed tissues. During inflammation, the expression of CD44 is upregulated on the endothelial cells and leukocytes, facilitating their adhesion to the endothelium and subsequent migration into the inflamed tissue. This process is mediated by the interaction of CD44 with its ligands, such as hyaluronic acid (HA) and osteopontin.[29] In addition, the combination of CD44 and HA can enhance the activation of T cells, and the up-regulation of CD44 will promote the differentiation of Th1/Th17.[30,31] Similarly, LYN, as an important regulator of immune receptor signaling, also plays an important role in the activation and suppression of the immune system.[32] LYN can regulate the differentiation of Th2 cells by inhibiting the expression of GATA-3 to affect the balance between Th1/Th2 cells.[33] When LYN is absent, the excessive activation of Toll-Like Receptor signaling mediated by dendritic cells will cause severe immune imbalance, leading to the occurrence of autoimmune diseases.[34] The precise mechanisms underlying the positive and negative regulation of immune responses by Lyn remain elusive. In the intestine, LYN deficiency can exacerbate the development of inflammation. Conversely, when LYN is upregulated, LYN promotes the production of IL-22 to enhance the barrier function of the intestinal mucosa.[35,36] ITGA5 is an important member of the integrin family, which is involved in the regulation of cell adhesion, diffusion, migration, growth, invasion, and other functions.[37] The association of ITGA5 with immune cell infiltration may be related to its role in regulating the recruitment and function of immune cells by fibroblasts.[38] In addition, ITGA5 is expressed on the surface of tumor-associated macrophages (TAM), which directly regulates the recruitment and alternative activation of immune cells.[39] In gastrointestinal tumors, the expression level of ITGA5 is positively correlated with immune markers of TAM, macrophage 2 (M2) cells, and Th2 cells, and is associated with the prognosis of the disease.[40] Therefore, CD44, LYN, and ITGA5 may play a key role in the immune infiltration of UC.

In this study, we used ssGSEA and CIBERSORT to initially analyze the immune infiltrative environment in UC and identified three hub genes associated with immune infiltration by integrated WGCNA and LASSO regression analysis. In addition, we preliminarily explored the relevance of hub genes to existing clinical information and validated them using multiple other data sets including UC clinical sample. However, there are some limitations to our work. First, we used data from public databases that do not allow us to understand patients' choice of treatment regimen before golimumab treatment. Both CIBERSORT and ssGSEA estimate the relative proportion of immune cells rather than the absolute cell count, and there are still a large number of other epithelial cell disturbances in the intestine, which may affect our results. The clinical sample size we collected for verification was relatively small. Our findings need to be validated using larger cohorts and prospective studies. In addition, we only preliminarily explored the correlation between key genes and immune infiltration, and further studies on the specific mechanisms are needed in cell and animal experiments.

Conclusion

We obtained three hub genes (CD44, LYN, and ITGA5) that are highly associated with UC immune infiltration. These genes are important in the pathogenesis of UC and are highly associated with body's inflammatory status . Therefore, these hub genes may be new immunotherapeutic targets for UC.

Abbreviations

UC, ulcerative colitis; CD, Crohn's disease; ssGSEA, single-sample gene set enrichment analysis; DEGs, Differential Expressed Genes, BP, Biological Process, CC, Cellular Component, MF, Molecular Function, LASSO, least absolute shrinkage and selection operator; IHC, immunohistochemical; IOD, integrated optical density; SII, systemic immune inflammation index.

Data Sharing Statement

The datasets had analyzed during the current study are available in the Gene Expression Omnibus (GEO). The data used in this study can be accessed at the following link: GSE92415 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE92415), GSE206285 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE206285), andGSE1687 (https://

Ethics Approval and Consent to Participate

This study was conducted in accordance with the Declaration of Helsinki. This study was approved by ethics committee of the Renmin Hospital of Wuhan University (WDRY2021-K025). All subjects signed a document of informed consent.

Acknowledgments

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Disclosure

The authors have no relevant financial or non-financial interests to disclose.

References

1. Feuerstein JD, Moss AC, Farraye FA. Ulcerative Colitis. *Mayo Clin Proc*. 2019;94(7):1357–1373. doi:10.1016/j.mayocp.2019.01.018
2. Alatab S, Sepanlou SG, Ikuta K, Collaborators GBDIBD. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. *Lancet Gastroenterol Hepatol*. 2020;5(1):17–30. doi:10.1016/S2468-1253(19)30333-4
3. Piovani D, Danese S, Peyrin-Biroulet L, Bonovas S. Inflammatory bowel disease: estimates from the global burden of disease 2017 study. *Aliment Pharmacol Ther*. 2020;51(2):261–270. doi:10.1111/apt.15542
4. Wei SC, Sollano J, Hui YT, et al. Epidemiology, burden of disease, and unmet needs in the treatment of ulcerative colitis in Asia. *Expert Rev Gastroenterol Hepatol*. 2021;15(3):275–289. doi:10.1080/17474124.2021.1840976
5. Ni J, Wu GD, Albenberg L, Tomov VT. Gut microbiota and IBD: causation or correlation? *Nat Rev Gastroenterol Hepatol*. 2017;14(10):573–584. doi:10.1038/nrgastro.2017.88
6. Mirkov MU, Verstockt B, Cleynen I. Genetics of inflammatory bowel disease: beyond NOD2. *Lancet Gastroenterol Hepatol*. 2017;2(3):224–234. doi:10.1016/S2468-1253(16)30111-X
7. de Souza HS, Fiocchi C. Immunopathogenesis of IBD: current state of the art. *Nat Rev Gastroenterol Hepatol*. 2016;13(1):13–27. doi:10.1038/nrgastro.2015.186
8. Camelo A, Barlow JL, Drynan LF, et al. Blocking IL-25 signalling protects against gut inflammation in a type-2 model of colitis by suppressing nuocyte and NKT derived IL-13. *J Gastroenterol*. 2012;47(11):1198–1211. doi:10.1007/s00535-012-0591-2
9. Guan Q. A comprehensive review and update on the pathogenesis of inflammatory bowel disease. *J Immunol Res*. 2019;2019:7247238. doi:10.1155/2019/7247238
10. Choy MC, Visvanathan K, De Cruz P. An overview of the innate and adaptive immune system in inflammatory bowel disease. *Inflamm Bowel Dis*. 2017;23(1):2–13. doi:10.1097/MIB.0000000000000955
11. Jeong DY, Kim S, Son MJ, et al. Induction and maintenance treatment of inflammatory bowel disease: a comprehensive review. *Autoimmun Rev*. 2019;18(5):439–454. doi:10.1016/j.autrev.2019.03.002
12. Marafini I, Sedda S, Dinallo V, Monteleone G. Inflammatory cytokines: from discoveries to therapies in IBD. *Expert Opin Biol Ther*. 2019;19(11):1207–1217. doi:10.1080/14712598.2019.1652267
13. Armuzzi A, Liguori G. Quality of life in patients with moderate to severe ulcerative colitis and the impact of treatment: a narrative review. *Dig Liver Dis*. 2021;53(7):803–808. doi:10.1016/j.dld.2021.03.002
14. Le T, Aronow RA, Kirshtein A, Shahriyari L. A review of digital cytometry methods: estimating the relative abundance of cell types in a bulk of cells. *Brief Bioinform*. 2021;22(4). doi:10.1093/bib/bbaa219
15. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–457. doi:10.1038/nmeth.3337

16. Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*. 2009;462(7269):108–112. doi:10.1038/nature08460

17. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf*. 2008;9:559. doi:10.1186/1471-2105-9-559

18. Yin L, Chen Y, Fu T, Liu L, Xia Q. Identification of candidate blood biomarkers for the diagnosis of septicaemic melioidosis based on WGCNA. *Artif Cells Nanomed Biotechnol*. 2022;50(1):252–259. doi:10.1080/21691401.2022.2126490

19. Rezaei Z, Ranjbaran J, Safarpour H, et al. Identification of early diagnostic biomarkers via WGCNA in gastric cancer. *Biomed Pharmacother*. 2022;145:112477. doi:10.1016/j.biopha.2021.112477

20. Sandborn WJ, Feagan BG, Marano C, et al. Subcutaneous golimumab induces clinical response and remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology*. 2014;146(1):85–95; quiz e14–85. doi:10.1053/j.gastro.2013.05.048

21. Pavlidis P, Tsakmaki A, Pantazi E, et al. Interleukin-22 regulates neutrophil recruitment in ulcerative colitis and is associated with resistance to ustekinumab therapy. *Nat Commun*. 2022;13(1):5820. doi:10.1038/s41467-022-33331-8

22. Arijs I, De Hertogh G, Lemaire K, et al. Mucosal gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infliximab treatment. *PLoS One*. 2009;4(11):e7984. doi:10.1371/journal.pone.0007984

23. Hu B, Yang XR, Xu Y, et al. Systemic immune-inflammation index predicts prognosis of patients after curative resection for hepatocellular carcinoma. *Clin Cancer Res*. 2014;20(23):6212–6222. doi:10.1158/1078-0432.CCR-14-0442

24. Jiang C, Zhang H, Wu W, et al. Immune Characteristics of LYN in Tumor Microenvironment of Gliomas. *Front Cell Dev Biol*. 2021;9:760929. doi:10.3389/fcell.2021.760929

25. Atreya I, Atreya R, Neurath MF. NF-kappaB in inflammatory bowel disease. *J Intern Med*. 2008;263(6):591–596. doi:10.1111/j.1365-2796.2008.01953.x

26. Barnabei L, Laplantine E, Mbongo W, Rieux-Laucat F, Weil R. NF-kappaB: at the borders of autoimmunity and inflammation. *Front Immunol*. 2021;12:716469. doi:10.3389/fimmu.2021.716469

27. Schinocca C, Rizzo C, Fasano S, et al. Role of the IL-23/IL-17 pathway in rheumatic diseases: an overview. *Front Immunol*. 2021;12:637829. doi:10.3389/fimmu.2021.637829

28. Chen C, Zhao S, Karnad A, Freeman JW. The biology and role of CD44 in cancer progression: therapeutic implications. *J Hematol Oncol*. 2018;11(1):64. doi:10.1186/s13045-018-0605-5

29. van Steen ACI, Gronloh MLB, Joosten S, et al. Endothelial ICAM-1 Adhesome Recruits CD44 for optimal transcellular migration of human CTLs. *J Immunol*. 2023;211:377–388. doi:10.4049/jimmunol.2200761

30. Nagy N, Kuipers HF, Marshall PL, Wang E, Kaber G, Bollyky PL. Hyaluronan in immune dysregulation and autoimmune diseases. *Matrix Biol*. 2019;78-79:292–313. doi:10.1016/j.matbio.2018.03.022

31. Guan H, Nagarkatti PS, Nagarkatti M. CD44 Reciprocally regulates the differentiation of encephalitogenic Th1/Th17 and Th2/regulatory T cells through epigenetic modulation involving DNA methylation of cytokine gene promoters, thereby controlling the development of experimental autoimmune encephalomyelitis. *J Immunol*. 2011;186(12):6955–6964. doi:10.4049/jimmunol.1004043

32. Scapini P, Pereira S, Zhang H, Lowell CA. Multiple roles of Lyn kinase in myeloid cell signaling and function. *Immunol Rev*. 2009;228(1):23–40. doi:10.1111/j.1600-065X.2008.00758.x

33. Charles N, Watford WT, Ramos HL, et al. Lyn kinase controls basophil GATA-3 transcription factor expression and induction of Th2 cell differentiation. *Immunity*. 2009;30(4):533–543. doi:10.1016/j.immuni.2009.02.008

34. Lamagna C, Scapini P, van Ziffle JA, DeFranco AL, Lowell CA. Hyperactivated MyD88 signaling in dendritic cells, through specific deletion of Lyn kinase, causes severe autoimmunity and inflammation. *Proc Natl Acad Sci U S A*. 2013;110(35):E3311–3320. doi:10.1073/pnas.1300617110

35. Bishop JL, Roberts ME, Beer JL, et al. Lyn activity protects mice from DSS colitis and regulates the production of IL-22 from innate lymphoid cells. *Mucosal Immunol*. 2014;7(2):405–416. doi:10.1038/mi.2013.60

36. Roberts ME, Bishop JL, Fan X, et al. Lyn deficiency leads to increased microbiota-dependent intestinal inflammation and susceptibility to enteric pathogens. *J Immunol*. 2014;193(10):5249–5263. doi:10.4049/jimmunol.1302832

37. Bianconi D, Unseld M, Prager GW. Integrins in the Spotlight of Cancer. *Int J Mol Sci*. 2016;17(12):2037. doi:10.3390/ijms17122037

38. Lu L, Xie R, Wei R, et al. Integrin α5 subunit is required for the tumor supportive role of fibroblasts in colorectal adenocarcinoma and serves as a potential stroma prognostic marker. *Mol Oncol*. 2019;13(12):2697–2714. doi:10.1002/1878-0261.12583

39. Kuninty PR, Bansal R, De Geus SWL, et al. ITGA5 inhibition in pancreatic stellate cells attenuates desmoplasia and potentiates efficacy of chemotherapy in pancreatic cancer. *Sci Adv*. 2019;5(9):eaax2770. doi:10.1126/sciadv.aax2770

40. Zhu H, Wang G, Zhu H, Xu A. ITGA5 is a prognostic biomarker and correlated with immune infiltration in gastrointestinal tumors. *BMC Cancer*. 2021;21(1):269. doi:10.1186/s12885-021-07996-1