Check for updates

OPEN ACCESS

EDITED BY Pradeep Kumar, Veer Bahadur Singh Purvanchal University, India

REVIEWED BY Palash Mandal, Charotar University of Science and Technology, India Giovanni Tarantino, University of Naples Federico II, Italy

*CORRESPONDENCE Fating Lu ⊠ lft922@hotmail.com Tian E. Zhang ⊠ zhte2003@cdutcm.edu.cn

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 07 November 2023 ACCEPTED 21 October 2024 PUBLISHED 06 November 2024

CITATION

Luo Z, Huang C, Chen J, Chen Y, Yang H, Wu Q, Lu F and Zhang TE (2024) Potential diagnostic markers and therapeutic targets for non-alcoholic fatty liver disease and ulcerative colitis based on bioinformatics analysis and machine learning. *Front. Med.* 11:1323859. doi: 10.3389/fmed.2024.1323859

COPYRIGHT

© 2024 Luo, Huang, Chen, Chen, Yang, Wu, Lu and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Potential diagnostic markers and therapeutic targets for non-alcoholic fatty liver disease and ulcerative colitis based on bioinformatics analysis and machine learning

Zheng Luo^{1†}, Cong Huang^{1,2†}, Jilan Chen^{1,2}, Yunhui Chen¹, Hongya Yang¹, Qiaofeng Wu¹, Fating Lu^{1*} and Tian E. Zhang^{1,2*}

¹Chengdu University of Traditional Chinese Medicine, Chengdu, China, ²Key Biology Laboratory for TCM Viscera-Manifestation Research of Sichuan University, Chinese Medical Center of Chengdu University of TCM, Chengdu, China

Background: Non-alcoholic fatty liver disease (NAFLD) and ulcerative colitis (UC) are two common health issues that have gained significant global attention. Previous studies have suggested a possible connection between NAFLD and UC, but the underlying pathophysiology remains unclear. This study investigates common genes, underlying pathogenesis mechanisms, identification of diagnostic markers applicable to both conditions, and exploration of potential therapeutic targets shared by NAFLD and UC.

Methods: We obtained datasets for NAFLD and UC from the GEO database. The DEGs in the GSE89632 dataset of the NAFLD and GSE87466 of the UC dataset were analyzed. WGCNA, a powerful tool for identifying modules of highly correlated genes, was employed for both datasets. The DEGs of NAFLD and UC and the modular genes were then intersected to obtain shared genes. Functional enrichment analysis was conducted on these shared genes. Next, we utilize the STRING database to establish a PPI network. To enhance visualization, we employ Cytoscape software. Subsequently, the Cytohubba algorithm within Cytoscape was used to identify central genes. Diagnostic biomarkers were initially screened using LASSO regression and SVM methods. The diagnostic value of ROC curve analysis was assessed to detect diagnostic genes in both training and validation sets for NAFLD and UC. A nomogram was also developed to evaluate diagnostic efficacy. Additionally, we used the CIBERSORT algorithm to explore immune infiltration patterns in both NAFLD and UC samples. Finally, we investigated the correlation between hub gene expression, diagnostic gene expression, and immune infiltration levels.

Results: We identified 34 shared genes that were found to be associated with both NAFLD and UC. These genes were subjected to enrichment analysis, which revealed significant enrichment in several pathways, including the IL-17 signaling pathway, Rheumatoid arthritis, and Chagas disease. One optimal candidate gene was selected through LASSO regression and SVM: CCL2. The ROC curve confirmed the presence of CCL2 in both the NAFLD and UC training sets and other validation sets. This finding was further validated using a nomogram in the validation set. Additionally, the expression levels of CCL2 for NAFLD and UC showed a significant correlation with immune cell infiltration.

Conclusion: This study identified a gene (*CCL2*) as a biomarker for NAFLD and UC, which may actively participate in the progression of NAFLD and UC. This discovery holds significant implications for understanding the progression of these diseases and potentially developing more effective diagnostic and treatment strategies.

KEYWORDS

bioinformatics, non-alcoholic fatty liver disease, ulcerative colitis, machine learning, diagnosis, immune infiltration

1 Introduction

Non-alcoholic fatty liver disease (NAFLD) is a prevalent condition characterized by the accumulation of fat in the liver cells, known as steatosis (1). NAFLD includes a range of diseases ranging from NAFL to NASH, which involves both fat accumulation and inflammation in the liver. NASH can progress to more severe forms of liver damage, including cirrhosis and HC. The global prevalence of NAFLD is estimated to be around 25%, making it one of the most common chronic liver diseases worldwide (2). One concerning aspect of NAFLD is its association with various extrahepatic complications. Research has shown that individuals with NAFLD have an increased risk of developing CVD, CKD, and T2DM (3).

Ulcerative colitis (UC), a chronic inflammatory bowel disease, caused by multiple reasons and mediated by abnormal immunity. The main manifestations are recurrent diarrhea, abdominal pain, and mucopurulent bloody stools (4). Looking toward future trends, it is estimated that there will be approximately 5 million cases of UC worldwide by 2023 (5). In recent years, more and more studies have pointed out that Irritable bowel syndrome (IBS) and IBD (including Crohn's disease and UC) have partial overlap in clinical manifestations and pathophysiological mechanisms. For example, studies have found that IBS-like symptoms are not uncommon in patients with IBD, especially when IBD is in remission, and this overlap can lead to misdiagnosis and mistreatment (6, 7). In addition, intestinal microecological disorder, intestinal barrier dysfunction and low-grade chronic inflammation may be the common underlying pathological mechanisms (8). Due to the high prevalence of IBS, this greatly increases the prevalence of NAFLD due to intestinal causes (9).

Recent studies have highlighted the co-existence of NAFLD and UC (7). NAFLD is a prevalent condition among individuals with UC, affecting up to one-third of UC patients globally. Moreover, research

has shown that UC patients are twice as likely to develop NAFLD compared to healthy individuals (8). In recent years, studies have also suggested a potential association between UC and NAFLD through the enteric-liver axis (10). Chronic intestinal inflammation may cause bacterial products (such as endotoxins) to enter the liver through the portal vein by disrupting the intestinal barrier function, thereby triggering an inflammatory response in the liver and promoting the occurrence and development of NAFLD (11, 12). In this complex process, the immune system's involvement is equally important (13). Therefore, based on the complex crosstalk between two diseases, Identifying factors related to the co-development of two syndromes and identifying common pathophysiological pathways can represent diagnostic and treatment strategies, which is of great significance.

Advancements in the field of bioinformatics have revolutionized our ability to study coexpressed genes that are associated with NAFLD and UC. These tools have opened up new opportunities for researchers to delve deeper into the molecular mechanisms underlying these diseases and identify potential targets for treatment (14). We analyzed the coexpressed genes of NAFLD and UC using bioinformatics tools. By examining common genes across both NAFLD and UC, we have identified diagnostic markers that could indicate disease presence or progression. Additionally, we provide valuable information about immune infiltration within both affected NAFLD and UC. This research has significant implications for improving patient outcomes through more targeted therapies explicitly tailored toward individuals with NAFLD and UC. The workflow can be seen in Figure 1.

2 Materials and methods

2.1 Data acquisition

Four datasets were downloaded from the GEO database (15). The GSE89632 (16) dataset comprises 24 HC patients and 39 NAFLD patients. The GSE87466 (17) dataset comprises 21 HC and 87 UC patients. Additionally, we also obtained the GSE48452 (18) dataset consisting of liver tissues from 32 NAFLD patients and 41 HC patients, as well as the GSE92415 (19) dataset consisting of liver tissues from 162 UC patients and 21 HC patients for external validation purposes. Details for the data sets were provided in Table 1.

2.2 Identification of DEGs

We utilized the "limma" R package to analyze the GSE89632 dataset and identify DEGs between the NAFLD and HC groups. The criteria for selecting DEGs were set at $|\log 2 \text{ FC}| \ge 0.886$ or 1 and p-adjust <0.05.

Abbreviations: AUC, Area under the curve; BP, Biological process; CC, Cellular component; CCL2, C-C motif chemokine 2; CKD, chronic kidney disease; OVD, cardiovascular disease; DEGs, differentially expressed genes; GEO, Gene Expression Omnibus; GO, Gene ontology; GSEA, gene set enrichment analysis; HC, hepatocellular carcinoma; IL-1 β , interleukin 1 beta; INF- α , interferon- α ; IBS, Irritable bowel syndrome; LASSO, Least Absolute Shrinkage and Selection Algorithm; MF, Molecular function; NAFL, Non- alcoholic simple fatty liver; NASH, non-alcoholic steatohepatitis; NF κ B, nuclear factor kappa B; oxLDL, oxidized low-density lipoprotein cholesterol; PAMP, permeability of the increased exposure of immune cells to microbial-associated molecular patterns; SVM-RFE, Support Vector Machine Recursive Feature Elimination; T2DM, type 2 diabetes mellitus; TLR4, Toll-like receptor 4; TNF- α , tumor necrosis factor alpha; WGCNA, Weighted gene co-expression network analysis.



TABLE 1 Detailed information of datasets used in this study.

ID	Organism	Platform	Normal vs. NAFLD	Type of samples
GSE89632	Human (samples of liver)	GPL14951	24 vs. 39 (62)	NAFLD
GSE87466	Human (Mucosal biopsy samples)	GPL13158	21 vs. 87 (108)	UC
GSE48452	Human (samples of liver)	GPL11532	41 vs. 32 (73)	NAFLD
GSE92415	Human (Mucosal biopsy samples)	GPL13158	21 vs. 162(183)	UC

To visualize the DEGs, a volcano plot was generated using the "ggplot2" package. Additionally, a heatmap was created to display the top 25 DEGs. Additionally, we applied the same methodology described above to analyze another dataset called GSE87466.

2.3 Weighted gene co-expression network analysis

The WGCNA method was utilized to identify functional modules. To ensure the accuracy of the analysis, only the top 25% of genes with the highest variation were chosen for further investigation. Next, to eliminate any ineligible genes and samples, the goodSamplesGenes function was employed, creating a scale-free coexpression network. Subsequently, the soft powers β were determined using the pickSoftThreshold function. A dynamic tree-cutting technique was applied to detect gene modules within the coexpression network. Additionally, these identified gene modules will be linked to clinical features. Finally, a Venn diagram on a bioinformatics platform (20) displayed the overlapping genes between NAFLD's green module genes with DEGs.

2.4 Functional enrichment analysis

We used GSEA with version 4.8.2 of the "clusterProfiler" package to identify KEGG pathways enriched by key genes in NAFLD and UC. Additionally, we conducted GO and KEGG enrichment analyses on the Metascape platform (21) to explore the functional biological roles of co-expressed genes. We applied a significance threshold of *p*-value <0.01 for these analyses. The results were then visualized using Sankey dot pathway enrichment diagrams on a bioinformatics platform.

2.5 Construction of PPI network and identification of hub genes

To analyze the interactions between commonly recognized genes, the string platform was used to construct PPI networks. Setting screening criteria with interaction scores exceeding 0.4 is significant. Next, Cytoscape software is utilized to visualize. We applied the cytoHubba plugin in Cytoscape, which calculates centrality measures based on maximal cliques within a



network, allowing for identification of highly connected nodes or hub genes.

calibration curve—box plots generated by ggplot2 visualized expression levels of diagnostic genes in different groups or conditions. Statistical significance was denoted as *p < 0.05 and **p < 0.01.

2.6 Machine learning for screening key genes and validation

The LASSO and the SVM algorithm were used to identify key genes in GSE89632 and GSE87466. Tw o independent datasets, GSE48452 and GSE92415, were used to validate the critical genes identified. The "glmnet" R package was employed for the LASSO logistic regression analysis. The smallest lambda value obtained from this analysis was considered optimal for predicting disease outcomes. SVM, powerful machine learning algorithms, and RFE that select a subset of relevant genes by iteratively eliminating less informative ones based on five-fold cross-validation were employed to acquire critical genes. The ROC curves were generated using the pROC package to evaluate diagnostic accuracy. The "Rms" package is utilized to construct a nomograph and assess its predictive capability through the

2.7 Immune infiltration analysis

The abundance of 22 subtypes of immune cells in NAFLD samples and UC samples was determined using the cibersort algorithm from the "IOBR" package in R software and the CIBERSOTRx platform.¹ This analysis provided insights into the cellular composition of the immune microenvironment in both NAFLD and UC. Additionally, we performed Spearman correlation analysis to investigate potential relationships between hub genes and different subsets of immune cells.

¹ http://cibersort.stanford.edu/

3 Results

3.1 Screening of DEGs in NAFLD and UC datasets

903 DEGs were found in the GSE89632 dataset (16 healthy individuals and 16 NAFLD patients were selected). Among these DEGs, 623 genes were upregulated, while 280 genes were down-regulated. The significance of these DEGs is further highlighted in Figure 2A, where the top 25 most significant DEGs are displayed as a heatmap. Figure 2C shows the DEGs by the volcano diagram. We obtained 882 UC-related

DEGs in the GSE87466 dataset (21 healthy individuals and 21 patients with UC were selected). This analysis showed more downregulated genes (643) than upregulated ones (239). The heatmap and volcano diagram are shown in Figures 2B,D.

3.2 The co-expression modules in NAFLD and UC $\,$

We used WGCNA to identify co-expressed gene profiles in the NAFLD (GSE89632) and UC datasets (GSE87466). After removing





abnormal samples, we obtained clustering dendrograms for NAFLD and UC (Figures 3A,B). For the NAFLD dataset, a soft-threshold power β of 16 was chosen based on an R2 greater than 0.9, indicating scale independence (Figure 3C). Similarly, a soft-threshold power β of 18 for the UC dataset was chosen (Figure 3D). Using dynamic hybrid shearing,

nine gene coexpression modules were produced for NAFLD, and 12 gene coexpression modules were produced for UC (Figures 3E,F). The correlation between these gene modules and NAFLD/Normal is depicted in Figure 3G. Notably, The first module, called the green module, consisted of 453 genes. We observed a robust correlation



Frontiers in Medicine

between this module and NAFLD. Likewise, the correlation between these gene modules and UC/Normal (Figure 3H). The second module, known as the turquoise module, contained 1740 genes. We discovered a significant correlation between this module and UC. Therefore, the green and turquoise modules were selected for subsequent analysis. Subsequently, the intersection of these four sets, including the DEGs of NAFLD, the DEGs of UC, the green module, and the turquoise, yielded 34 common genes (Figure 3I). Further research on these specific genes could provide valuable insights into understanding the underlying mechanisms linking these two conditions together.

3.3 Functional correlation analysis

The GSEA analysis results revealed the top five in the NAFLD active UP group. These pathways included Bladder cancer, IL-17 signaling pathway, Osteoclast differentiation, Malaria, and TNF signaling pathway. The active down group included Base excision repair, Lipoic acid metabolism, Mismatch repair, Non-homologous end-joining, and Nucleotide excision repair (Figure 4A). Similarly, in the UC group, the active UP-enriched pathways were found to be the IL-17 signaling pathway, Malaria, Primary immunodeficiency, TNF signaling pathway, and Viral protein interaction with cytokine and cytokine receptor. The active down group included Ascorbate and alternate metabolism, Chemical carcinogenesis-DNA adducts, Drug metabolismcytochrome P450, Pentose and glucuronate interconversions, and Porphyrin metabolism (Figure 4B). Next, With the metascape platform, further GO and KEGG enrichment analysis of critical genes showed their involvement in various BPs such as inflammatory response, chemotaxis, and taxis (Figure 4C). Regarding CCs, essential genes were principally associated with the secretory granule membrane, secretory granule lumen, and cytoplasmic vesicle lumen (Figure 4D). The MFs analysis indicated cytokine activity, receptor-ligand activity, and signaling receptor activator activity (Figure 4E). The essential genes' top three significant KEGG pathways were enriched in the IL-17 signaling pathway, Rheumatoid arthritis, and Chagas disease (Figure 4F).

3.4 PPI network analysis and Core genes selection

A PPI network consisting of 34 protein targets was constructed using the STRING database. They were deleting disconnected nodes in the network. There were 24 nodes and 81 edges in Figure 5A. The cytoHubba plugin in Cytoscape was employed to score the importance of each gene based on their connectivity within the network. After applying this analysis method, the top 10 key genes with the highest node scores were identified, including *IL1B*, *IL6*, *CXCL8 (IL8)*, *CCL2*, *S100A12*, *IL1RN*, *S100A9*, *CCL3*, *CD44*, and *S100A8* (Figure 5B).

3.5 Machine learning for screening core genes and validation

24 core genes were screened using the two machine learning. LASSO logistic regression identified 7 diagnostic core genes of NAFLD in the model. On the other hand, the SVM-RFE method identified a more extensive set of 18 diagnostic core genes for NAFLD (Figures 6A,C,E). LASSO logistic regression and SVM identified 6 and 24 diagnostic core genes, respectively, for UC (Figures 6B,D,F). These two diseases' common diagnostic core genes are considered diagnostic markers for NAFLD with UC (Figure 6G). Finally, one diagnostic marker was obtained: *CCL2*.

In order to evaluate the role of one biomarker in NAFLD and UC diagnosis, the nomogram containing one biomarker (*CCL2*) was generated (Figures 7A,C). The calibration curve generated from this nomogram revealed a slight difference between actual and predicted values. This suggests that the nomogram has a high diagnostic value (Figures 7B,D). Moreover, ROC analysis was used to determine the AUC and 95% CI of the candidate gene in the training and validation sets. The findings were as follows: *CCL2* (AUC: 0.9961, 95% CI: 0.938–1.000) in GSE89632, *CCL2* (AUC: 0.6748, 95% CI: 0.656–0.781) in GSE48452, *CCL2* (AUC: 0.9138, 95% CI: 0.857–0.857) in GSE87466, *CCL2* (AUC: 0.932, 95% CI: 0.857–0.905) in GSE92415



(Figures 7E–H). In the validation set, it was observed that the expression level of *CCL2* was significantly higher than that of the normal group (Figures 7I,J). The above results show that *CCL2* has a high diagnostic value.

3.6 Immune cell correlation analysis

The enrichment analysis results revealed a significant overlap between the genes associated with NAFLD and UC, particularly in terms of their involvement in inflammatory response and immune regulation. Consequently, a more in-depth analysis of immune infiltration is required for both NAFLD and UC. First, Both CIBERSORTx and CIBERSORT algorithms were used to evaluate the proportions of 22 immune cells in GSE89632 (16 normal controls and 16 patients with NAFLD) and GSE87466 (21 normal controls and 21 patients with UC). The proportions of the immune cell composition showed clustering and individual differences (Figures 8A–D). T cells CD4 memory resting, Mast cells restings, $T\gamma\delta$, Macrophagess M1, and Macrophagess M2 were significantly upregulated in NAFLD samples; however, the levels of Plasma cells, Monocytes, and NK cells activated were significantly decreased. In UC samples, T cells CD4 memory activated, T cells follicular helper and Macrophagess M1 were significantly upregulated; however, the levels of Plasma cells, T cells CD8, and Macrophagess M2 were significantly decreased (Figures 8E,F). Additionally, our analysis





identified feature genes closely associated with immune cell infiltration. In NAFLD samples, CCL2 was positively correlated with monocytes and mast cells activated; however, it was negatively correlated with mast cells resting and macrophage M2 (Figure 8G). In UC samples, CCL2 was positively correlated with Neutrophils, and Macrophagess M0 negatively correlated with mast cells resting and Macrophagess M2 (Figure 8H). These genes likely play a critical role in shaping the local microenvironment within both NAFLD and UC-affected tissues.

4 Discussion

The enterohepatic axis is thought to play a vital role in the comorbidities of NAFLD and UC. Recent studies have shown that the interaction between the gut microbiota and the host influences the development and course of both diseases through multiple pathways (22). In addition, metabolic syndrome is commonly seen in NAFLD patients and is characterized by insulin resistance, obesity, etc. These factors may also influence the development of UC by altering microbiome composition (23). Therefore, based on the physiological mechanisms of the hepatoenteric axis of these two diseases, a deeper understanding of these complex interactions through bioinformatics approaches could help to shed light on the pathogenesis of the coexistence of NAFLD and UC and provide potential targets for the development of new therapeutic strategies.

The results of the GO enrichment analysis indicate that key genes are primarily enriched in inflammatory response, chemotaxis, and taxis. This suggests that inflammatory processes play a significant role in the co-pathogenesis of NAFLD and UC. Oxic lipid entities such as palmitic acid, oxLDL, and free cholesterol, which are increased in NAFLD, can activate innate immune pathways (e.g., TLR4, IL-1 β , TNF α , NF κ B) to drive hepatic inflammation (24). Pro-inflammatory cytokines such as TNFa, IL-6, IL-12, and IL-23 play a crucial role in the occurrence and progression of UC. These cytokines are closely associated with the inflammatory response seen in UC patients (25). One important factor contributing to this inflammatory response is the increased permeability of the intestinal barrier. When the intestinal barrier becomes more permeable, immune cells are exposed to higher levels of PAMPs. This exposure triggers an inflammatory response and allows harmful substances, such as bacterial toxins or inflammatory molecules, to enter systemic circulation more easily, which may provide a beneficial way to initiate and further develop NAFLD (25, 26). In the GSEA and KEGG analysis, we find that the IL-17 and TNF signaling pathways have emerged as crucial players in the development and progression of two diseases. One key finding is that IL-17, a pro-inflammatory cytokine, is pivotal in promoting the transition from NAFL to NASH (27). Research has shown that the IL 23/IL17 axis is critically involved in UC pathogenesis. The interaction between these two molecules promotes the differentiation and activation of Th17 cells, which contribute significantly to immune responses and inflammation (28). Additionally, the TNF signaling pathway has been implicated in both NAFLD and UC pathogenesis based on numerous pieces of evidence (29, 30).

Furthermore, we used the Cytoscape plugin cytoHubba to score and screen 34 genes. This analysis identified the top 10 key genes: IL1B, IL6, CXCL8 (IL8), CCL2, S100A12, IL1RN, S100A9, CCL3, CD44, and S100A8. These genes are believed to play crucial roles in NAFLD and UC. IL1B, a pro-inflammatory cytokine, has been found to induce the formation of lipid droplets in hepatocytes and promote the recruitment of neutrophils in the liver. Additionally, IL1B contributes to the progression from liver inflammation to liver fibrosis (31). Patients with UC have increased levels of IL1B mRNA in their intestine tissues. This increase reduces occludin expression by enterocytes and increases TJ permeability. Consequently, it promotes the occurrence and development of UC. Another cytokine, IL-6, has been implicated in liver injury and apoptosis when overexpressed for long periods (32). Inhibiting IL-6 is an effective treatment for UC (33). CCL2 signaling is associated with metabolic disorders during the development of NASH and contributes to lipid accumulation in hepatocytes (34). The CXCL8-CXCR1/2 axis participates in UC pathogenesis through multiple signaling pathways, including PI3k/ Akt, MAPKs, and NF-KB (35). Based on these findings, these genes, IL1B, IL-6, CCL2, and CXCL8, have been shown to contribute not only to NAFLD but also to UC occurrence and development. They hold potential as new treatment targets.

One diagnostic marker was obtained from 24 core genes using two algorithms to investigate the diagnostic markers of NAFLD complicated by UC. Numerous studies have indicated that Chemokines are a group of signaling proteins that play crucial roles in various biological processes involved in the development and progression of NAFLD. These processes include inflammation, immune cell migration, and inflammatory mediator secretion. By regulating chemotaxis, which is the movement of immune cells toward sites of inflammation, chemokines contribute to the overall pathophysiology of NAFLD (36, 37). Conversely, genetic depletion or pharmacological inhibition of CCL2 in mice has been shown to improve steatosis progression, alleviate hepatic inflammatory response, and reduce liver injury (38, 39). Another study has shown that the expression level of CCL2 in the liver positively correlates with the severity of fatty liver disease. Elevated CCL2 levels may promote





Nomogram construction and prediction accuracy evaluation. (A,C) Nomogram for diagnosing NAFLD with UC. (B,D) Calibration curves assessing the predictive accuracy of the nomograms. (E–H) ROC curves for candidate gene (*CCL2*) in the training (GSE89632 and GSE87466) and validation sets (GSE48452 and GSE92415). (I,J) Validation of key genes. *CCL2* were significantly up-regulated in NAFLD (GSE48452) and UC (GSE92415). *p < 0.05; *p < 0.01.

the migration and infiltration of monocytes, thereby exacerbating liver inflammation and fibrosis processes (40–42). Moreover, CCL2 is crucial in recruiting macrophages during the inflammatory process (43). Zheng et al. improve the inflammatory status of the liver by suppressing CCL2 expression (44). Therefore, these findings suggest that the CCL2 gene may serve as a biomarker for NAFLD, offering new avenues for early screening and intervention. Similarly, in the study of the etiology and pathogenesis of UC, the role of the CCL2 gene has attracted significant attention. Previous studies have shown that CCL2 expression levels are associated with various inflammatory diseases, including UC (45). CCL2 is a potent chemotactic cytokine that controls the chemotaxis and infiltration of monocytes/ macrophages in the gut (46). CCL2 enhances the expression and production of other inflammatory cytokines and the infiltration of inflammatory cells and macrophages (47). Previous data suggest that ulcerative colitis is improved by inhibiting the CCL2/NF- κ B/IL-18





pathway, thereby preserving colon function (45). In addition, studies by Yamada et al. (48) have found that palmitic acid, a long-chain fatty acid produced by intestinal flora, can promote the increase of CCL2 secretion, leading to a significant increase in the number of macrophages in the liver and aggravating high-fat diet-induced mouse steatosis. However, this change will be reversed after the removal of intestinal flora. These results suggest that some metabolites of intestinal flora can affect the process of NAFLD through macrophages. Therefore, CCL2 may be an important gene in the comorbidities of NAFLD and UC, mainly promoting the occurrence of the comorbidities of NAFLD and UC by inducing inflammation on the physiological basis of enterohepatic axis.

There are immune cells in the liver that play a crucial role in developing NAFLD (49). On the other hand, UC is an autoimmune disease caused by an overactive immune system (50). To better understand how immune cell infiltration affects NAFLD with UC, we assessed immune infiltration using CIBERSORT. Previous studies have shown that dysregulation of M1/M2 macrophages can lead to chronic inflammation, cancer, and NAFLD (51). Increasing evidence suggests that $\gamma \Delta$ T cells in the liver respond to liver-targeted injury and regulate the progression of liver disease (52). The dysfunction of CD4+ T cells has emerged as a significant pathological factor in the advancement of NAFLD and NASH. Both human and mouse NASH models have revealed peripheral and intrahepatic CD4+ T cell accumulation (53). Additionally, macrophages are essential effector cells of the innate immune system that contribute to intestinal mucosal homeostasis (54). Previous research has indicated potential therapeutic benefits from targeting macrophages in UC treatment (55). CD4+ T cells consist of various functionally diverse subsets such as Th1, Th2, Th17, and Treg cells. All these subsets have been implicated in initiating or propagating intestinal inflammation in UC patients (56).

The correlation between the six hub genes and immune cells revealed consistent expression of these genes with multiple immune cells. For instance, all six hub genes were overexpressed in NAFLD patients compared to HCs. Additionally, the expression of these genes was positively correlated with an increase in resting mast cells, which were found to be elevated in NAFLD patients compared to HCs. Based on this information, the high expression of these hub genes may contribute to the increase in multiple immune cells associated with NAFLD. Furthermore, these genes may play a role in shaping the characteristic immune microenvironment of NAFLD. Our research suggests that the CCL2 gene may be an essential core gene among these six identified genes. CCL2 is a chemokine that attracts inflammatory monocytes to stressed or injured tissues. Infiltrating inflammatory monocytes and upregulation of CCL2 have been strongly implicated in liver disease pathogenesis based on animal models (57, 58). Similarly, studies have shown that macrophages can secrete chemokines and mediate immune cell recruitment during the inflammatory cascade process (59). The FBXW7/EZH2/CCL2/CCL7 pathway has been shown to exacerbate colitis severity by recruiting CX3CR1 proinflammatory MPhs (60).

In summary, in this study, we explored the potential mechanisms of comorbidities between NAFLD and UC based on bioinformatics analysis. The results show that inflammation plays a vital role in the co-pathogenesis of NAFLD and UC, primarily through inflammatory response, immune regulation, and other pathways, affecting both pathological processes. We found that a critical diagnostic gene, CCL2, is abnormally expressed in both NAFLD and UC, suggesting that this molecule may be a common pathogenic factor for both diseases. Through in-depth analysis of related genes and signaling pathways, we propose that the dual role of inflammatory factors such as CCL2 in NAFLD and UC may provide new targets for future diagnosis and treatment. However, there are some limitations. First, the study data came from an existing database, so the authenticity and completeness of the results depend on the data. Second, the results do not reflect all actual cellular network characteristics in living organisms, and our findings still need to be validated *in vivo* and *in vitro* to guide clinical practice better.

5 Conclusion

In this study, we have identified 34 key genes that are associated with the comorbidity between NAFLD and UC, which were mainly involved in the IL-17 signaling pathway. 10 hub genes (including *IL1B*, *IL6*, *CXCL8* (*IL8*), *CCL2*, *S100A12*, *IL1RN*, *S100A9*, *CCL3*, *CD44*, and *S100A8*) may have a significant impact on the pathophysiological mechanisms of these two diseases. The machine learning analysis also revealed a single feature gene *CCL2* that could potentially serve as a diagnostic biomarker for NAFLD and UC. This diagnostic marker not only affects immune cells, but also has the potential to become a therapeutic target when NAFLD coexists with UC.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: GSE89632, GSE87466, GSE48452, and GSE92415.

Author contributions

ZL: Writing – original draft. CH: Writing – original draft. JC: Software, Conceptualization, Project administration, Resources, Writing – original draft. YC: Visualization, Software, Writing – review & editing. HY: Methodology, Software, Investigation, Project administration, Validation, Writing – review & editing. QW: Writing

References

1. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*. (2016) 64:73–84. doi: 10.1002/hep.28431

2. Powell EE, Wong VW, Rinella M. Non-alcoholic fatty liver disease. *Lancet.* (2021) 397:2212-24. doi: 10.1016/S0140-6736(20)32511-3

3. Adams LA, Anstee QM, Tilg H, Targher G. Non-alcoholic fatty liver disease and its relationship with cardiovascular disease and other extrahepatic diseases. *Gut.* (2017) 66:1138–53. doi: 10.1136/gutjnl-2017-313884

4. Ordás I, Eckmann L, Talamini M, Baumgart DC, Sandborn WJ. Ulcerative colitis. Lancet. (2012) 380:1606–19. doi: 10.1016/S0140-6736(12)60150-0

5. Le Berre C, Honap S, Peyrin-Biroulet L. Ulcerative colitis. *Lancet.* (2023) 402:571–84. doi: 10.1016/S0140-6736(23)00966-2

6. Halpin SJ, Ford AC. Prevalence of symptoms meeting criteria for irritable bowel syndrome in inflammatory bowel disease: systematic review and meta-analysis. *Am J Gastroenterol.* (2012) 107:1474–82. doi: 10.1038/ajg.2012.260

7. Ford AC. Overlap between irritable bowel syndrome and inflammatory bowel disease. *Gastroenterol Hepatol.* (2020) 16:211–3.

 Szałwińska P, Włodarczyk J, Spinelli A, Fichna J, Włodarczyk M. IBS-symptoms in IBD patients-manifestation of concomitant or different entities. J Clin Med. (2020) 10:31. doi: 10.3390/jcm10010031

9. Wu S, Yuan C, Yang Z, Liu S, Zhang Q, Zhang S, et al. Non-alcoholic fatty liver is associated with increased risk of irritable bowel syndrome: a prospective cohort study. *BMC Med.* (2022) 20:262. doi: 10.1186/s12916-022-02460-8

– review & editing. FL: Writing – review & editing. TZ: Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Key Research and Development Program of Science and Technology Department of Sichuan Province (grant number 2023YFS0333) and the Innovative Research Team Project on "kidney brain related of Zang in TCM in biology" from Chengdu University of TCM (2023).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1323859/ full#supplementary-material

10. Papaefthymiou A, Potamianos S, Goulas A, Doulberis M, Kountouras J, Polyzos SA. Inflammatory bowel disease-associated fatty liver disease: the potential effect of biologic agents. *J Crohns Colitis.* (2022) 16:852–62. doi: 10.1093/ecco-jcc/jjab212

11. Albillos A, de Gottardi A, Rescigno M. The gut-liver axis in liver disease: pathophysiological basis for therapy. *J Hepatol.* (2020) 72:558–77. doi: 10.1016/j. jhep.2019.10.003

12. Benedé-Ubieto R, Cubero FJ, Nevzorova YA. Breaking the barriers: the role of gut homeostasis in metabolic-associated Steatotic liver disease (MASLD). *Gut Microbes*. (2024) 16:2331460. doi: 10.1080/19490976.2024.2331460

13. Amerikanou C, Papada E, Gioxari A, Smyrnioudis I, Kleftaki SA, Valsamidou E, et al. Mastiha has efficacy in immune-mediated inflammatory diseases through a microRNA-155 Th17 dependent action. *Pharmacol Res.* (2021) 171:105753. doi: 10.1016/j.phrs.2021.105753

14. Hu Y, Zeng N, Ge Y, Wang D, Qin X, Zhang W, et al. Identification of the shared gene signatures and biological mechanism in type 2 diabetes and pancreatic Cancer. *Front Endocrinol.* (2022) 13:847760. doi: 10.3389/fendo.2022.847760

15. Clough E, Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Res.* (2024) 52:D138–d144. doi: 10.1093/nar/gkad965

16. Arendt BM, Comelli EM, Ma DWL, Lou W, Teterina A, Kim TH, et al. Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. *Hepatology*. (2015) 61:1565–78. doi: 10.1002/hep.27695

17. Li K, Strauss R, Ouahed J, Chan D, Telesco SE, Shouval DS, et al. Molecular comparison of adult and pediatric ulcerative colitis indicates broad similarity of molecular pathways in disease tissue. *J Pediatr Gastroenterol Nutr.* (2018) 67:45–52. doi: 10.1097/MPG.00000000001898

18. Ahrens M, Ammerpohl O, von Schönfels W, Kolarova J, Bens S, Itzel T, et al. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab.* (2013) 18:296–302. doi: 10.1016/j.cmet.2013.07.004

19. Sandborn WJ, Feagan BG, Marano C, Zhang H, Strauss R, Johanns J, et al. Subcutaneous golimumab induces clinical response and remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology*. (2014) 146:85–95. doi: 10.1053/j. gastro.2013.05.048

20. Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics*. (2014) 15:293. doi: 10.1186/1471-2105-15-293

21. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* (2019) 10:1523. doi: 10.1038/s41467-019-09234-6

22. Lau E, Carvalho D, Freitas P. Gut microbiota: association with NAFLD and metabolic disturbances. *Biomed Res Int.* (2015) 2015:979515. doi: 10.1155/2015/979515

23. Reccia I, Kumar J, Akladios C, Virdis F, Pai M, Habib N, et al. Non-alcoholic fatty liver disease: a sign of systemic disease. *Metabolism*. (2017) 72:94–108. doi: 10.1016/j. metabol.2017.04.011

24. McGettigan B, McMahan R, Orlicky D, Burchill M, Danhorn T, Francis P, et al. Dietary lipids differentially shape nonalcoholic steatohepatitis progression and the transcriptome of Kupffer cells and infiltrating macrophages. *Hepatology.* (2019) 70:67–83. doi: 10.1002/hep.30401

25. Samoilă I, Dinescu S, Costache M. Interplay between cellular and molecular mechanisms underlying inflammatory bowel diseases development-a focus on ulcerative colitis. *Cells.* (2020) 9:1647. doi: 10.3390/cells9071647

26. Santana PT, Rosas SLB, Ribeiro BE, Marinho Y, de Souza HSP. Dysbiosis in inflammatory bowel disease: pathogenic role and potential therapeutic targets. *Int J Mol Sci.* (2022) 23:3464. doi: 10.3390/ijms23073464

27. Chackelevicius CM, Gambaro SE, Tiribelli C, Rosso N. Th17 involvement in nonalcoholic fatty liver disease progression to non-alcoholic steatohepatitis. *World J Gastroenterol.* (2016) 22:9096–103. doi: 10.3748/wjg.v22.i41.9096

28. Noviello D, Mager R, Roda G, Borroni RG, Fiorino G, Vetrano S. The IL23-IL17 immune Axis in the treatment of ulcerative colitis: successes, defeats, and ongoing challenges. *Front Immunol.* (2021) 12:611256. doi: 10.3389/fimmu.2021.611256

29. Ghorbaninejad M, Heydari R, Mohammadi P, Shahrokh S, Haghazali M, Khanabadi B, et al. Contribution of NOTCH signaling pathway along with TNF- α in the intestinal inflammation of ulcerative colitis. *Gastroenterol Hepatol Bed Bench.* (2019) 12:S80–s86.

30. Luo Y, Yang P, Li Z, Luo Y, Shen J, Li R, et al. Liraglutide improves non-alcoholic fatty liver disease in diabetic mice by modulating inflammatory signaling pathways. *Drug Des Devel Ther.* (2019) 13:4065–74. doi: 10.2147/DDDT.S224688

31. Mirea AM, Tack CJ, Chavakis T, Joosten LAB, Toonen EJM. IL-1 family cytokine pathways underlying NAFLD: towards new treatment strategies. *Trends Mol Med.* (2018) 24:458–71. doi: 10.1016/j.molmed.2018.03.005

32. Dogru T, Ercin CN, Erdem G, Sonmez A, Tapan S, Tasci I. Increased hepatic and circulating interleukin-6 levels in human nonalcoholic steatohepatitis. *Am J Gastroenterol.* (2008) 103:3217–8. doi: 10.1111/j.1572-0241.2008.02161_17.x

33. Schreiber S, Aden K, Bernardes JP, Conrad C, Tran F, Höper H, et al. Therapeutic Interleukin-6 trans-signaling inhibition by Olamkicept (sgp130Fc) in patients with active inflammatory bowel disease. *Gastroenterology.* (2021) 160:2354–2366.e11. doi: 10.1053/j.gastro.2021.02.062

34. Holt AP, Haughton EL, Lalor PF, Filer A, Buckley CD, Adams DH. Liver myofibroblasts regulate infiltration and positioning of lymphocytes in human liver. *Gastroenterology.* (2009) 136:705–14. doi: 10.1053/j.gastro.2008.10.020

35. Zhu Y, Yang S, Zhao N, Liu C, Zhang F, Guo Y, et al. CXCL8 chemokine in ulcerative colitis. *Biomed Pharmacother*. (2021) 138:111427. doi: 10.1016/j. biopha.2021.111427

36. Roh YS, Seki E. Chemokines and chemokine receptors in the development of NAFLD. Adv Exp Med Biol. (2018) 1061:45–53. doi: 10.1007/978-981-10-8684-7_4

37. McMahan RH, Porsche CE, Edwards MG, Rosen HR. Correction: free fatty acids differentially downregulate chemokines in liver sinusoidal endothelial cells: insights into non-alcoholic fatty liver disease. *PLoS One.* (2016) 11:e0168301. doi: 10.1371/journal. pone.0168301

38. Kanda H, Tateya S, Tamori Y, Kotani K, Hiasa K, Kitazawa R, et al. MCP-1 contributes to macrophage infiltration into adipose tissue, insulin resistance, and hepatic steatosis in obesity. *J Clin Invest.* (2006) 116:1494–505. doi: 10.1172/JCI26498

39. Baeck C, Wehr A, Karlmark KR, Heymann F, Vucur M, Gassler N, et al. Pharmacological inhibition of the chemokine CCL2 (MCP-1) diminishes liver macrophage infiltration and steatohepatitis in chronic hepatic injury. *Gut.* (2012) 61:416–26. doi: 10.1136/gutjnl-2011-300304

40. Thibaut R, Gage MC, Pineda-Torra I, Chabrier G, Venteclef N, Alzaid F. Liver macrophages and inflammation in physiology and physiopathology of non-alcoholic fatty liver disease. *FEBS J.* (2022) 289:3024–57. doi: 10.1111/febs.15877

41. Krenkel O, Puengel T, Govaere O, Abdallah AT, Mossanen JC, Kohlhepp M, et al. Therapeutic inhibition of inflammatory monocyte recruitment reduces steatohepatitis and liver fibrosis. *Hepatology*. (2018) 67:1270–83. doi: 10.1002/hep.29544

42. Kang J, Postigo-Fernandez J, Kim KJ, Zhu C, Yu J, Meroni M, et al. Notch-mediated hepatocyte MCP-1 secretion causes liver fibrosis. *JCI Insight*. (2023) 8:e165369. doi: 10.1172/jci.insight.165369

43. Ryu J, Hadley JT, Li Z, Dong F, Xu H, Xin X, et al. Adiponectin alleviates dietinduced inflammation in the liver by suppressing MCP-1 expression and macrophage infiltration. *Diabetes*. (2021) 70:1303–16. doi: 10.2337/db20-1073

44. Zheng Y, Ying H, Shi J, Li L, Zhao Y. Alanyl-glutamine dipeptide attenuates nonalcoholic fatty liver disease induced by a high-fat diet in mice by improving gut microbiota Dysbiosis. *Nutrients*. (2023) 15:3988. doi: 10.3390/nu15183988

45. ElMahdy MK, Antar SA, Elmahallawy EK, Abdo W, Hijazy HHA, Albrakati A, et al. A novel role of Dapagliflozin in mitigation of acetic acid-induced ulcerative colitis by modulation of monocyte chemoattractant protein 1 (MCP-1)/nuclear factor-kappa B (NF-κB)/Interleukin-18 (IL-18). *Biomedicines*. (2021) 10:40. doi: 10.3390/biomedicines10010040

46. Lee SH, Kwon JE, Cho ML. Immunological pathogenesis of inflammatory bowel disease. *Intest Res.* (2018) 16:26-42. doi: 10.5217/ir.2018.16.1.26

47. Deshmane SL, Kremlev S, Amini S, Sawaya BE. Monocyte chemoattractant protein-1 (MCP-1): an overview. *J Interf Cytokine Res.* (2009) 29:313–26. doi: 10.1089/ jir.2008.0027

48. Yamada S, Kamada N, Amiya T, Nakamoto N, Nakaoka T, Kimura M, et al. Gut microbiota-mediated generation of saturated fatty acids elicits inflammation in the liver in murine high-fat diet-induced steatohepatitis. *BMC Gastroenterol.* (2017) 17:136. doi: 10.1186/s12876-017-0689-3

49. Barreby E, Chen P, Aouadi M. Macrophage functional diversity in NAFLD - more than inflammation. *Nat Rev Endocrinol.* (2022) 18:461–72. doi: 10.1038/ s41574-022-00675-6

50. Magro F, Gionchetti P, Eliakim R, Ardizzone S, Armuzzi A, Barreiro-de Acosta M, et al. Third European evidence-based consensus on diagnosis and Management of Ulcerative Colitis. Part 1: definitions, diagnosis, extra-intestinal manifestations, pregnancy, cancer surveillance, surgery, and Ileo-anal pouch disorders. *J Crohns Colitis*. (2017) 11:649–70. doi: 10.1093/ecco-jcc/jjx008

51. Sica A, Mantovani A. Macrophage plasticity and polarization: *in vivo* veritas. *J Clin Invest.* (2012) 122:787–95. doi: 10.1172/JCI59643

52. Rajoriya N, Fergusson JR, Leithead JA, Klenerman P. Gamma Delta T-lymphocytes in hepatitis C and chronic liver disease. *Front Immunol.* (2014) 5:400. doi: 10.3389/fimmu.2014.00400

53. Sutti S, Albano E. Adaptive immunity: an emerging player in the progression of NAFLD. *Nat Rev Gastroenterol Hepatol.* (2020) 17:81–92. doi: 10.1038/ s41575-019-0210-2

54. Bain CC, Mowat AM. The monocyte-macrophage axis in the intestine. Cell Immunol. (2014) 291:41-8. doi: 10.1016/j.cellimm.2014.03.012

55. Bloemendaal FM, Koelink PJ, van Schie K, Rispens T, Peters CP, Buskens CJ, et al. TNF-anti-TNF immune complexes inhibit IL-12/IL-23 secretion by inflammatory macrophages via an fc-dependent mechanism. *J Crohns Colitis*. (2018) 12:1122–30. doi: 10.1093/ecco-jcc/jjy075

56. Choy MC, Visvanathan K, De Cruz P. An overview of the innate and adaptive immune system in inflammatory bowel disease. *Inflamm Bowel Dis.* (2017) 23:2–13. doi: 10.1097/MIB.00000000000955

57. Liu B, Xiang L, Ji J, Liu W, Chen Y, Xia M, et al. Sparcl1 promotes nonalcoholic steatohepatitis progression in mice through upregulation of CCL2. *J Clin Invest*. (2021) 131:e144801. doi: 10.1172/JCI144801

58. Ferrari-Cestari M, Okano S, Patel PJ, Horsfall LU, Keshvari S, Hume DA, et al. Serum CC-chemokine ligand 2 is associated with visceral adiposity but not fibrosis in patients with non-alcoholic fatty liver disease. *Dig Dis.* (2023) 41:439–46. doi: 10.1159/000527784

59. Asano K, Takahashi N, Ushiki M, Monya M, Aihara F, Kuboki E, et al. Intestinal CD169(+) macrophages initiate mucosal inflammation by secreting CCL8 that recruits inflammatory monocytes. *Nat Commun.* (2015) 6:7802. doi: 10.1038/ncomms8802

60. He J, Song Y, Li G, Xiao P, Liu Y, Xue Y, et al. Fbxw7 increases CCL2/7 in CX3CR1hi macrophages to promote intestinal inflammation. *J Clin Invest.* (2019) 129:3877–93. doi: 10.1172/JCI123374