Check for updates

Expression of immune related genes and possible regulatory mechanisms in different stages of non-alcoholic fatty liver disease

Risheng He[†], Canghai Guan[†], Xudong Zhao, Liang Yu* and Yunfu Cui*

Department of Pancreatobiliary Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, China

Background: Non-alcoholic fatty liver disease (NAFLD), which includes simple steatosis (SS) and non-alcoholic steatohepatitis (NASH), is a significant contributor to liver disease on a global scale. The change of immunity-related genes (IRGs) expression level leads to different immune infiltrations. However, the expression of IRGs and possible regulatory mechanisms involved in NAFLD remain unclear. The objective of our research is to investigate crucial genes linked to the development of NAFLD and the transition from SS to NASH.

Methods: Dataset GSE89632, which includes healthy controls, SS patients, and NASH patients, was obtained using the GEO database. To examine the correlation between sets of genes and clinical characteristics, we employed weighted gene co-expression network analysis (WGCNA) and differential expression analysis. Hub genes were extracted using a network of protein-protein interactions (PPI) and three different machine learning algorithms. To validate the findings, another dataset that is publicly accessible and mice that were subjected to a high-fat diet (HFD) or MCD diet were utilized. Furthermore, the ESTIMATE algorithm and ssGSEA were employed to investigate the immune landscape in the normal versus SS group and SS versus NASH group, additionally, the relationship between immune infiltration and the expression of hub genes was also examined.

Results: A total of 28 immune related key genes were selected. Most of these genes expressed reverse patterns in the initial and progressive stages of NAFLD. GO and KEGG analyses showed that they were focused on the cytokine related pathways and immune cell activation and chemotaxis. After screening by various algorithms, we obtained two hub genes, including JUN and CCL20. Validation of these findings was confirmed by analyzing gene expression patterns in both the validation dataset and the mouse model. Ultimately, two hub genes were discovered to have a significant correlation with the infiltration of immune cells.

Conclusion: We proposed that there were dynamic changes in the expression levels of IRGs in different stages of NAFLD disease, which led to different immune landscapes in SS and NASH. The findings of our research could serve as a guide for the accurate management of various phases of NAFLD.

1 Introduction

Over the past few decades, there has been a notable rise in the occurrence of metabolic disorders, such as obesity. It is crucial to note that obesity not only has its own associated comorbidities but also adversely affects overall health and increases susceptibility to other conditions (1, 2). As a result, people who are obese have an increased likelihood of experiencing comorbidities like insulin resistance, type 2 diabetes, high blood pressure, abnormal lipid levels, heart disease, and fatty liver disease. Non-alcoholic fatty liver disease (NAFLD) is acknowledged as the liver-related aspect of the metabolic syndrome (3). NAFLD is a widespread chronic liver disease that includes different conditions like simple steatosis (SS) and non-alcoholic steatohepatitis (NASH), which is characterized by inflammation (4). Projections indicate a significant rise in the prevalence of NAFLD, with the Chinese population estimated to exceed 300 million cases by 2030, over 100 million cases in the United States, and 15-20 million cases in major European countries (5). Hence, the escalating global incidence of NAFLD warrants considerable clinical scrutiny. Notably, the clinical presentation of NAFLD is inconspicuous during the stage of SS, only becoming apparent in the subsequent stage of NASH (6). As the disease advances, the liver undergoes significant pathological alterations, characterized by lobular inflammation and hepatocyte ballooning, accompanied by fibrosis in some cases (7). Failure to effectively manage the disease may lead to its progression to cirrhosis and, ultimately, liver cancer (8). Despite extensive research on this disease, its pathogenesis and the mechanisms underlying the progression from SS to NASH are still poorly understood.

In spite of the fact that it is primarily a metabolic disorder, inflammatory processes mediated by immune cells are involved in NAFLD, and inflammation is especially important at the stage of NASH, when it becomes integral to disease progression (9, 10). Increasing amounts of research validate the crucial involvement of the immune system in the different phases of NAFLD advancement, exhibiting alterations in immune cell infiltration levels and cytokine levels within the liver microenvironment throughout the progression of the disease (11, 12). In this study, we performed an extensive bioinformatics analysis comparing normal liver tissues with SS tissues, as well as SS tissues with NASH tissues. Based on comprehensive bioinformatics analyses, we have identified the immune status of the liver at different stages of the disease, as well as the hub genes and the mechanisms that regulate the immune microenvironment.

2 Materials and methods

2.1 Data acquisition and preliminary processing

The mRNA sequencing dataset GSE89632 and GSE135251, which were obtained from the GEO database in NCBI (http://www.ncbi.nlm.nih.gov/geo/), included expression profiles that were generated using the platforms GPL14951 and GPL18573, respectively. These data sets contained normal samples, SS samples and NASH samples. And log2 was performed to process raw counts. To further investigate, we obtained immunity-related genes (IRGs) from ImmPort, a curated immune database used for the management and identification of genes related to immunity (13). Figure 1 depicted the main search process of the article.

2.2 Weighted gene co-expression network analysis and module gene selection in normal vs. SS patients

The WGCNA method was used to construct gene coexpression networks and identify functional modules (14).The cut height was set to 120 and the goodSamplesGenes function was used to filter outlier samples, resulting in the construction of a scale-free coexpression network. The most highly expressed 10000 genes were selected for the following analysis. The pickSoftThreshold function determined an appropriate 'soft' threshold power (β) for calculating intergenic adjacency. Weight coexpression network used the blockwiseModules function. The function plotDendroAndColors visualized the clustering among samples. The labeledHeatmap function showed the correlation between the
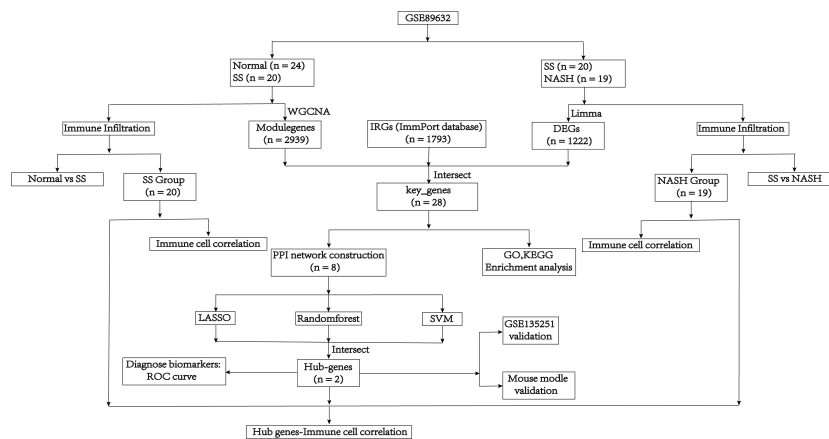
FIGURE 1
Flow diagram of the analysis process.

group and gene Modules. The function plotEigengeneNetworks displayed the association among different gene Modules. Finally, gene significance (GS) and module membership (MM) correlations were calculated, and the corresponding module gene information was extracted for further analysis.

2.3 Identification of differentially expressed genes between SS samples and NASH samples

Differentially expressed genes (DEGs) were selected from the SS vs. NASH group using the R limma package (15), based on the threshold criteria of |log2FC| > 0.25 and p-value < 0.05. Using the R software, the volcano plot for the DEGs and the expression heatmap for the top 15 up and top 15 down genes were created using the 'ggplot2' and 'pheatmap' packages, respectively. The draw_venn method in the tinyarray R package was used to display the key genes, which were obtained by intersecting the DEGs, modulegenes, and IRGs. The differential expression of these genes in normal vs. SS group and SS vs. NASH group was visualized using the ggplot2 and ComplexHeatmap packages. Furthermore, the corrplot library was employed to examine the associations among these genes.

2.4 Enrichment analysis

To perform enrichment analysis on important genes, we utilized the enrichGO and enrichKEGG functions from the R package clusterProfiler. The enrichGO function was employed for Gene Ontology (GO) analysis, while the enrichKEGG function was used for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis (16). A p-value cutoff of 0.05 was applied. The R package ggplot2 and stringr were used for visualization.

2.5 Protein–protein interaction network construction

Using the online tool STRING (https://string-db.org/), a PPI network was built by considering all essential genes and applying a filter condition (combined score > 0.4). Afterwards, we obtained the interaction data and enhanced the PPI network using Cytoscape software to improve its visual representation (17, 18). The identification of significant gene clusters and the acquisition of cluster scores were achieved using Minimal Common Oncology Data Elements (MCODE), with filter criteria including a degree cut-off of 2, a node score cut-off of 0.2, a k-core of 2, and a maximum depth of 100. The cytoHubba plugins utilized the Maximal Clique Centrality (MCC) algorithm to determine the significance of genes within the primary gene cluster.

2.6 Screening of hub genes

Three machine learning algorithms, containing lasso regression (19), random forest (RF) (20) and SVM (21), were used to screen the hub genes from the most significant gene cluster screened by PPI. We used Lasso Cox regression to detect changes in regression coefficients of the important genes. The optimal parameter λ was determined using 10-fold cross-validation with the R package glmnet. Finally, we selected genes based on lambda.min. And we used the R package plotmo to visualize the coefficient contraction of LASSO Cox regression. The RF algorithm utilized the R package random Forest to assess the significance of genes in the most prominent gene cluster identified by PPI. The top five genes were selected. Support Vector Machines (SVM) is a supervised machine learning method used for regression or classification tasks, which necessitates a labeled training dataset SVM-RFE, a method in machine learning, trained a subset of characteristics from various groups to reduce the feature set and identify the most influential

features. In the end, we identified the hub genes by intersecting the genes screened using the three algorithms.

2.7 Verification of hub genes expression and diagnostic efficacy

The expression level of the central genes was determined using the GSE89632 dataset, and the diagnostic value of these genes in distinguishing between the normal and SS groups, as well as the SS and NASH groups, was evaluated by constructing ROC curves using the pROC package. And the above results were verified with the GSE135251 dataset.

2.8 Immune infiltration analysis and correlation analysis

The ESTIMATE algorithm was utilized to deduce the immune cell infiltration by analyzing the transcriptome data and the immune microenvironment scores (22). The scores comprised of the immune score, the stromal score, and the estimate score. To better recognize the immune cell characteristics in tissue of normal vs. SS group and SS vs. NASH group, we compared the differences of immune cell subsets in the samples. The ssGSEA method was employed to compare the distinct composition of 28 immune cells. The means in the normal vs. SS group and SS vs. NASH group were compared using the t_test function from the R package rstatix. Subsequently, the results were visualized using the ggboxplot function from the R package ggpubr (23, 24). The correlation between the distribution of immune cells in SS and NASH patients was uncovered using the corrplot R package. Furthermore, we examined the relationship between the expression of hub genes and the infiltration of immune cells in SS and NASH patients.

2.9 Construction of SS and NASH model mice

A total of 18 male C57/BL6 mice, aged 6-8 weeks, were randomly assigned to three groups. The first group received a high-fat diet (HFD) consisting of 60% fat (n = 6), the second group was fed on an MCD diet (n = 6), and the third group was given a control diet (CD) with 10% fat (n = 6) for a duration of 8 weeks. At the Second Affiliated Hospital of Harbin Medical University, the animals were kept in a controlled environment where the temperature was regulated. They were provided with food

and water freely and followed a 12-hour cycle of light and darkness. The animal experiments received ethical clearance from the Ethics Committee of Second Affiliated Hospital of Harbin Medical University (SYDW2023-077).

2.10 Histological analysis

Liver tissues were isolated from mice and immediately fixed with 4% formalin (Sigma- Aldrich, St. Louis, MO). Afterward, the dehydrated samples were embedded in paraffin. Histological changes were examined by H&E staining. Images were acquired using the Eclipse E100 microscope (Nikon, Japan). Mice models were evaluated by two qualified pathologists.

2.11 Quantitative RT-PCR analysis

Total RNA was extracted from liver tissues using Trizol reagent (Invitrogen), and then reverse transcribed into cDNA using the Transcriptor First Strand cDNA Synthesis Kit (Roche, Penzberg, Germany). FastStart Universal SYBR Green Master (Roche) was used to amplify each sample in a reaction mixture of 20 μl. The fold changes were converted using the 2-ΔΔCt technique. Expression levels were determined by calculating and normalizing them to the endogenous GAPDH. The primer sequences were shown in Table 1.

2.12 Statistical analysis

The findings were presented as average ± standard deviation from a minimum of 3 separate trials. The differences between groups were compared using t-test, and GraphPad Prism 8.0 and R software version 4.2.3 were utilized for data analyses. Statistical significance was determined by comparing p-values, where $p < 0.05$ denoted significance (*$p < 0.05$, **$p < 0.01$, ***$p<0.001$, ****$p<0.0001$).

3 Result

3.1 Weighted gene co-expression network construction

We selected data of normal and SS patients in the dataset to identify regulatory genes related to the occurrence of SS. Correlation networks were used for identifying clusters of highly correlated genes across microarray samples. We employed WGCNA to construct and analyzed active SS-associated networks. After clustering the samples,

TABLE 1   The sequences of primers.

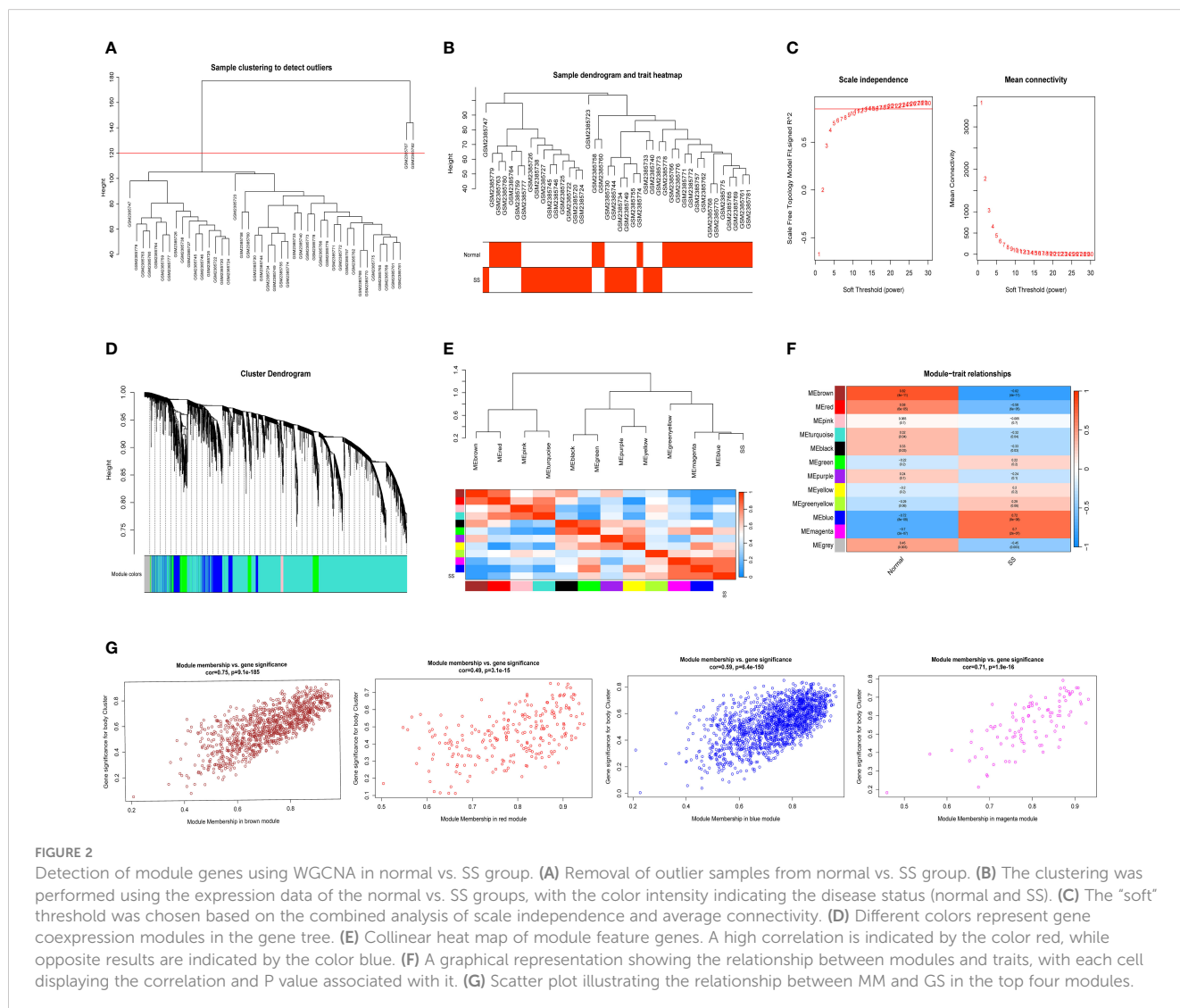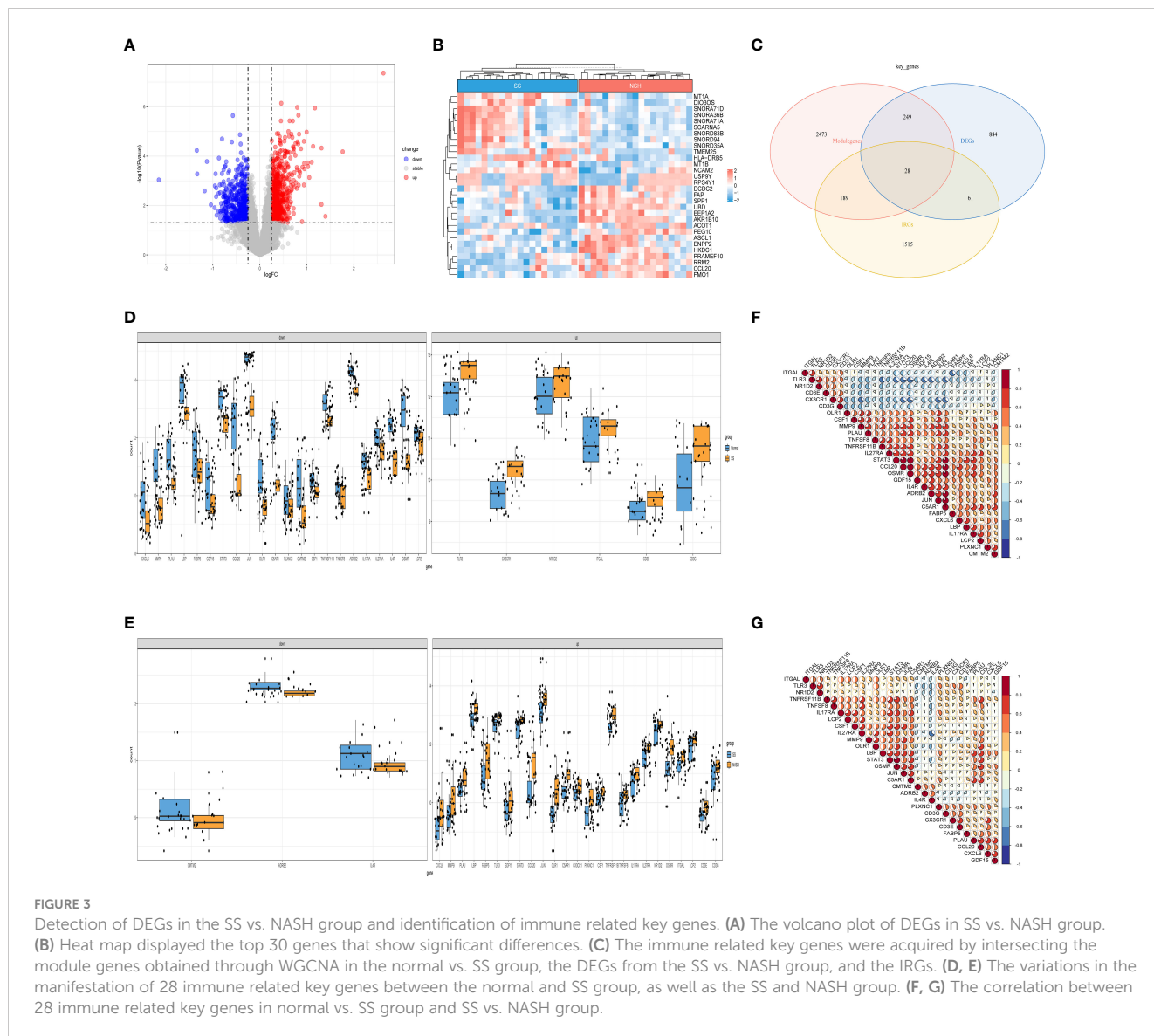| Genes	Forward Primer (5′-3′)	Reverse Primer (5′-3′)
GAPDH	GTGCCGCCTGGAGAAAC	AAGGTGGAAGAGTGGGAGT
Jun	GGGAGCATTTGGAGAGTCCC	TTTGCAAAAGTTCGCTCCCG
Ccl20	CCAGGCAGAAGCAAGCAAC	TTTGGATCAGCGCACACAGA

we established a suitable threshold (cutHeight = 120) to eliminate the clearly abnormal samples (Figure 2A). After removing outliers, we drew a sample clustering tree (Figure 2B). Here, we selected top 10000 genes for subsequent analysis. Using a soft-threshold power of β = 14 (R2 = 0.85), the adjacency matrix was created, ensuring gene distribution adhered to a scale-free network (Figure 2C). This retained valuable connectivity information. A total of 11 modules were generated and identified under the parameter settings of minModuleSize = 30 and mergeCutHeight = 0.25 (Figure 2D). The connectivity was calculated among the modules and we added group information to them, then we performed the cluster analysis, meanwhile, the heat map of their correlation was also plotted (Figure 2E). In order to further examine the connection between the models and phenotype, we computed the correlation coefficients of each model with the SS trait. The findings indicated that SS had a statistically significant correlation with 6 modules. Among them, SS is highly associated with four modules: 'brown' (r = -0.82, p = 4e−11), 'red' (r = -0.58, p = 6e−05), 'blue' (r = 0.72, p = 8e−08), and 'magenta' (r = 0.7, p = 2e−07) (Figure 2F). Next, we conducted an analysis of module membership (MM) and gene significance (GS) correlation

for these 4 modules. Interestingly, we found a strong positive correlation between MM and GS in these 4 modules (Figure 2G). After combining the genes of these four modules, we ultimately acquired a sum of 2939 genes within the modules.

3.2 Identification of immune related key genes involved in the onset and progression of NAFLD

Next, we selected data of SS and NASH patients in the dataset to identify regulatory genes related to the development of SS. DEGs were identified by converting the fold changes (FC) of gene expression to log2 values and applying the cutoff criteria of |log2FC| ≥ 0.25 and p-value < 0.05. Based on these criteria, a total of 1222 genes were identified as DEGs that potentially contribute to the progression of SS. Among these genes, 694 were found to be up-regulated while 528 were down-regulated. Volcano plots of DEGs were displayed in Figure 3A, while Figure 3B presented a heat map showcasing the top 30 genes. By intersecting the DEGs, module

FIGURE 2
Detection of module genes using WGCNA in normal vs. SS group. **(A)** Removal of outlier samples from normal vs. SS group. **(B)** The clustering was performed using the expression data of the normal vs. SS groups, with the color intensity indicating the disease status (normal and SS). **(C)** The "soft" threshold was chosen based on the combined analysis of scale independence and average connectivity. **(D)** Different colors represent gene coexpression modules in the gene tree. **(E)** Collinear heat map of module feature genes. A high correlation is indicated by the color red, while opposite results are indicated by the color blue. **(F)** A graphical representation showing the relationship between modules and traits, with each cell displaying the correlation and P value associated with it. **(G)** Scatter plot illustrating the relationship between MM and GS in the top four modules.

FIGURE 3

Detection of DEGs in the SS vs. NASH group and identification of immune related key genes. **(A)** The volcano plot of DEGs in SS vs. NASH group. **(B)** Heat map displayed the top 30 genes that show significant differences. **(C)** The immune related key genes were acquired by intersecting the module genes obtained through WGCNA in the normal vs. SS group, the DEGs from the SS vs. NASH group, and the IRGs. **(D, E)** The variations in the manifestation of 28 immune related key genes between the normal and SS group, as well as the SS and NASH group. **(F, G)** The correlation between 28 immune related key genes in normal vs. SS group and SS vs. NASH group.

genes, and IRGs, we identified a total of 28 genes that were present in all three gene clusters during the onset and progression of NAFLD (Figure 3C). Box plots displayed the variations in gene expression between the normal vs. SS group, as well as the SS vs. NASH group. We could observe that 6 genes were upregulated in SS patients, while 22 genes were upregulated in normal samples (Figure 3D). By contrast, most of these 28 genes were highly expressed in NASH patients compared to patients with SS (Figure 3E). In addition, we also analyzed the correlation between these genes, and the results were shown in Figures 3F, G, which suggested that most of the genes might interact with each other and participate in the same pathway.

3.3 Enrichment analyses of 28 immune related key genes

In order to explore the biological functions and pathways of these 28 immune related key genes, GO and KEGG enrichment analyses were performed. We obtained a total of 298 biological processes that were significantly related, along with 23 KEGG signaling pathways. To uncover the biological functions of immune-related key genes, a GO analysis was conducted (Figure 4A). As observed, the majority of genes in the GO category were primarily involved in functions such as 'leukocyte migration', 'myeloid leukocyte activation', 'cytokine-mediated signaling pathway', 'macrophage activation', and 'granulocyte chemotaxis' (BP); 'external side of plasma membrane', 'secretory granule membrane', 'plasma membrane signaling receptor complex', 'specific granule membrane', and 'alpha-beta T cell receptor complex' (CC); 'cytokine activity', 'cytokine receptor binding', 'receptor ligand activity', 'immune receptor activity', and 'cytokine receptor activity' (MF). Figure 4B showed the correlation between the top 10 biological functions and genes. According to the KEGG pathway enrichment analysis, the Cytokine-cytokine receptor interaction, Th17 cell differentiation, Rheumatoid arthritis, IL-17 signaling pathway, TNF signaling pathway, and other pathways were found to be highly associated with immune response and inflammation (Figure 4C).
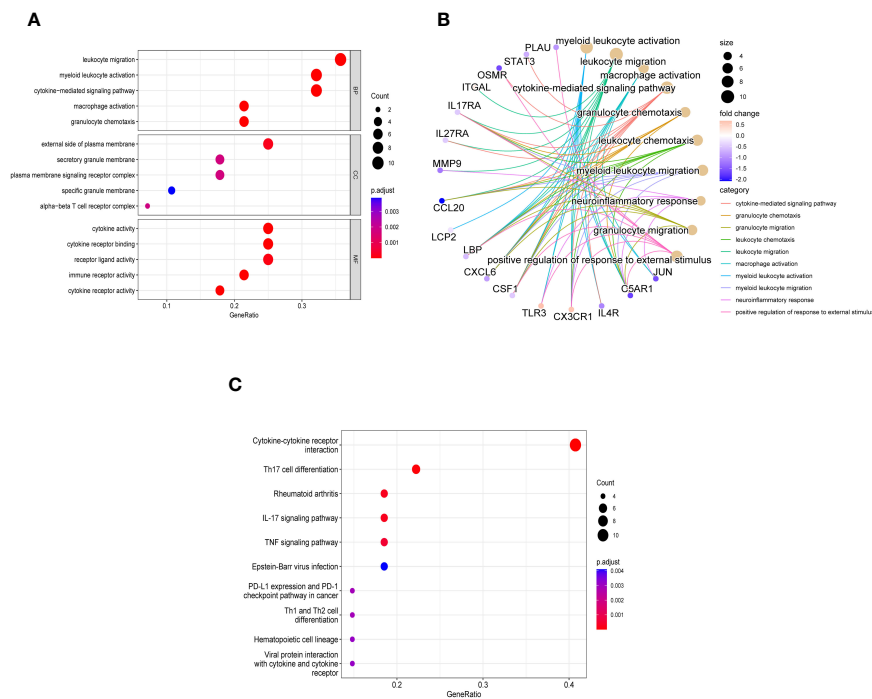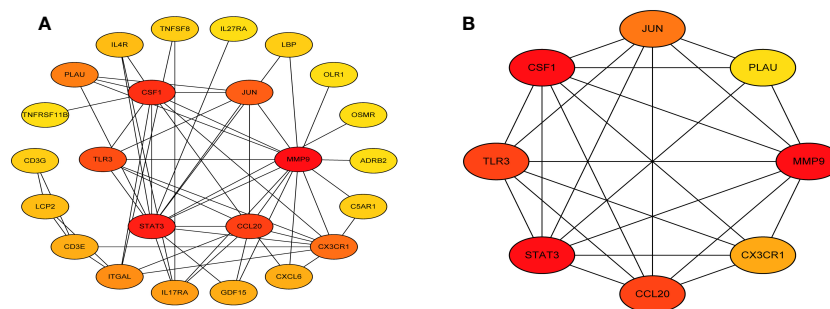
FIGURE 4
Enriched items in GO and KEGG analyses of 28 immune related key genes. **(A)** The enriched terms in GO analysis. **(B)** Correlation between the top 10 biological functions and genes. **(C)** KEGG analysis.

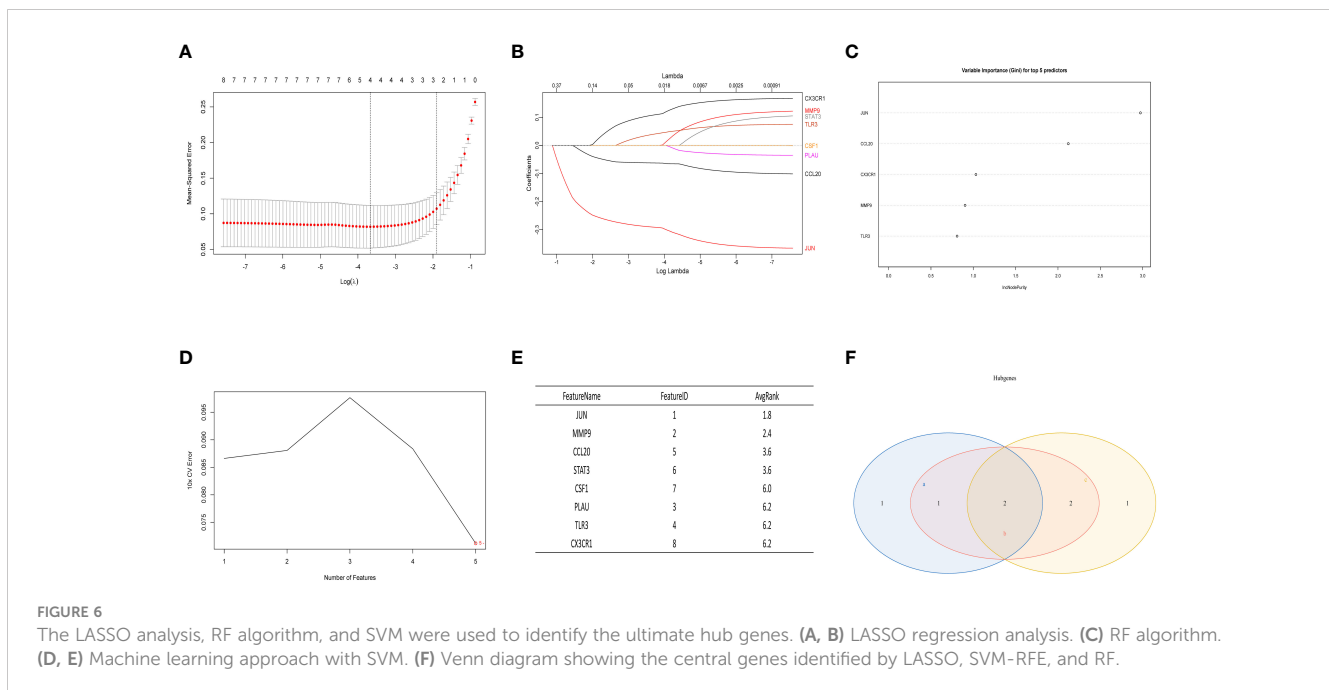## 3.4 Analysis of the network of interactions between proteins

Next, we accessed the STRING database and constructed a PPI network for the immune related key genes using Cytoscape software, 4 genes were kicked out because they had no interaction with the other genes, including CMTM2, PLXNC1, NR1D2, and FABP5 (Figure 5A). MCODE was used to explore the most significant cluster (cluster 1, containing 8 genes). Using the MCC algorithm, the interaction network (included 8 nodes and 48 edges) of these 8 genes were obtained through the CytoHubba plugin of the software Cytoscape (Figure 5B).

3.5 Integrated LASSO analysis, RF algorithm, and SVM for screening hub genes

In the normal vs. SS group, the LASSO Cox regression model was employed to identify the most valuable diagnostic gene signature among the mentioned genes, resulting in the identification of 4 potential genes (Figures 6A, B). Next, the RF algorithm evaluated the significance of each gene and determined the ranking of these 8 genes. From this ranking, we selected the top 5 genes with the highest importance (as shown in Figure 6C). Simultaneously, we utilized a machine learning technique with

FIGURE 5
Visual representation of the protein-protein interaction networks. **(A)** PPI network of immune related key genes. **(B)** Gene clustering based on the MCODE algorithm.

FIGURE 6
The LASSO analysis, RF algorithm, and SVM were used to identify the ultimate hub genes. **(A, B)** LASSO regression analysis. **(C)** RF algorithm.
(D, E) Machine learning approach with SVM. **(F)** Venn diagram showing the central genes identified by LASSO, SVM-RFE, and RF.

SVM to conduct a thorough analysis of the distinct genes, and the findings indicated that opting for the top five genes is suitable (Figures 6D, E). Finally, we intersected the three screening results mentioned above and obtained the optimal gene signature consisting of 2 diagnostic genes, including JUN and CCL20 (Figure 6F).

3.6 Expression and diagnostic efficacy identification of hub genes

After conducting an examination, we observed the expression of two central genes derived from the aforementioned analysis in the normal vs. SS group as well as the SS vs. NASH group. Surprisingly, their expressions were entirely contrasting, indicating that these two genes potentially perform contradictory functions during various phases of the ailment (Figures 7A, B). In order to confirm the diagnostic significance of the two central genes, we generated ROC curves and determined the area under the curve (AUC) for these genes. In normal vs. SS group, the AUC of both two genes were 0.925, while the combined diagnostic efficacy of the two genes was better than that of any single one (AUC: 0.9417) (Figure 7C). In SS vs. NASH group, the AUC of JUN and CCL20 were 0.7579 and 0.8421, respectively, however, the combined diagnostic efficacy of the two genes was not improved (AUC: 0.8289) (Figure 7D).
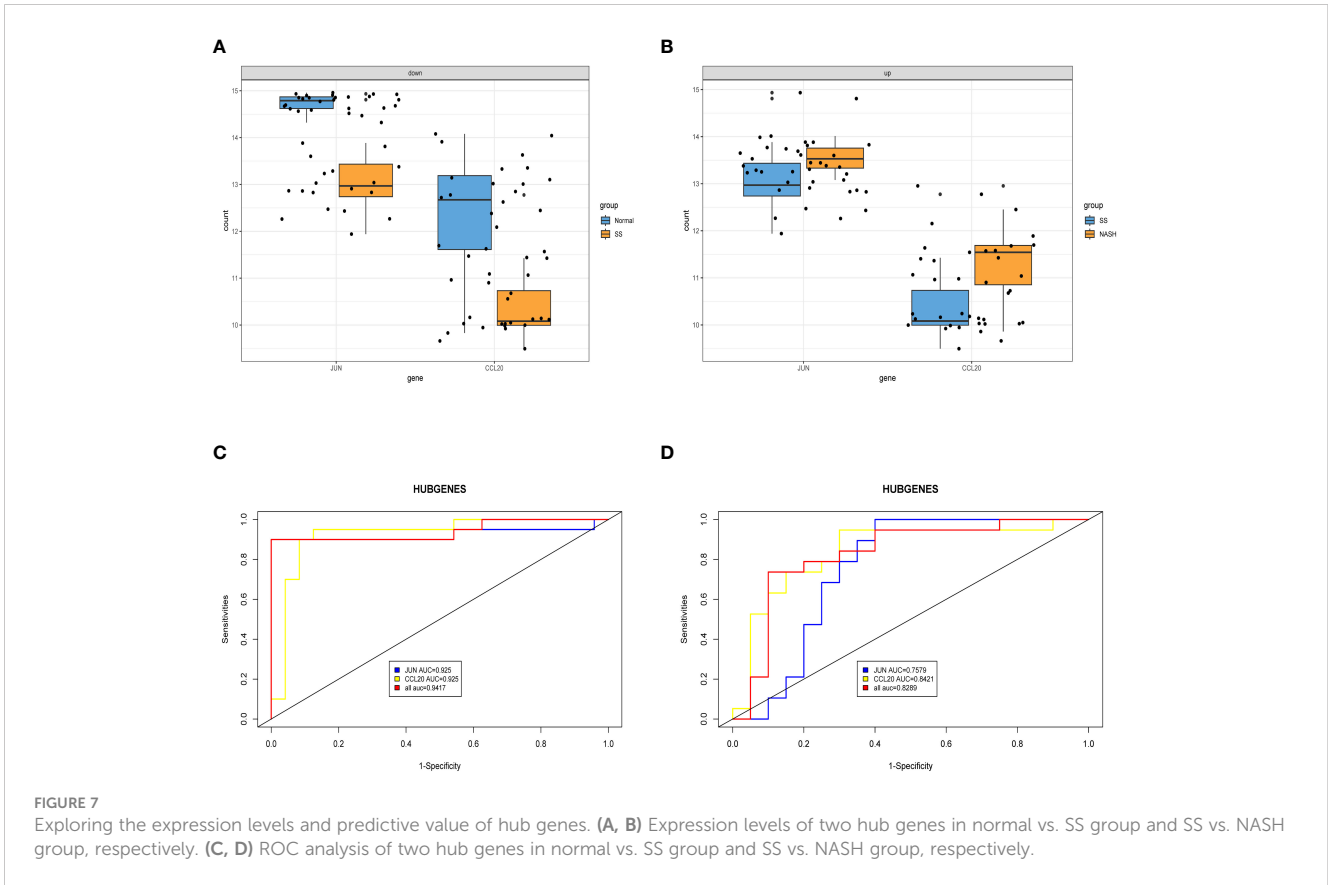
3.7 Validation of hub genes expression and diagnostic efficacy through a GEO dataset and mouse model

For the purpose of confirming the expression patterns and diagnostic effectiveness of hub genes, the dataset GSE135251 was

utilized in the current investigation. The findings indicated that the expression of hub genes aligned with the aforementioned outcomes (Table 2). In addition, in the dataset GSE135251, both genes had good diagnostic efficacy (Figures 8A, B). During the animal study, H&E staining showed obvious fat accumulation in the SS and NASH groups, and a large number of immune cell infiltration in the NASH group (Figure 8C). In comparison to the normal group, we observed a significant decrease in the expression of the two hub genes in the SS group. On the contrary, compared with SS group, the expression levels of these two genes were significantly up-regulated in NASH group. Figures 8D, E displayed the relative mRNA expression of the two genes.

3.8 Analysis of immune infiltration and correlation analysis

Immune dysregulation in the liver microenvironment could potentially be linked to the onset and progression of NAFLD. Hence, in order to comprehend the development of SS and its progression to NASH, the ESTIMATE algorithm was utilized to deduce the infiltration of immune cells. Figures 9A, C, E demonstrated that individuals with SS exhibited decreased immune score, stromal score, and ESTIMATE score in comparison to the healthy controls. In comparison, NASH patients exhibited elevated immune score, stromal score, and ESTIMATE score in contrast to SS patients (Figures 9B, D, F). Furthermore, we employed ssGSEA to assess the variances in the abundance of 28 immune cell subpopulations infiltrating the hepatic tissue between the normal and SS groups, as well as the SS and NASH groups, using data from the GSE89632 dataset (Figures 10A, B). Compared to healthy controls, SS patients exhibited higher levels of immune cell infiltration, including

FIGURE 7
Exploring the expression levels and predictive value of hub genes. **(A, B)** Expression levels of two hub genes in normal vs. SS group and SS vs. NASH group, respectively. **(C, D)** ROC analysis of two hub genes in normal vs. SS group and SS vs. NASH group, respectively.

monocytes, central memory CD4 T cells, gamma delta T cells, central memory CD8 T cells, CD56 bright natural killer cells, activated CD8 T cells, effector memory CD8 T cells, natural killer cells, and effector memory CD4 T cells, in the normal vs. SS group. On the contrary, MDSCs, activated CD4 T cells, mast cells, neutrophils, and eosinophils were enriched in healthy controls. In SS vs. NASH group, high infiltration level of activated CD4 T cells was observed in NASH patients compared to SS patients. Based on Figures 10C, D, we conducted a correlation analysis on immune cells in both SS and NASH patients, revealing noteworthy associations among the majority of immune cells in these individuals. We conducted separate correlation analyses to investigate the connection between our identified hub genes JUN and CCL20 and the content of immune cells. Among individuals with SS, there was a robust positive association between JUN expression and Plasmacytoid dendritic cell (r = 0.890), whereas the CCL20 expression exhibited a noteworthy inverse relationship with CD56 bright natural killer cell (r = -0.686) (Figure 10E). Among individuals with NASH, there was a notable inverse

relationship between JUN expression and macrophage, with a correlation coefficient of -0.611. Conversely, the CCL20 expression exhibited a robust positive correlation with activated CD4 T cell, with a correlation coefficient of 0.767 (as shown in Figure 10F). Collectively, this suggested that these central genes facilitated the immune response throughout the progression of the disease in individuals with NAFLD.
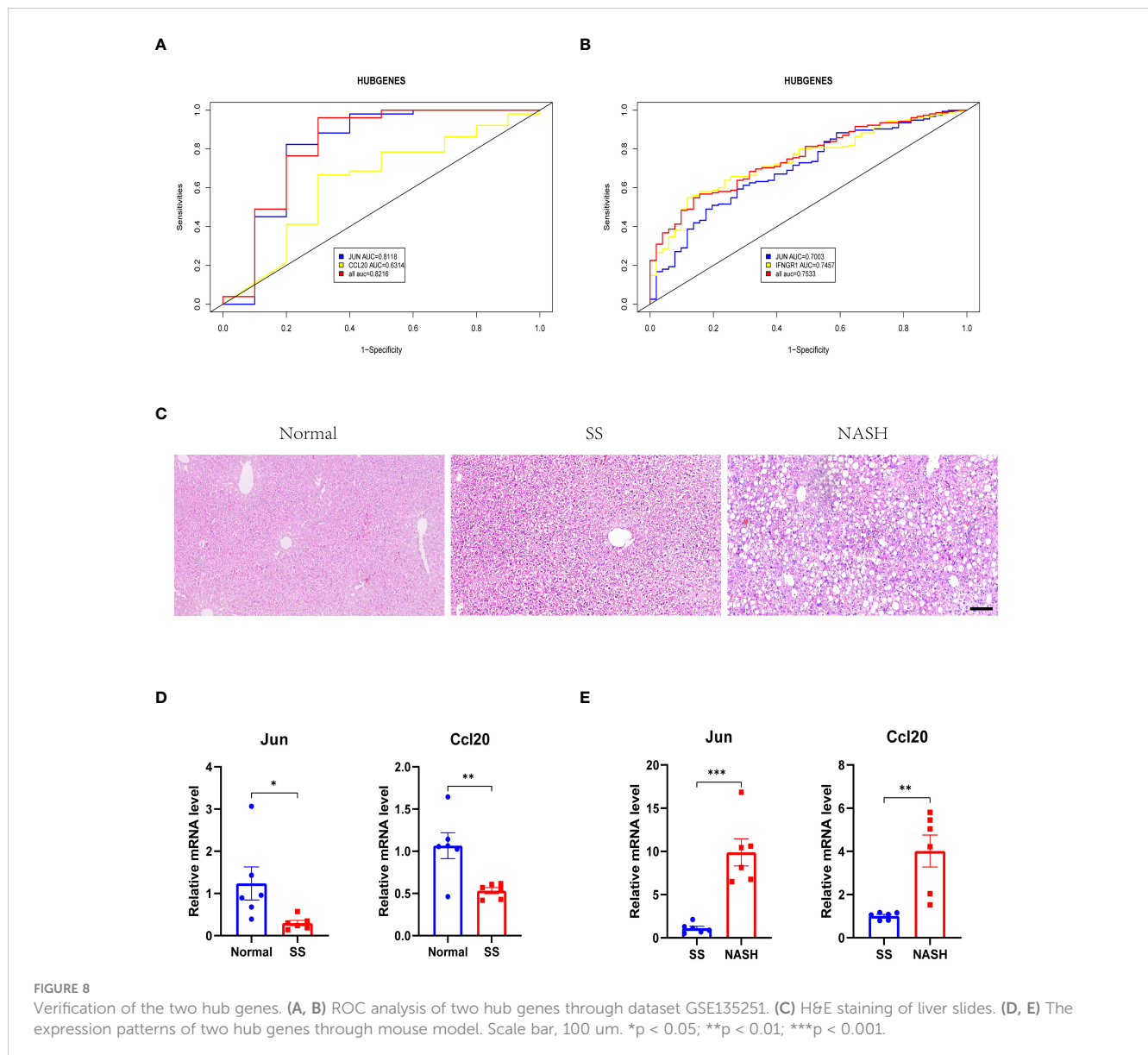
4 Discussion

NAFLD is a complex disease caused by multiple factors, especially obesity. Previous studies have extensively explored its pathogenesis and found that the mechanism of its development was affected by a variety of factors, such as age, menopause, and type 2 diabetes (T2D) [25, 26]. Since the accumulation of fat may affect the infiltration of immune cells, the abnormal function of the immune system in NAFLD has been paid more and more attention [27, 28]. Song et al. found that C/EBPα was significantly upregulated in

TABLE 2 The gene expression pattern in dataset GSE135251 for hub genes.

Normal vs. SS			SS vs. NASH		
Gene Symbol	logFC	Adjusted p-Value	Gene Symbol	logFC	Adjusted p-Value
JUN	-1.4096286	6.73e-05	JUN	0.4997750	9.93e-04
CCL20	-0.1656848	2.63e-01	CCL20	0.5086908	2.36e-04

FIGURE 8

Verification of the two hub genes. **(A, B)** ROC analysis of two hub genes through dataset GSE135251. **(C)** H&E staining of liver slides. **(D, E)** The expression patterns of two hub genes through mouse model. Scale bar, 100 um. *p < 0.05; **p < 0.01; ***p < 0.001.

NASH compared with healthy controls, possibly contributing to disease progression by regulating intrahepatic immune and inflammatory responses (29). In addition, multiple research studies validate that both inherent and acquired immune disorders are significant contributors to the development and advancement of NAFLD, for example, during NASH, there is a large infiltration of neutrophils and NKT cells in the liver, which is closely related to the development of the disease in the direction of increased inflammation (30–32). Moreover, platelets in the liver interact with Kupffer cells to induce the secretion of alpha granules, which contains large amounts of growth factors as well as cytokines, thereby exacerbating liver inflammation (33). Modifying the expression of IRGs brings about variations in the infiltration level and functional state of immune cells (34). Disorders of IRGs have been described in a variety of diseases, such as osteosarcoma and Alzheimer's disease (35, 36). Considering these factors, we implemented a thorough and extensive assessment system to examine the immune-associated hub genes and molecular pathways in the onset and progression of NAFLD using bioinformatics. Our objective is to expand the understanding of the physiological pathology and molecular mechanisms of NAFLD, and offer insights for clinical diagnosis and treatment approaches.

For this particular investigation, we obtained the gene expression matrix of SS tissue in comparison to normal liver tissue and NASH tissue in comparison to SS tissue from the GSE89632 dataset. In normal vs. SS group, we obtained 11 SS related modules using WGCNA, and after further screening, we identified 4 modules that were strongly correlated with SS, and finally obtained 2939 module genes (Figure 2). In SS vs. NASH group, we performed differential analysis using another analytical method and obtained 1222 DEGs, of which 694 genes were up-regulated and 528 genes were down-regulated. By intersecting module genes, DEGs, and IRGs, we identified 28 immune-related genes that played a role in both the onset and progression of
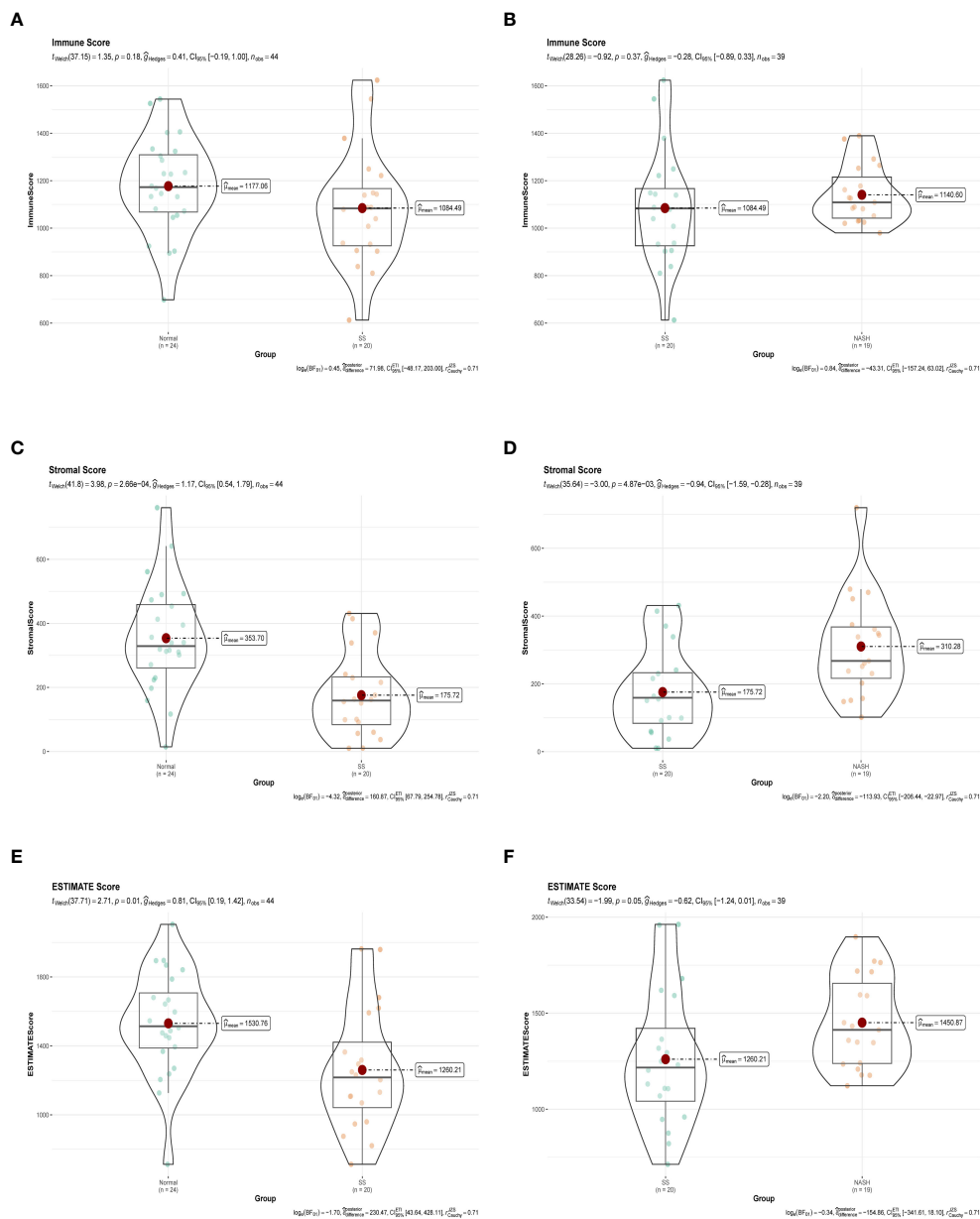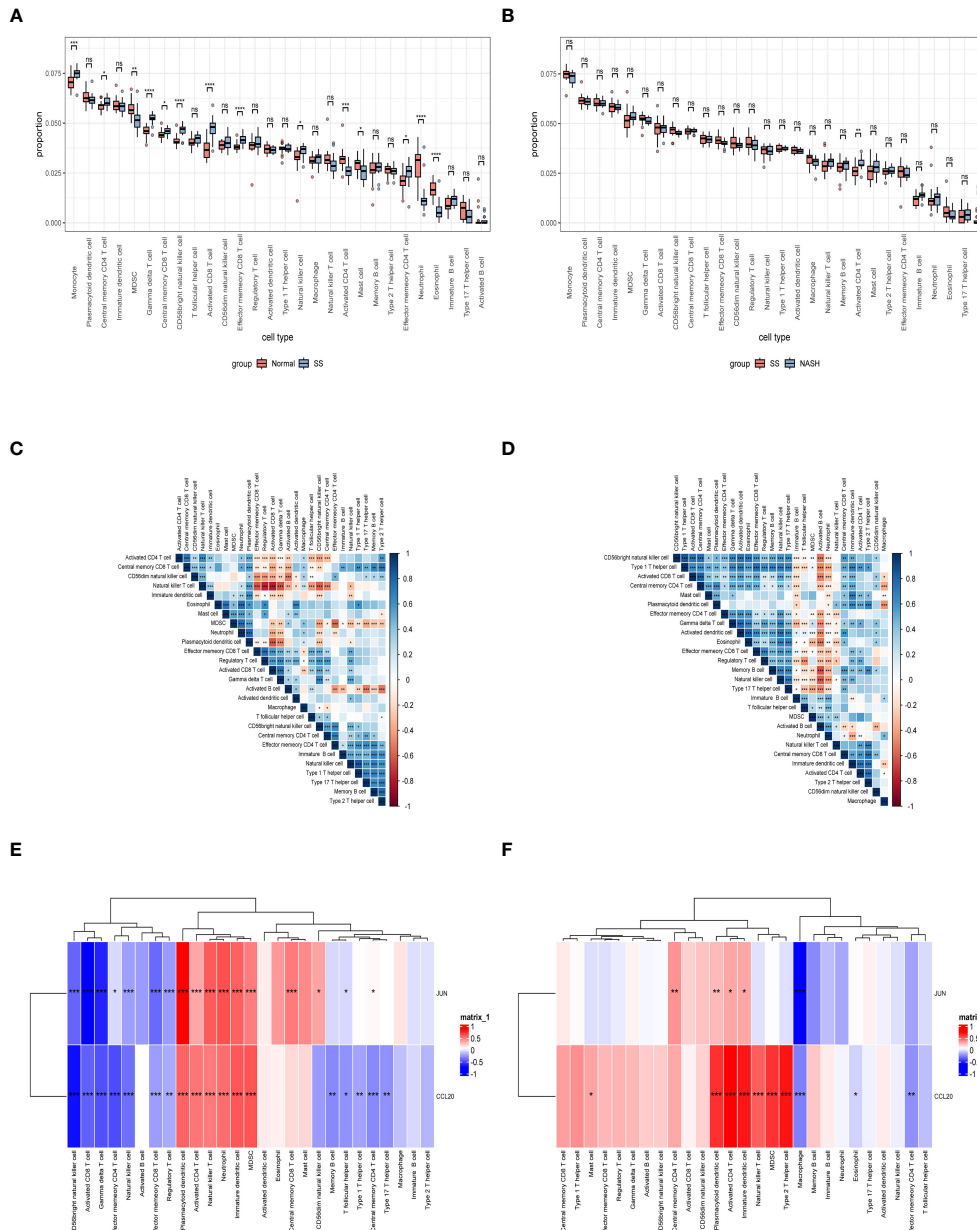
FIGURE 9
Immune status assessment by ESTIMATE algorithm. **(A, B)** Immune Score in normal vs. SS group and SS vs. NASH group, respectively. **(C, D)** Stromal Score in normal vs. SS group and SS vs. NASH group, respectively. **(E, F)** ESTIMATE Score in normal vs. SS group and SS vs. NASH group, respectively.

NAFLD (Figure 3C). Previous studies performed high-throughput sequencing of liver tissue from normal mice and mice at different stages of NASH, they found that compared with normal liver tissue, the expression patterns of genes in the liver tissue of model mice and the signaling pathways involved changed as the disease progressed (37). In our study, we performed a follow-up analysis to determine whether the expression patterns of these 28 genes also changed as the NAFLD progressed. Additional examination of the comparative expression of these 28 genes uncovered a fascinating observation that these genes did not consistently exhibit increased or decreased regulation at the initiation of NAFLD and the advancement to NASH (Figures 3D, F). This suggested that the immune microenvironment might be diametrically opposed

at different stages of the disease. Analysis of the immune infiltration score further confirmed our conjecture (Figure 9). Dynamic changes in immune cell infiltration in the immune microenvironment at different stages of the disease had previously been demonstrated in a variety of diseases (38, 39). In addition, this also suggested that the treatment might not be consistent at different stages of the disease.

In order to obtain a deeper understanding of the possible roles of these genes associated with the immune system, we utilized bioinformatics tools to conduct GO and KEGG enrichment analyses on the set of 28 genes. Analysis of the genes revealed their involvement in immune-related pathways, particularly cytokine-related pathways and the activation and chemotaxis of

FIGURE 10

Exploration of immune infiltration in NAFLD and its association with central genes through ssGSEA. **(A, B)** Analysis of the immunocyte infiltration degrees regarding 28 immunocyte subunits in normal vs. SS group and SS vs. NASH, respectively. **(C, D)** Correlation between different immune cells in SS and NASH patients, respectively. The correlation between immune cell infiltration and two hub genes in patients with SS and NASH, respectively. *p < 0.05; **p < 0.01; ***p < 0.001; ns: no significance.

immune cells (Figure 4). This could explain the lower infiltration of activated CD4 T cells during the early stage of the disease and their higher level in NASH. And through a variety of bioinformatics analysis methods, such as PPI, LASSO, RF, and SVM, we further screened from these 28 genes and obtained 2 hub genes, including JUN and CCL20.

JUN plays a vital role in activator protein 1 (AP-1), being essential for liver development and contributing to the onset and progression of diverse liver disorders (40). Previous studies focused on the differential expression of this molecule between NAFLD and normal healthy controls, but the changes in the expression level of

this molecule at different stages of the disease and its remodeling effect on the liver immune microenvironment were rarely mentioned (41–43). During this investigation, it was discovered that JUN exhibited a decrease in expression levels in SS in comparison to individuals without any health issues. This was consistent with the findings of Qu et al., however, they did not further explore the expression pattern of JUN in NASH stage (44). Our study had found that as the disease advanced to NASH, JUN was notably up-regulated. Furthermore, this particular molecule played a crucial role in governing the infiltration of various immune cells throughout different phases of the disease.

CCR6[+] cells are driven to migrate through tissues by the high affinity binding of CCL20 to its receptor CCR6, which is currently the sole known ligand for CCR6 (45). Increased infiltration of CCR6[+] lymphocytes with upregulated CCL20 expression was found in tissue microenvironment during inflammation, infection and malignant lesions in various organs, such as stomach, intestine, liver and lung (46). In addition, one study found that postmenopausal women with T2D were more likely to have upregulated CCL20 expression levels, which might be closely related to a more pronounced liver inflammatory response and susceptibility to NASH in these patients (25). In this study, we found that CCL20 was down-regulated in SS compared with healthy controls, but significantly up-regulated when the disease progressed to NASH. In our research, it was discovered that this particular molecule exhibited a strong positive correlation with the infiltration of activated CD4 T cells in NAFLD. Earlier research had validated that stimulated CD4 T cell types, like Th1 and Th17, possess the ability to release diverse cytokines, including IFN-γand IL-17, thereby enhancing liver inflammation (47–49). Hence, we hypothesized that this compound might have a crucial function in the progression of SS to NASH.

HFD-induced mice and MCD-induced mice are currently recognized animal models that can mimic human NAFLD, and these models are highly similar to human SS and NASH in histologic appearance and liver transcriptome characteristics (50, 51). Therefore, in our study we constructed the above two models. Afterwards, we confirmed the gene expression by conducting *in vivo* experiments, the results aligned with the sequencing findings. Simultaneously, we also confirmed the expression of these two compounds by analyzing an additional dataset, GSE135251, and the outcome aligned with our discoveries. We then further explored the diagnostic capabilities of JUN and CCL20 and found that they were able to distinguish between different stages of NAFLD disease. In summary, JUN and CCL20 had been identified as key targets for mediating the onset of NAFLD and leading to its progression from SS to NASH, and may be markers for predicting disease progression.

In conclusion, we proposed that there were dynamic changes in the expression levels of IRGs in different stages of NAFLD disease, and they regulated the immune microenvironment of the liver mainly through the cytokine related pathways and immune cell activation and chemotaxis, which suggested that the treatment might not be consistent at different stages of the disease. Among them, JUN and CCL20 might be the key molecules that promoted the occurrence and progression of the disease, and had potential as diagnostic markers and therapeutic targets. However, our study was based on the public database, therefore, further experiments were needed to explore and verify the mechanism of the results obtained from the analysis. We believe these will further deepen our understanding of the disease and provide references for the treatment of the disease.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Ethics statement

The animal study was approved by the Ethics Committee of Second Affiliated Hospital of Harbin Medical University. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

RH: Writing – original draft, Data curation. CG: Writing – original draft, Data curation. XZ: Writing – review & editing. LY: Writing – review & editing. YC: Writing – review & editing.

Funding

Acknowledgments

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.

References

1. Park J, Morley TS, Kim M, Clegg DJ, Scherer PE. Obesity and cancer–mechanisms underlying tumour progression and recurrence. *Nat Rev Endocrinol*. (2014) 10:455–65. doi: 10.1038/nrendo.2014.94

2. Drucker DJ. Diabetes, obesity, metabolism, and SARS-CoV-2 infection: the end of the beginning. *Cell Metab*. (2021) 33:479–98. doi: 10.1016/j.cmet.2021.01.016

3. Sarma S, Sockalingam S, Dash S. Obesity as a multisystem disease: Trends in obesity rates and obesity-related complications. *Diabetes Obes Metab*. (2021) 23 Suppl 1:3–16. doi: 10.1111/dom.14290

4. Pierantonelli I, Svegliati-Baroni G. Nonalcoholic fatty liver disease: basic pathogenetic mechanisms in the progression from NAFLD to NASH. *Transplantation*. (2019) 103:e1–e13. doi: 10.1097/TP.0000000000002480

5. Estes C, Anstee QM, Arias-Loste MT, Bantel H, Bellentani S, Caballeria J, et al. Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016-2030. *J Hepatol*. (2018) 69:896–904. doi: 10.1016/j.jhep.2018.05.036

6. Friedman SL, Neuschwander-Tetri BA, Rinella M, Sanyal AJ. Mechanisms of NAFLD development and therapeutic strategies. *Nat Med*. (2018) 24:908–22. doi: 10.1038/s41591-018-0104-9

7. Peiseler M, Schwabe R, Hampe J, Kubes P, Heikenwälder M, Tacke F. Immune mechanisms linking metabolic injury to inflammation and fibrosis in fatty liver disease - novel insights into cellular communication circuits. *J Hepatol*. (2022) 77:1136–60. doi: 10.1016/j.jhep.2022.06.012

8. Anstee QM, Reeves HL, Kotsiliti E, Govaere O, Heikenwalder M. From NASH to HCC: current concepts and future challenges. *Nat Rev Gastroenterol Hepatol*. (2019) 16:411–28. doi: 10.1038/s41575-019-0145-7

9. Huby T, Gautier EL. Immune cell-mediated features of non-alcoholic steatohepatitis. *Nat Rev Immunol*. (2022) 22:429–43. doi: 10.1038/s41577-021-00639-3

10. Hoogerland JA, Staels B, Dombrowicz D. Immune-metabolic interactions in homeostasis and the progression to NASH. *Trends Endocrinol Metab*. (2022) 33:690–709. doi: 10.1016/j.tem.2022.07.001

11. Dudek M, Pfister D, Donakonda S, Filpe P, Schneider A, Laschinger M, et al. Auto-aggressive CXCR6+ CD8 T cells cause liver immune pathology in NASH. *Nature*. (2021) 592:444–9. doi: 10.1038/s41586-021-03233-8

12. Arab JP, Arrese M, Trauner M. Recent insights into the pathogenesis of nonalcoholic fatty liver disease. *Annu Rev Pathol*. (2018) 13:321–50. doi: 10.1146/annurev-pathol-020117-043617

13. Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J, et al. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci Data*. (2018) 5:180015. doi: 10.1038/sdata.2018.15

14. 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

15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. (2015) 43:e47. doi: 10.1093/nar/gkv007

16. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. (2012) 16:284–7. doi: 10.1089/omi.2011.0118

17. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, et al. A travel guide to Cytoscape plugins. *Nat Methods*. (2012) 9:1069–76. doi: 10.1038/nmeth.2212

18. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. (2019) 47:D607–13. doi: 10.1093/nar/gky1131

19. Maksimov MO, Pan SJ, James Link A. Lasso peptides: structure, function, biosynthesis, and engineering. *Nat Prod Rep*. (2012) 29:996–1006. doi: 10.1039/c2np20070h

20. Guo L, Wang Z, Du Y, Mao J, Zhang J, Yu Z, et al. Random-forest algorithm based biomarkers in predicting prognosis in the patients with hepatocellular carcinoma. *Cancer Cell Int*. (2020) 20:251. doi: 10.1186/s12935-020-01274-z

21. Lin X, Yang F, Zhou L, Yin P, Kong H, Xing W, et al. A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information. *J Chromatogr B Analyt Technol BioMed Life Sci*. (2012) 910:149–55. doi: 10.1016/j.jchromb.2012.05.020

22. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. (2013) 4:2612. doi: 10.1038/ncomms3612

23. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep*. (2017) 18:248–62. doi: 10.1016/j.celrep.2016.12.019

24. Yu B, Yin YX, Tang YP, Wei KL, Pan ZG, Li KZ, et al. Diagnostic and predictive value of immune-related genes in Crohn's disease. *Front Immunol*. (2021) 12:643036. doi: 10.3389/fimmu.2021.643036

25. Raverdy V, Chatelain E, Lasailly G, Caiazzo R, Vandel J, Verkindt H, et al. Combining diabetes, sex, and menopause as meaningful clinical features associated with NASH and liver fibrosis in individuals with class II and III obesity: A retrospective cohort study. *Obes (Silver Spring)*. (2023) 31:3066–76. doi: 10.1002/oby.23904

26. Gawrieh S, Karns R, Kleiner DE, Olivier M, Jenkins T, Inge TH, et al. Comparative analysis of global hepatic gene expression in adolescents and adults with non-alcoholic fatty liver disease. *Arch Clin BioMed Res*. (2023) 7:112–9. doi: 10.26502/acbr.50170323

27. Hotamisligil GS. Inflammation, metaflammation and immunometabolic disorders. *Nature*. (2017) 542:177–85. doi: 10.1038/nature21363

28. Ertunc ME, Hotamisligil GS. Lipid signaling and lipotoxicity in metaflammation: indications for metabolic disease pathogenesis and treatment. *J Lipid Res*. (2016) 57:2099–114. doi: 10.1194/jlr.R066514

29. Song Z, Wang Y, Lin P, Yang K, Jiang X, Dong J, et al. Identification of key modules and driving genes in nonalcoholic fatty liver disease by weighted gene co-expression network analysis. *BMC Genomics*. (2023) 24:414. doi: 10.1186/s12864-023-09458-3

30. Tilg H, Adolph TE, Dudek M, Knolle P. Non-alcoholic fatty liver disease: the interplay between metabolism, microbes and immunity. *Nat Metab*. (2021) 3:1596–607. doi: 10.1038/s42255-021-00501-9

31. Yahoo N, Dudek M, Knolle P, Heikenwälder M. Role of immune responses in the development of NAFLD-associated liver cancer and prospects for therapeutic modulation. *J Hepatol*. (2023) 79:538–51. doi: 10.1016/j.jhep.2023.02.033

32. Wu L, Gao X, Guo Q, Li J, Yao J, Yan K, et al. The role of neutrophils in innate immunity-driven nonalcoholic steatohepatitis: lessons learned and future promise. *Hepatol Int*. (2020) 14:652–66. doi: 10.1007/s12072-020-10081-7

33. Malehmir M, Pfister D, Gallage S, Szydlowska M, Inverso D, Kotsiliti E, et al. Platelet GPIbα is a mediator and potential interventional target for NASH and subsequent liver cancer. *Nat Med*. (2019) 25:641–55. doi: 10.1038/s41591-019-0379-5

34. Lu Y, Li K, Hu Y, Wang X. Expression of immune related genes and possible regulatory mechanisms in Alzheimer's disease. *Front Immunol*. (2021) 12:768966. doi: 10.3389/fimmu.2021.768966

35. Liu W, Xie X, Qi Y, Wu J. Exploration of immune-related gene expression in osteosarcoma and association with outcomes. *JAMA Netw Open*. (2021) 4:e2119132. doi: 10.1001/jamanetworkopen.2021.19132

36. Chen Y, Dai J, Tang L, Mikhailova T, Liang Q, Li M, et al. Neuroimmune transcriptome changes in patient brains of psychiatric and neurological disorders. *Mol Psychiatry*. (2023) 28:710–21. doi: 10.1038/s41380-022-01854-7

37. Zhao Y, Yakufu M, Ma C, Wang B, Yang J, Hu J. Transcriptomics reveal a molecular signature in the progression of nonalcoholic steatohepatitis and identifies PAI-1 and MMP-9 as biomarkers in *in vivo* and *in vitro* studies. *Mol Med Rep*. (2024) 29:15. doi: 10.3892/mmr.2023.13138

38. de Visser KE, Joyce JA. The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. *Cancer Cell*. (2023) 41:374–403. doi: 10.1016/j.ccell.2023.02.016

39. Zhang C, Li J, Cheng Y, Meng F, Song JW, Fan X, et al. Single-cell RNA sequencing reveals intrahepatic and peripheral immune characteristics related to disease phases in HBV-infected patients. *Gut*. (2023) 72:153–67. doi: 10.1136/gutjnl-2021-325915

40. Schulien I, Hockenjos B, Schmitt-Graeff A, Perdekamp MG, Follo M, Thimme R, et al. The transcription factor c-Jun/AP-1 promotes liver fibrosis during non-alcoholic steatohepatitis by regulating Osteopontin expression. *Cell Death Differ*. (2019) 26:1688–99. doi: 10.1038/s41418-018-0239-8

41. Wang J, Ma J, Nie H, Zhang XJ, Zhang P, She ZG, et al. Hepatic regulator of G protein signaling 5 ameliorates nonalcoholic fatty liver disease by suppressing transforming growth factor beta-activated kinase 1-c-jun-N-terminal kinase/p38 signaling. *Hepatology*. (2021) 73:104–25. doi: 10.1002/hep.31242

42. Min RWM, Aung FWM, Liu B, Arya A, Win S. Mechanism and therapeutic targets of c-jun-N-terminal kinases activation in nonalcoholic fatty liver disease. *Biomedicines*. (2022) 10:2035. doi: 10.3390/biomedicines10082035

43. Yan FJ, Wang X, Wang SE, Hong HT, Lu J, Ye Q, et al. C-Jun/C7ORF41/NF-κB axis mediates hepatic inflammation and lipid accumulation in NAFLD. *Biochem J*. (2020) 477:691–708. doi: 10.1042/BCJ20190799

44. Qu S, Huang H, Diao Y, Liu B, Tang B, Huo S, et al. Identification of biomarkers associated with immune-propionate metabolism in nonalcoholic fatty liver disease. *Cell Mol Biol (Noisy-le-grand)*. (2023) 69:254–63. doi: 10.14715/cmb/2023.69.10.38

45. Meitei HT, Jadhav N, Lal G. CCR6-CCL20 axis as a therapeutic target for autoimmune diseases. *Autoimmun Rev*. (2021) 20:102846. doi: 10.1016/j.autrev.2021.102846

46. Lee B, Namkoong H, Yang Y, Huang H, Heller D, Szot GL, et al. Single-cell sequencing unveils distinct immune microenvironments with CCR6-CCL20 crosstalk in human chronic pancreatitis. *Gut*. (2022) 71:1831–42. doi: 10.1136/gutjnl-2021-324546

47. Rolla S, Alchera E, Imarisio C, Bardina V, Valente G, Cappello P, et al. The balance between IL-17 and IL-22 produced by liver-infiltrating T-helper cells critically controls NASH development in mice. *Clin Sci (Lond)*. (2016) 130:193–203. doi: 10.1042/CS20150405

48. Gomes AL, Teijeiro A, Burén S, Tummala KS, Yilmaz M, Waisman A, et al. Metabolic inflammation-associated IL-17A causes non-alcoholic steatohepatitis and hepatocellular carcinoma. *Cancer Cell*. (2016) 30:161–75. doi: 10.1016/j.ccell.2016.05.020

49. Mao T, Yang R, Luo Y, He K. Crucial role of T cells in NAFLD-related disease: A review and prospect. *Front Endocrinol (Lausanne)*. (2022) 13:1051076. doi: 10.3389/fendo.2022.1051076

50. Ding Y, Dai X, Bao M, Xing Y, Liu J, Zhao S, et al. Hepatic transcriptome signatures in mice and humans with nonalcoholic fatty liver disease. *Anim Model Exp Med*. (2023) 6:317–28. doi: 10.1002/ame2.12338

51. Alshawsh MA, Alsalahi A, Alshehade SA, Saghir SAM, Ahmeda AF, Al Zarzour RH, et al. A comparison of the gene expression profiles of non-alcoholic fatty liver disease between animal models of a high-fat diet and methionine-choline-deficient diet. *Molecules*. (2022) 27:858. doi: 10.3390/molecules27030858