Published online 2021 October 20.

Research Article

Identification of a 17-gene-signature in Non-alcoholic Steatohepatitis and Its Relationship with Immune Cell Infiltration

Junwang Zhang¹, Xumei Kang², Ling Zhang³, Huiming Wang¹ and Zhihua Deng^{1,*}

¹Gastroenterology Department, Second Hospital of Shanxi Medical University, Taiyuan City, China
²Imaging Department, Second Hospital of Shanxi Medical University, Taiyuan City, China
³Nursing Department, Second Hospital of Shanxi Medical University, Taiyuan City, China

^{*} Corresponding author: Gastroenterology Department, Second Hospital of Shanxi Medical University, No. 236, Wuyi Road, Xinghualing District, Taiyuan City, Shanxi Province, 030001, China. Email: junwangzhang@126.com

Received 2021 May 19; Revised 2021 September 03; Accepted 2021 September 04.

Abstract

Background: Non-alcoholic steatohepatitis (NASH) is a risk factor for hepatocellular carcinoma, but the understanding of the regulatory mechanisms driving NASH is not comprehensive.

Objectives: We aimed to identify the potential markers of NASH and explore their relationship with immune cell populations. **Methods:** Five gene expression datasets for NASH were downloaded from the Gene Expression Omnibus and European Bioinformatics Institute Array Express databases. Differentially expressed genes (DEGs) between NASH and controls were screened. Gene Ontology-Biological Process (GO-BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed for functional enrichment analysis of DEGs. Among the candidate genes selected from the protein-protein interaction (PPI) network and module analysis, DEG signatures were further identified using least absolute shrinkage and selection operator regression analysis. The Spearman correlation coefficient was calculated to assess the correlation between DEG signatures and immune cell abundance based on the CIBERSORT algorithm.

Results: We screened 403 upregulated, and 158 downregulated DEGs for NASH, and they were mainly enriched in GO-BP, including the inflammatory response, innate immune response, signal transduction, and KEGG pathways, such as the pathways involved in cancer (e.g., the PI3K-Akt signaling pathway), and focal adhesion. We then screened 73 candidate genes from the PPI network and module analysis and finally identified 17 DEG signatures. By evaluating their relationship with immune cell populations, 12 DEG signatures were found to correlate with activated dendritic cells, resting dendritic cells, M2 macrophages, monocytes, neutrophils, and resting memory CD4 T cells, which were significantly different between the NASH and control tissues.

Conclusions: We identified a 17-DEG-signature as a candidate biomarker for NASH and analyzed its relationship with immune infiltration in NASH.

Keywords: Non-Alcoholic Steatohepatitis, Gene Expression Signatures, Immune Infiltration, Protein-Protein Interaction

1. Background

Non-alcoholic steatohepatitis (NASH), which develops from steatosis, is characterized by lipid metabolic disorder and chronic inflammation, eventually leading to liver cirrhosis and hepatocellular carcinoma (HCC) if left untreated (1). As a progressive form of non-alcoholic fatty liver disease (NAFLD), NASH has been predicted to become the leading cause of end-stage liver disease requiring liver transplantation (2). The "two-hit" theory is widely used to explain the pathogenesis of NASH. The first hit is hepatic steatosis resulting from the accumulation of triglycerides, and the second hit is the oxidative stress response to liver damage, inflammation, and fibrosis (3, 4). According to global statistics, approximately 25% of the general population is suffering from NAFLD, and the estimated prevalence of NASH among NAFLD patients is approximately 60% (5, 6). Although patients with simple steatosis (SS) have an optimistic survival and prognosis, the overall morbidity and mortality rates of patients with NASH increase annually (7). To date, liver biopsy remains the gold standard for the diagnosis of NASH, but due to the high incidence of NASH worldwide, liver biopsy is not suitable for large-scale screening (8). Therefore, there is an urgent need to identify novel biomarkers to distinguish NASH from SS and healthy conditions and to better define the severity of liver injury and inflammation.

The development of NASH stems from an inflammatory process that is driven by multiple immune cells. Studies have reported that in the liver, a large number

Copyright © 2021, Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/) which permits copy and redistribute the material just in noncommercial usages, provided the original work is properly cited.

of innate and adaptive immune cells, including lymphocytes, macrophages, neutrophils, and dendritic cells, are involved in the development of chronic inflammation (9). Non-parenchymal hepatic cells, including liver sinusoidal endothelial cells and hepatic stellate cells, have been shown to participate in NASH progression (10). Damaged liver cells release signaling molecules, including injuryassociated molecular patterns and pathogen-associated molecular patterns, thereby arousing an immune response (11). Moreover, these liver immune-associated cells can sense excess metabolites and bacterial products and transduce these signals to activate immune responses in NASH (12). Although the activation of the immune system and the recruitment of pro-inflammatory factors in NASH are widely recognized, the signaling and molecular regulatory mechanisms that facilitate this process remain unclear.

Therefore, this study aimed to identify the characteristic genes of NASH and explore the relationship between them and immune cells. To obtain optimized feature genes, we screened differentially expressed genes (DEGs) between NASH samples and healthy controls, followed by the function and pathway enrichment analysis of these DEGs. We then attempted to screen candidate genes from protein-protein interaction (PPI) network-based module analysis and performed least absolute shrinkage and selection operator (LASSO) regression to identify the gene signatures of NASH.

2. Objectives

In this study, we identified several predictive biomarkers for NASH diagnosis, as well as their potential regulatory mechanisms involved in immune response activation and NASH development.

3. Methods

3.1. Data Acquisition and Processing

Four microarray datasets, including GSE89632 (containing 20 SS, 19 NASH, and 24 normal liver biopsy tissue samples detected from Illumina HumanHT-12 WG-DASL v4.0 R2 expression BeadChip), GSE63067 (containing nine NASH and seven normal liver biopsy tissue samples detected from Affymetrix Human Genome U133 Plus 2.0 Array), GSE107231 (containing five NAFLD and five normal liver biopsy tissue samples detected from Agilent-067406 Human CBC lncRNA + mRNA microarray v4.0), and GSE72756 (containing five NAFLD and five normal liver biopsy tissue samples detected from Agilent-045997 Arraystar human lncRNA microarray v3), were downloaded from Gene Expression Omnibus (GEO) (13). In addition, the E-MEXP-3291 containing 10 SS samples, 16 NASH samples, and 19 healthy controls, detected from Affymetrix GeneChip Human Gene 1.0 ST Array, was obtained from European Bioinformatics Institute (EBI) Array Express (14). The expression data of the samples from these five datasets were acquired accordingly, of which GSE89632 and E-MEXP-3291 were used for analysis while GSE63067, GSE107231, and GSE72756 were employed for validation. Sample employment and study procedures have been shown in Figure 1. The probe was transformed into a gene symbol according to the platform annotation information, and when multiple probes corresponded to one gene, the mean value was adopted.

3.2. Screening for DEGs in NASH

The limma package (version 3.34.7) in R 3.6.1(15) was applied to select DEGs between NASH samples and healthy controls in the both GSE89632 and E-MEXP-3291 datasets. Batch effects on the expression data rooted into two detection platforms and were removed by the sva package (version 3.38.0)(16) and visualized by principal component analysis (PCA) diagrams. An expression heatmap was generated using the pheatmap package (version 1.0.12). Intersecting DEGs from the two databases were selected for further analysis.

3.3. Enrichment Analysis of DEGs

The online tool of DAVID (version 6.8) (17, 18) was used for the gene ontology-biological process (GO-BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment of DEGs. Thresholds were set at P< 0.05 and gene count ≥ 2 .

3.4. PPI Network Establishment and Module Analysis

In this section, STRING version 11.0 (19) was used to establish functional interactions between proteins with a threshold of combined score = 0.70 (high confidence), and then the established network was visualized through Cytoscape version 3.6.1 (20). Moreover, the CytoNCA plug-in version 2.1.6 (21) was used to calculate the degree centrality, betweenness centrality, and closeness centrality of nodes in the network. The hub proteins involved in the PPI network were obtained by ranking the topological properties of each node.

MCODE plug-in version 1.5.1 (22) was used to identify module genes in the PPI network based on the following settings: degree cut-off=2, node score cut-off=0.2, K-core = 2, and max depth=100. Then the clusterProfiler package in



Figure 1. The flowchart of the study. NASH, non-alcoholic steatohepatitis; DEGs, differentially expressed genes; FDR, false discovery rate; FC, fold change; PPI, protein-protein interaction; LASSO, least absolute shrinkage and selection operator.

R (version 3.8.1) (23) was used to conduct KEGG pathway enrichment analysis for the genes in each module and identify key module genes according to the significance of enrichment results at the threshold of P < 0.05.

3.5. Selection and Validation of Core Genes

To identify hub genes in the PPI network and key module genes as candidate genes, LASSO regression analysis was performed using Lars package in R 3.6.1 (version 1.2) for the screening of DEG signatures. Thereafter, the expression of DEG signatures was validated on the GSE63067, GSE107231, and GSE72756 datasets.

3.6. Identification of Immune Infiltration-related DEG Signatures

The CIBERSORT algorithm was used to evaluate the abundance of immune cell populations in all samples, and then differences in the proportion of individual immune cells were compared between NASH and healthy controls. The Cor function in R 3.6.1 was used to compute the Spearman correlation coefficient between DEG signatures and the abundance of immune cells. Finally, DEG signatures with an adjusted P of < 0.05 and |Spearman R| > 0.6 were used to generate a scatter plot.

3.7. Statistical Analysis

Gene expression comparison between NASH and control samples was performed using the limma package with a false discovery rate (FDR) of < 0.05 and $|log_2$ fold change (FC)| > 0.263, as thresholds. Also, LASSO regression analysis was performed to screen DEG signatures. The Wilcoxon test was used to compare the difference in immune cell infiltration between NASH and normal samples, and the Spearman correlation coefficient was calculated to assess the relationship between DEG signatures and immune cell infiltration. Statistical significance was set at P < 0.05, and P values in module and correlation analyses were adjusted using the Benjamini & Hochberg (BH) method.

4. Results

4.1. Comparison of DEGs Between NASH and Healthy Control Samples

According to expression profile data and the threshold set, we obtained 3194 upregulated and 2444 downregulated genes at GSE89632 and 2703 upregulated and 2039 downregulated genes at E-MEXP-3291 comparing NASH and healthy control samples considering FDR < 0.05 and $|log_2|$ (FC) > 0.263. By intersecting the DEGs obtained from the two datasets, 403 upregulated genes (Figure 2A) and 158 downregulated genes (Figure 2B) were further screened. We further removed the batch effect of samples from the GSE89632 and E-MEXP-3291 datasets using the sva package. Appendix 1 shows PCA diagrams before and after the removal of the batch effect in the two datasets. The results suggested that the sample distribution of the two datasets was not significantly different after the removal of the batch effect. Thereafter, expression data were extracted to generate a sample clustering heatmap, as shown in Figure 2C. The diagram illustrates that the expression levels of DEGs were significantly different between NASH and normal samples, and DEGs were identified based on their respective expression profiles.

4.2. Functional and Pathway Enrichment Analysis on DEGs

In this regard, GO-BP functional analysis and KEGG pathway enrichment analysis were performed based on P < 0.05 and gene count > 2, leading to the emergence of 69 GO-BP and 36 KEGG pathways for upregulated DEGs and 22 GO-BP and three KEGG pathways for downregulated DEGs. According to the ranking of P values, the top 20 pathways of GO-BP and KEGG were selected (Figure 3A and 3B, respectively). Enrichment results showed that downregulated DEGs were more likely to participate in the biological processes involved in the inflammatory and innate immune responses whereas upregulated DEGs were more engaged with signal transduction, positive regulation of GTPase activity, and cell adhesion. Of the enriched KEGG pathways, upregulated DEGs were more engaged in the pathways involved in cancer, the PI3K-Akt signaling pathway, and focal adhesion while downregulated DEGs were most likely to participate in metabolic pathways.

4.3. Analysis Using Constructed PPI Network

We further established a PPI network based on intersecting DEGs and obtained 555 pairs of interactions containing 246 genes (Figure 4A). Then 12 DEGs (*SYK*, *PIK*3CG, *PLCG1*, *SOS1*, *CASP3*, *CD2*, *CDH1*, *PAK1*, *CCL5*, *DDX58*, *CD3G*, and *STAT3*), which ranked as the top 20 based on the topological properties of degree, betweenness, and closeness, were identified as hub genes in the PPI network (Figure 4B). We also performed module mining analysis on the PPI network and identified 17 submodules and their corresponding genes (Appendix 2). Accordingly, KEGG pathway analysis revealed that only the genes in 11 submodules were significantly enriched in KEGG pathways, and these enriched pathways varied in each submodule. The results of the top five pathways with BH-adjusted P values in each submodule have been displayed in Figure 4C.

4.4. Screening and Validation of DEG Signatures

Twelve hub genes in the PPI network were included in the gene set from the 11 submodules identified. Therefore, the expression data of candidate genes from these submodules were extracted to recognize optimized DEG combinations. After LASSO regression analysis, 17 DEG signatures were identified (Figure 5A), and their expressions were validated in the normal, SS, and NASH samples from the E-MEXP-3291 (Figure 5B) and GSE89632 (Figure 5C) datasets. The results suggested that these gene signatures were differentially expressed between healthy and NASH samples in the two datasets, which was statistically significant. In addition, we found that in the E-MEXP-3291 dataset, except for VAMP3, ACVR2B, ACTN1, TNFSF10, and SOS1, the other 12 gene signatures were differentially expressed between SS and NASH samples while TNFSF10 and CASP1 were differentially expressed between the normal and SS groups. In the GSE89632 dataset, CXCL10, CXCL9, CDH1, CASP1, MRAS, and SOS1 were differentially expressed between SS and NASH samples (P < 0.05) while 12 of the 17-DEG signatures identified were differentially expressed between the normal and SS groups (P < 0.05). The expression of the 17-gene-signature varied progressively from normal to SS and then to NASH in these two datasets. Subsequently, the expression data of the 17-DEG-signature were extracted from the GSE63067, GSE107231, and GSE72756 datasets for validation (Appendix 3). The results suggested that the expression of CASP1 was significantly different between NASH and control samples in the GSE63067 dataset while MRAS was significantly upregulated in NASH compared with control in the GSE107231 dataset. Finally, in the GSE72756 dataset, the expression levels of ACVR2B, CXCL9, and CXCL10 were significantly different between NASH and healthy control samples.

4.5. Immune Infiltration Analysis of DEG Signatures

We further evaluated the abundance of 22 types of tissue-infiltrating immune cells in the combined dataset (Figure 6A). We finally filtered eight NASH samples and nine healthy controls with a standard of P < 0.05 for the following analysis. Furthermore, the Wilcoxon test was performed to compare the abundance of each immune



Figure 2. Screening of differentially expressed genes (DEGs) between non-alcoholic steatohepatitis and normal samples in the E-MEXP-3291 and GSE89632 datasets. A-B. Venn diagrams showing commonly upregulated (A) and downregulated (B) genes from the GSE89632 and E-MEXP-3291 datasets. The DEGs were selected with FRD < 0.05 and |log₂ fold change (FC)| > 0.263 as thresholds. C. The heatmap shows the differences in the expression of DEGs between NASH and normal samples from the GSE89632 and E-MEXP-3291 datasets.

cell type between the two groups. Overall, seven types of immune cells showed significant differences between the two groups (Figure 6B), showing higher numbers of resting memory CD4 T cells, M2 macrophages, and resting dendritic cells while lower numbers of activated NK cells, monocytes, activated dendritic cells, and neutrophils in NASH than in normal tissues. Next, we analyzed the correlation between the frequency of these seven immune cell types and the 17-DEG-signature and identified 28 pairs of relationships between these cells and genes at adjusted P< 0.05 and |Spearman R| > 0.6 (Table 1). The results suggested that a total of 12 DEG signatures correlated with six types of immune cells, including activated dendritic cells, resting dendritic cells, M2 macrophages, monocytes, neutrophils, and resting memory CD4 T cells. Considering relationships between DEG signatures and immune cells, activated dendritic cells positively correlated with *AVPR1A* expression; M2 macrophages positively correlated with *VAMP3* expression; neutrophils positively correlated with *SMAD1*, *AVPR1A*, *CXCL2*, and *CXCR1* expression, and finally, resting memory CD4 T cells were positively associated with *CXCL10*, *SYK*, *CDH1*, and *FGFR2* expression. All other relationships were negative correlations.

5. Discussion

The onset of NASH is usually accompanied by lipid accumulation, liver cell damage, immune system dysfunction, and fibrosis, which may eventually lead to HCC (24). Therefore, the identification of biomarkers for the early diagnosis of NASH is of great significance in controlling the inflammatory response and the prevention of disease de-



Figure 3. Gene ontology-biological process (GO-BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment of differentially expressed genes (DEGs) between normal and NASH samples. The top 20 GO-BP (A) and KEGG pathways (B) for both upregulated and downregulated DEGs ranked by P values. Significantly enriched terms were selected based on the criteria of P < 0.05 and gene count ≥ 2 . The x-axis indicates the gene count while the y-axis indicates the terms of the GO-BP and KEGG pathways with a significant correlation. The triangles and circles in B represent upregulated and downregulated DEGs, respectively.

terioration in patients with NASH. To explore the feature genes of NASH, we first identified 403 upregulated and 158 downregulated DEGs based on the analysis of 35 NASH samples and 43 healthy controls. Furthermore, we screened candidate genes by the PPI network-based module and LASSO regression analyses, resulting in the identification of 17 DEG signatures, including CXCL2, CXCL10, CXCR1, CXCL9, VAMP3, SMAD1, ACVR2B, CDH1, SYK, AVPR1A, ACTN1, GAS6, TN-FSF10, CASP1, MRAS, FGFR2, and SOS1. Finally, by evaluating the relationship between the 17-DEG-signature and immune cell abundance in NASH and control samples, 12 DEG signatures were found to significantly correlate with six types of immune cells, including M2 macrophages and neutrophils. These results provide novel insights into the immune-regulatory mechanisms of NASH.

By performing enrichment analysis, we found that 561

DEGs in NASH were mainly enriched in the inflammatory response, innate immune response, signal transduction, cancer-related pathways, PI3K-Akt signaling pathway, and focal adhesion. Hepatic steatosis could trigger the release of stress signals from hepatocytes, thereby activating inflammatory pathways. On the other hand, the pattern recognition receptors expressed on the surface of cells can sense cellular damage and pathogen invasion and activate the innate immune response in NASH (25). Secretome gene analysis revealed a highly linked network of intrahepatic signals in NASH, and the signal transduction triggered by TGF- β -Smad3 was also shown to be involved in diet-induced NASH in mice (26, 27). Liu et al. pointed out that the inhibition of PI3K/AKT/mTOR pathways could enhance the autophagic flux of macrophages and thus inhibit the inflammatory response in mice with NASH (28).



Figure 4. Construction and module mining analysis of the protein-protein interaction (PPI) network. A, Establishment of a PPI network based on intersected differentially expressed genes (DEGs). This PPI network contains 246 genes (upregulated DEGs are shown in orange, and downregulated DEGs are shown in blue) and 555 pairs of interactions. B, The Venn diagram shows the top 20 genes ranked by gene count in terms of the topological properties of degree, betweenness, and closeness, and the intersected 12 DEGs were identified as hub genes in the PPI network. C, The heatmap shows the top five KEGG pathways significantly enriched in DEGs, ranked by Pvalue in each submodule.

The activation of IRS2/PI3K/Akt signal transduction could also increase hepatic glycogen storage and improve insulin resistance while the dysregulation of the insulinrelated PI3K/AKT-p70S6K pathway was involved in fibrosis development in NASH (29, 30). Additionally, Dattaroy et al. observed changes in the expression of focal adhesion proteins upon SsnB administration in animal models of liver fibrosis (31). These findings explain the possible involvement of enriched pathways and biological processes in disease regulation, thus further demonstrating that the obtained DEGs have certain pathogenic characteristics and are involved in NASH-related inflammatory and innate immune responses.

In this study, we highlighted the significance of the 17-DEG-signature in the development and progression of NASH and verified their expression by internal and external validation. By comparing the differences in the expression of these DEGs between normal, SS, and NASH samples, we found that *CASP1* was significantly upregulated in the SS and NASH groups. Dixon et al. showed that caspase-



Figure 5. Screening of differentially expressed gene (DEG) signatures using least absolute shrinkage and selection operator (LASSO) regression analysis and the validation of their expression in normal, simple steatosis (SS), and non-alcoholic steatohepatitis (NASH) samples from the E-MEXP-3291 and GSE89632 datasets. A. LASSO regression analysis was used to identify 17 DEG signatures as the optimized gene set. B. The expression differences of 17 DEG signatures between normal, SS, and NASH samples from the E-MEXP-3291 dataset. C. The expression differences of 17 DEG signatures between the normal, SS, and NASH groups in the GSE89632 dataset.

1 (*CASP1*)-deficient mice might be less prone to high fatinduced SS and inflammation associated with NASH (32). This finding explains our results and further suggests that *CASP1* is a risk factor for SS development and then its progression to NASH. Additionally, we identified several members of the CXC chemokine family as potential biomarkers for NASH, including *CXCL2*, *CXCL9*, *CXCL10*, and *CXCR1*. The role of the CXC chemokine family in NASH has been widely studied, and the hepatic infiltration of neutrophils and upregulation of *CXCL1* are known to be involved in NASH (33). Another member of this family, *CXCR3*, plays an important role in the development of NASH by inducing cytokine production and macrophage infiltration (34). In addition, Semba et al. found that hepatocytes and sinusoidal endothelium, which express *CXCL9*, were confined to areas of inflammatory cell infiltration in mice with NASH (35). Our study further found that the expression of *CXCL9* was significantly elevated during SS progression to NASH, suggesting that the overexpression of *CXCL9* may correlate with the activation of inflammatory responses in this condition. Regarding *CXCL10*, there was an increasing trend in the expression of this gene from normal to SS to NASH. A related study confirmed our results and reported that *CXCL10* was upregulated in NASH, and that the level of circulating *CXCL10* was associated with lobular inflammatory mediator of NASH (36). These data illustrated that the DEG signatures identified here could be important therapeutic targets for NASH, but their integrative effects on the disease pathogenesis remain to be explored.



Figure 6. Immune cell infiltration analysis in normal and non-alcoholic steatohepatitis (NASH) samples in the combined dataset. A. Histograms show the infiltration abundance of 22 types of immune cells in eight NASH samples and nine normal controls from the combined dataset (E-MEXP-3291 and GSE89632). B. The violin diagram shows the differences in the abundance of 22 types of immune cells between the NASH and control groups. Blue and red bars indicate normal and NASH samples, respectively.

We found that NASH and normal samples had significant differences in the abundance of immune cells. Among these cells, neutrophils showed the greatest difference between the two groups, and NASH samples showed enhanced infiltration of neutrophils. van der Windt et al. found early neutrophilic infiltration, macrophage influx, and production of inflammatory cytokines in mice with high fat diet-induced NASH (37). Wu et al. also proposed that enhanced neutrophil infiltration was a major histological feature of NASH (38). Furthermore, the overexpression of *CXCL1* in the liver may drive NASH progression through facilitating the activation of stress kinases and neutrophil infiltration (39). In addition to neutrophilrelated DEGs, we also identified 12 DEG signatures related

Table 1. Significant Correlations Between the 17-DEG-Signatures and Infiltrations by Immune Cells				
Immune Cell	Gene	R ^a	P-Value	Q-Value ^b
Activated dendritic cells	SYK	-0.697731584	0.001844931	0.01833562
Activated dendritic cells	CXCL10	-0.684242924	0.002449187	0.01943022
Activated dendritic cells	VAMP3	-0.659718088	0.003958514	0.024026602
Activated dendritic cells	MRAS	-0.619252109	0.00802977	0.039814276
Activated dendritic cells	AVPR1A	0.70141031	0.00170334	0.01833562
Resting dendritic cells	CXCR1	-0.760303501	0.000396191	0.009438255
Resting dendritic cells	CXCL2	-0.666025867	0.003512852	0.023302488
M2 macrophages	ACVR2B	-0.617647059	0.009698737	0.046165989
M2 macrophages	CXCL2	-0.607843137	0.011192078	0.046651756
M2 macrophages	VAMP3	0.644607843	0.006395802	0.034595476
Monocytes	FGFR2	-0.637254902	0.007189792	0.037199359
Monocytes	SOS1	-0.607843137	0.011192078	0.046651756
Neutrophils	FGFR2	-0.760269916	0.000396565	0.009438255
Neutrophils	CDH1	-0.727161387	0.000941323	0.016002484
Neutrophils	SYK	-0.703862793	0.001613999	0.01833562
Neutrophils	VAMP3	-0.669528023	0.003283597	0.023302488
Neutrophils	CXCL10	-0.665849297	0.003524746	0.023302488
Neutrophils	MRAS	-0.602084724	0.010545631	0.046651756
Neutrophils	SMAD1	0.690374133	0.002157132	0.01833562
Neutrophils	AVPR1A	0.692826617	0.002048599	0.01833562
Neutrophils	CXCL2	0.7627224	0.000369997	0.009438255
Neutrophils	CXCR1	0.902513965	7.20E-07	8.57E-05
Resting memory CD4 T cells	AVPR1A	-0.710784314	0.001926552	0.01833562
Resting memory CD4 T cells	CXCR1	-0.669117647	0.004239989	0.024026602
Resting memory CD4 T cells	CXCL10	0.669117647	0.004239989	0.024026602
Resting memory CD4 T cells	SYK	0.713235294	0.001831279	0.01833562
Resting memory CD4 T cells	CDH1	0.75245098	0.00074371	0.014750255
Resting memory CD4 T cells	FGFR2	0.801470588	0.000158696	0.009438255

Abbreviation: DEG, differentially expressed gene.

^a R stands for Spearman correlation coefficient.

 $^{\rm b}$ Q-value represents P-value adjusted by the Benjamini & Hochberg method, and Q < 0.05 indicates statistical significance.

to dendritic cells, M2 macrophages, monocytes, and memory CD4 T cells in NASH. Many of these relationships between DEG signatures and immune cells were initially found to be involved in the development of NASH. Related studies have reported that the loss of p38 α expression in macrophages can ameliorate steatohepatitis in mice through reducing the secretion of proinflammatory cytokines, including TNF- α , *CXCL10*, and IL-6 (40). Moreover, the immune infiltration of macrophages plays a crucial role in the CXCL10-mediated inflammatory response in mice with NASH (41). In this study, we found an increased expression of CXCL10, as well as decreased infiltration of M2 macrophages in NASH samples. Combined with the above findings, we hypothesized that CXCL10 may activate the inflammatory response by regulating macrophage polarization. However, the current understanding of cell-gene interactions that may stimulate the NASH-related inflammatory response is limited, and our study provided a systematic network of interactive relations as the theoretical basis for future research.

In the current study; however, the expression of these 17 gene signatures was not validated in the tissues of pa-

tients with NASH due to the lack of clinical biopsy samples. Furthermore, the lack of the experimental verification of the relationships between DEG signatures and immune cell populations was another limitation of this study. Additionally, the sample size in this study was not large enough to fully explain the key roles of these gene signatures in the development of NASH. Therefore, in future studies, we will expand the sample size to validate the significance of these 17 DEG signatures in NASH and further investigate their regulatory mechanisms in triggering immune responses in NASH.

5.1. Conclusions

We identified 561 DEGs in NASH tissue samples and explored their biological functions and related pathways involved in the development of NASH. The 17-DEG-signature was further identified from candidate genes in the PPI network and related submodules. Finally, 12 DEG signatures were found to be associated with activated dendritic cells, resting dendritic cells, M2 macrophages, monocytes, neutrophils, and resting memory CD4 T cells and probably govern the increase of these immune cells in NASH samples compared to healthy controls. Our results identified valuable biomarkers in NASH, which can role as potential therapeutic targets regulating the mechanisms underlying NASH-related inflammatory responses.

Supplementary Material

Supplementary material(s) is available here [To read supplementary materials, please refer to the journal website and open PDF/HTML].

Footnotes

Authors' Contribution: Conception and design of the research: Junwang Zhang and Zhihua Deng; Acquisition of data: Huimin Wang and Ling Zhang; Analysis and interpretation of data: Xumei Kang and Junwang Zhang; Statistical analysis: Ling Zhang and Huimin Wang; Funding acquisition: Junwang Zhang; Drafting the manuscript: Xumei Kang and Junwang Zhang; Revision of manuscript for important intellectual content: Zhihua Deng.

Conflict of Interests: The authors declare that they have no conflict of interest.

Ethical Approval: This study analyzed the expression profile datasets obtained from public databases, and no animal or human experiments were involved. Therefore, no ethical review or informed consent was needed.

Funding/Support: This study was supported by the Natural Science Foundation of Shanxi Province (No. 201801D121327).

References

- Min-DeBartolo J, Schlerman F, Akare S, Wang J, McMahon J, Zhan Y, et al. Thrombospondin-I is a critical modulator in non-alcoholic steatohepatitis (NASH). *PLoS One.* 2019;**14**(12). e0226854. doi: 10.1371/journal.pone.0226854. [PubMed: 31891606]. [PubMed Central: PMC6938381].
- Xiong X, Wang Q, Wang S, Zhang J, Liu T, Guo L, et al. Mapping the molecular signatures of diet-induced NASH and its regulation by the hepatokine Tsukushi. *Mol Metab.* 2019;**20**:128–37. doi: 10.1016/ji.molmet.2018.12.004. [PubMed: 30595550]. [PubMed Central: PMC6358550].
- Ye J, Lin Y, Wang Q, Li Y, Zhao Y, Chen L, et al. Integrated Multichip Analysis Identifies Potential Key Genes in the Pathogenesis of Nonalcoholic Steatohepatitis. *Front Endocrinol (Lausanne)*. 2020;**11**:601745. doi: 10.3389/fendo.2020.601745. [PubMed: 33324350]. [PubMed Central: PMC7726207].
- Baciu C, Pasini E, Angeli M, Schwenger K, Afrin J, Humar A, et al. Systematic integrative analysis of gene expression identifies HNF4A as the central gene in pathogenesis of non-alcoholic steatohepatitis. *PLoS One*. 2017;**12**(12). e0189223. doi: 10.1371/journal.pone.0189223. [PubMed: 29216278]. [PubMed Central: PMC5720788].
- Younossi Z, Tacke F, Arrese M, Chander Sharma B, Mostafa I, Bugianesi E, et al. Global Perspectives on Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis. *Hepatology*. 2019;69(6):2672– 82. doi: 10.1002/hep.30251. [PubMed: 30179269].
- Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. *Hepatology*. 2018;67(1):328–57. doi: 10.1002/hep.29367. [PubMed: 28714183].
- Alonso C, Fernandez-Ramos D, Varela-Rey M, Martinez-Arranz I, Navasa N, Van Liempd SM, et al. Metabolomic Identification of Subtypes of Nonalcoholic Steatohepatitis. *Gastroenterology*. 2017;**152**(6):1449-1461 e7. doi: 10.1053/j.gastro.2017.01.015. [PubMed: 28132890]. [PubMed Central: PMC5406239].
- Dongiovanni P, Meroni M, Longo M, Fargion S, Fracanzani AL. miRNA Signature in NAFLD: A Turning Point for a Non-Invasive Diagnosis. *Int J Mol Sci.* 2018;**19**(12). doi: 10.3390/ijms19123966. [PubMed: 30544653]. [PubMed Central: PMC6320931].
- Luci C, Bourinet M, Leclere PS, Anty R, Gual P. Chronic Inflammation in Non-Alcoholic Steatohepatitis: Molecular Mechanisms and Therapeutic Strategies. *Front Endocrinol (Lausanne)*. 2020;**11**:597648. doi: 10.3389/fendo.2020.597648. [PubMed: 33384662]. [PubMed Central: PMC7771356].
- Hammoutene A, Rautou PE. Role of liver sinusoidal endothelial cells in non-alcoholic fatty liver disease. J Hepatol. 2019;70(6):1278–91. doi: 10.1016/j.jhep.2019.02.012. [PubMed: 30797053].
- Chen Y, Tian Z. Roles of Hepatic Innate and Innate-Like Lymphocytes in Nonalcoholic Steatohepatitis. *Front Immunol.* 2020;**11**:1500. doi: 10.3389/fimmu.2020.01500. [PubMed: 32765518]. [PubMed Central: PMC7378363].
- Cai J, Zhang XJ, Li H. The Role of Innate Immune Cells in Nonalcoholic Steatohepatitis. *Hepatology*. 2019;**70**(3):1026–37. doi: 10.1002/hep.30506. [PubMed: 30653691].
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data setsupdate. *Nucleic Acids Res.* 2013;41(Database issue):D991–5. doi: 10.1093/nar/gks1193. [PubMed: 23193258].
- Brazma A, Parkinson H, ArrayExpress team E. ArrayExpress service for reviewers/editors of DNA microarray papers. *Nat Biotechnol.* 2006;24(11):1321-2. doi: 10.1038/nbt1106-1321. [PubMed: 17093465].

- 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(7). e47. doi: 10.1093/nar/gkv007. [PubMed: 25605792]. [PubMed Central: PMC4402510].
- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882–3. doi: 10.1093/bioinformatics/bts034. [PubMed: 22257669].
- Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44–57. doi: 10.1038/nprot.2008.211. [PubMed: 19131956].
- Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;**37**(1):1–13. doi: 10.1093/nar/gkn923. [PubMed: 19033363].
- Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 2017;**45**(D1):D362–8. doi: 10.1093/nar/gkw937. [PubMed: 27924014]. [PubMed Central: PMC5210637].
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-504. doi: 10.1101/gr.1239303. [PubMed: 14597658]. [PubMed Central: PMC403769].
- Tang Y, Li M, Wang J, Pan Y, Wu FX. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems*. 2015;**127**:67–72. doi: 10.1016/j.biosystems.2014.11.005. [PubMed: 25451770].
- Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003;4:2. doi: 10.1186/1471-2105-4-2. [PubMed: 12525261]. [PubMed Central: PMC149346].
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7. doi: 10.1089/omi.2011.0118. [PubMed: 22455463].
- Zeng F, Zhang Y, Han X, Zeng M, Gao Y, Weng J. Predicting Non-Alcoholic Fatty Liver Disease Progression and Immune Deregulations by Specific Gene Expression Patterns. *Front Immunol.* 2020;**11**:609900. doi: 10.3389/fimmu.2020.609900. [PubMed: 33574818]. [PubMed Central: PMC7870871].
- Arrese M, Cabrera D, Kalergis AM, Feldstein AE. Innate Immunity and Inflammation in NAFLD/NASH. *Dig Dis Sci.* 2016;61(5):1294–303. doi: 10.1007/s10620-016-4049-x. [PubMed: 26841783]. [PubMed Central: PMC4948286].
- Xiong X, Kuang H, Ansari S, Liu T, Gong J, Wang S, et al. Landscape of Intercellular Crosstalk in Healthy and NASH Liver Revealed by Single-Cell Secretome Gene Analysis. *Mol Cell*. 2019;75(3):644–660 e5. doi: 10.1016/j.molcel.2019.07.028. [PubMed: 31398325]. [PubMed Central: PMC7262680].
- Huang M, Kim HG, Zhong X, Dong C, Zhang B, Fang Z, et al. Sestrin 3 Protects Against Diet-Induced Nonalcoholic Steatohepatitis in Mice Through Suppression of Transforming Growth Factor beta Signal Transduction. *Hepatology*. 2020;71(1):76–92. doi: 10.1002/hep.30820. [PubMed: 31215672]. [PubMed Central: PMC6920605].
- 28. Liu B, Deng X, Jiang Q, Li G, Zhang J, Zhang N, et al. Scoparone improves hepatic inflammation and autophagy in mice with nonalcoholic steatohepatitis by regulating the ROS/P38/Nrf2 axis and PI3K/AKT/mTOR pathway in macrophages. *Biomed Pharmacother*. 2020;**125**:109895. doi: 10.1016/j.biopha.2020.109895. [PubMed:

32000066].

- Cai CX, Buddha H, Castelino-Prabhu S, Zhang Z, Britton RS, Bacon BR, et al. Activation of Insulin-PI3K/Akt-p70S6K Pathway in Hepatic Stellate Cells Contributes to Fibrosis in Nonalcoholic Steatohepatitis. *Dig Dis Sci.* 2017;62(4):968–78. doi: 10.1007/s10620-017-4470-9. [PubMed: 28194671].
- Xu H, Zhou Y, Liu Y, Ping J, Shou Q, Chen F, et al. Metformin improves hepatic IRS2/PI3K/Akt signaling in insulin-resistant rats of NASH and cirrhosis. J Endocrinol. 2016;229(2):133–44. doi: 10.1530/JOE-15-0409. [PubMed: 26941037].
- Dattaroy D, Seth RK, Sarkar S, Kimono D, Albadrani M, Chandrashekaran V, et al. Sparstolonin B (SsnB) attenuates liver fibrosis via a parallel conjugate pathway involving P53-P21 axis, TGF-beta signaling and focal adhesion that is TLR4 dependent. *Eur J Pharmacol.* 2018;**841**:33–48. doi: 10.1016/j.ejphar.2018.08.040. [PubMed: 30194936]. [PubMed Central: PMC7193950].
- Dixon LJ, Flask CA, Papouchado BG, Feldstein AE, Nagy LE. Caspase-1 as a central regulator of high fat diet-induced non-alcoholic steatohepatitis. *PLoS One*. 2013;8(2). e56100. doi: 10.1371/journal.pone.0056100. [PubMed: 23409132].
- Hwang S, Wang X, Rodrigues RM, Ma J, He Y, Seo W, et al. Protective and Detrimental Roles of p38alpha Mitogen-Activated Protein Kinase in Different Stages of Nonalcoholic Fatty Liver Disease. *Hepatology*. 2020;**72**(3):873–91. doi: 10.1002/hep.31390. [PubMed: 32463484]. [PubMed Central: PMC7704563].
- Zhang X, Han J, Man K, Li X, Du J, Chu ES, et al. CXC chemokine receptor 3 promotes steatohepatitis in mice through mediating inflammatory cytokines, macrophages and autophagy. *J Hepatol.* 2016;64(1):160–70. doi: 10.1016/j.j.jhep.2015.09.005. [PubMed: 26394162].
- Semba T, Nishimura M, Nishimura S, Ohara O, Ishige T, Ohno S, et al. The FLS (fatty liver Shionogi) mouse reveals local expressions of lipocalin-2, CXCL1 and CXCL9 in the liver with non-alcoholic steatohepatitis. *BMC Gastroenterol.* 2013;13:120. doi: 10.1186/1471-230X-13-120. [PubMed: 23875831].
- Zhang X, Shen J, Man K, Chu ES, Yau TO, Sung JC, et al. CXCL10 plays a key role as an inflammatory mediator and a non-invasive biomarker of non-alcoholic steatohepatitis. *J Hepatol.* 2014;61(6):1365–75. doi: 10.1016/j.jhep.2014.07.006. [PubMed: 25048951].
- van der Windt DJ, Sud V, Zhang H, Varley PR, Goswami J, Yazdani HO, et al. Neutrophil extracellular traps promote inflammation and development of hepatocellular carcinoma in nonalcoholic steatohepatitis. *Hepatology*. 2018;68(4):1347–60. doi: 10.1002/hep.29914. [PubMed: 29631332]. [PubMed Central: PMC6173613].
- 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(5):652–66. doi:10.1007/s12072-020-10081-7. [PubMed: 32880077].
- Hwang S, He Y, Xiang X, Seo W, Kim SJ, Ma J, et al. Interleukin-22 Ameliorates Neutrophil-Driven Nonalcoholic Steatohepatitis Through Multiple Targets. *Hepatology*. 2020;**72**(2):412–29. doi: 10.1002/hep.31031. [PubMed: 31705800]. [PubMed Central: PMC7210045].
- Zhang X, Fan L, Wu J, Xu H, Leung WY, Fu K, et al. Macrophage p38alpha promotes nutritional steatohepatitis through M1 polarization. *J Hepatol.* 2019;**71**(1):163–74. doi: 10.1016/j.jhep.2019.03.014. [PubMed: 30914267].
- Tomita K, Freeman BL, Bronk SF, LeBrasseur NK, White TA, Hirsova P, et al. CXCL10-Mediates Macrophage, but not Other Innate Immune Cells-Associated Inflammation in Murine Nonalcoholic Steatohepatitis. *Sci Rep.* 2016;**6**:28786. doi: 10.1038/srep28786. [PubMed: 27349927]. [PubMed Central: PMC4923862].