

# **Cellular and Molecular Biology**

E-ISSN: 1165-158X / P-ISSN: 0145-5680

www.cellmolbiol.org



# Identification of biomarkers associated with immune-propionate metabolism in nonal-

# coholic fatty liver disease

Shuye Qu<sup>1,2#</sup>, Hui Huang<sup>1,2#</sup>, Yan Diao<sup>3#</sup>, Bowei Liu<sup>1,2</sup>, Baozhu Tang<sup>1,2</sup>, Shijiao Huo<sup>1,2</sup>, Yu Lei<sup>1,2</sup>, Xiuchen Xuan<sup>1,2</sup>, Wenling Mou<sup>4</sup>, Ping Liu<sup>4</sup>, Jiye Zhang<sup>4</sup>, Ying Liu<sup>4,\*</sup>, Yanze Li<sup>1,2</sup>

<sup>1</sup> Department of Biochemistry and Molecular Biology, Harbin Medical University, Harbin, China
<sup>2</sup> Translational Medicine Center of Northern China, Harbin, China
<sup>3</sup> Department of Clinical Laboratory, Heilongjiang Province Hospital, Harbin, China
<sup>4</sup> Department of Gastroenterology, Heilongjiang Provincial Hospital, Harbin, China

<sup>#</sup>These authors contributed equally to this work.

ARTICLE INFO	ABSTRACT
Original paper	The mechanisms of the effect of propionate metabolism and immunity on nonalcoholic fatty liver disease
	(NAFLD) have not been adequately studied. Firstly, differentially expressed-propionate metabolism-related
Article history:	genes (DE-PMRGs) were selected by overlapping PMRGs and differentially expressed genes (DEGs) between
Received: August 10, 2023	the simple steatosis (SS) and health control (HC) groups. Then, common genes were selected by overlapping
Accepted: September 22, 2023	DE-PMRGs and key module genes obtained from weighted gene co-expression network analysis (WGCNA).
Published: October 31, 2023	Subsequently, the biomarkers were screened out by machine learning algorithms. The expression of the bio-
Keywords:	markers was validated by quantitative Real-time PCR. In total, 5 biomarkers (JUN, LDLR, CXCR4, NNMT,
	and ANXA1) were acquired. The nomogram constructed based on 5 biomarkers had good predictive power
	for the risk of SS. Next, 5 biomarkers, 11 miRNAs, and 149 lncRNAs were encompassed in the ceRNA regu-
Non-alcoholic fatty liver disease;	latory network. The expression of biomarkers was significantly higher in the HC group than in the SS group,
propionate metabolism; biomar-	which was consistent with the results in the GSE89632 and GSE126848 datasets. In this study, 5 immune and
ker; diagnosis	propionate metabolism-related biomarkers (JUN, LDLR, CXCR4, NNMT, and ANXA1) were screened out to
	provide a basis for exploring the prediction of diagnosis of NAFLD.

Doi: http://dx.doi.org/10.14715/cmb/2023.69.10.38

Copyright: © 2023 by the C.M.B. Association. All rights reserved.

#### Introduction

Non-alcoholic fatty liver disease (NAFLD) refers to hepatic steatosis caused by fat accumulation in the liver that is not caused by viral infection, alcohol consumption, or genetic factors (1). The main pathological feature of NAFLD is the accumulation of triglycerides and free cholesterol (2). Its pathogenesis is related to a variety of factors, such as altered intestinal flora, obesity, and disturbances in body lipid homeostasis (3). NAFLD is not only associated with 25-40% of cases of elevated liver enzymes but also with the development of cardiovascular disease and type 2 diabetes (1, 4). As the disease progresses, NAFLD can lead to non-alcoholic steatohepatitis, cirrhosis and even the development of liver cancer (5). The prevalence of NAFLD in the world is currently about 30% and is expected to increase in the coming years (6). Liver biopsy is the gold standard for the current diagnosis and prognosis of NAFLD, and the main treatment for NAFLD is weight loss and vitamin supplementation (7, 8). Moreover, to date, the available non-invasive or minimally invasive biomarkers are insufficient, so finding key diagnostic biomarkers in NAFLD is important for the diagnosis and targeted treatment of NAFLD.

Approximately 40% of NAFLD patients in Korea are non-obese or thin (9), therefore the occurrence of NAFLD may be related not only to physical dysfunction but also to other causes. The interaction between the intestine and the liver regulates the homeostasis between the organs of the body, which is important for the health of the individual (10). Acetic acid, propionic acid and butyric acid are produced by intestinal microorganisms, and reduced levels of these short-chain fatty acids can trigger NAFLD by promoting gluconeogenesis and inflammation (11-13). Among these, propionate is the main product of dietary fiber fermentation in the colon, which is thought to reduce adipogenesis, serum cholesterol levels, and other tissue carcinogenic effects (14). It was found through repeated observations that when animals were fed fermentable fibers, they were prevented from steatosis induced by a high-fat diet, probably because propionate when absorbed from the portal vein, can alter hepatic metabolic processes to reduce lipid content (15). In addition, it has been found that propionate may reduce atherosclerosis through immune-dependent regulation of intestinal cholesterol metabolism (16). Non-alcoholic steatohepatitis, as a progressive form of NAFLD, is marked by inflammation and hepatocellular damage (17). NAFLD itself is also conside-

<sup>\*</sup> Corresponding author. Email: anzi1975@sina.com; liyanzeqq@hotmail.com Cellular and Molecular Biology, 2023, 69(10): 254-263

red to be a chronic inflammatory disease (18). However, the combined role of immune features and propionate in NAFLD has not been reported.

Therefore, in this study, based on the transcriptome sequencing data of NAFLD in GEO, five immune-propionate metabolism-related NAFLD diagnostic genes were screened by a series of bioinformatics tools such as PPI network and WGCNA analysis. In addition, we created a ceRNA network of biomarkers and performed drug prediction of diagnostic genes. It provides an important reference for the study of the regulatory mechanism of immune and propionate metabolism-related genes in NAFLD.

## **Materials and Methods**

## Source of data

The GSE89632 and GSE126848 datasets were sourced from the GEO database. The GSE89632 dataset includes 20 simple steatosis (SS) samples and 24 health control (HC) samples, and the GSE126848 dataset includes 15 SS samples and 14 HC samples. In this study, GSE89632 was utilized as the training set and GSE126848 was used as the validation set. Then, 394 propionate metabolism-related genes (PMRGs) were obtained from the GeneCards database.

# Identification of DE-PMRGs and construction of PPI networks

Differentially expressed genes1 (DEGs1) and DEGs2 between the SS and HC groups were selected in the GSE89632 and GSE126848 datasets with P value < 0.05and  $|\log_{2}FC| > 0.5$ , respectively. The differential analysis was carried out by using the limma package (v 3.42.2) (19) and DESeq2 package (v 1.34.0) (20), respectively. The results of the differential analysis were illustrated by volcano maps and heatmaps (Top 100) plotted by the ggplot2 package (v 3.3.6) (21) and the pheatmap package (v 1.0.12) respectively. Then, the DEGs were obtained by taking the intersection of DEGs1 and DEGs2. Finally, DE-PMRGs were screened out by overlapping PMRGs and DEGs. A PPI network had been created on the basis of DE-PMRGs via the STRING database (medium confidence = 0.4). Then, the visualization of the PPI network diagram was enabled by Cytoscape (v 3.6.1) (22).

## Functional enrichment analysis of DE-PMRGs

Gene Ontology (GO), REACTOME and the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of DE-PMRGs were executed via the g: Profiler online website.

## Immune cell infiltration analysis

To analyze the difference in immune infiltration degree between SS and HC groups in the GSE89632 dataset, the proportion of 20 immune cell subtypes (memory B cells and naive CD4 T cells content was 0, so it was not involved in) was calculated for each sample by the CIBERSORT algorithm (v 1.0.3) (23) (P value < 0.05). The results of the analysis were presented by the proportion and violin diagrams.

## Filtering for key module genes by WGCNA

The co-expression network was constructed by WGC-NA (v 1.69-3) (24) on the basis of the GSE89632 data-

set. Firstly, the samples were clustered and outliers were removed to ensure the accuracy of the analysis. Then, the optimal soft threshold ( $\beta$ ) was chosen to make the network approximate a scale-free distribution. In the following, the cluster dendrogram was gained via calculating adjacency and similarity. The modules were partitioned by a dynamic tree-cutting algorithm. Next, we evaluated the correlations between each module and differential immune cells and selected the module with |cor| > 0.5 and P values < 0.05 as the key module. Finally, the genes in the key module were identified as key module genes for follow-up analyses.

## Screening biomarkers by machine learning

Firstly, common genes were screened out by overlapping DE-PMRGs and key module genes. Three machine learning models were constructed by RF, SVM and GLM machine learning algorithms using the R package caret (v 6.0.93) (25). The performance analysis of the three models was carried out using the DALEX package (v 2.4.2) (26). The residual distribution of the models was plotted, and the one with the smallest residual was selected as the optimal model. Finally, the feature genes in the optimal model were selected as biomarkers for subsequent analysis.

## **Construction of the nomogram**

The nomogram was structured and visualized on the basis of biomarkers by rms package (v 5.1-4). Next, a calibration curve was plotted to judge the model prediction performance. At the same time, decision and impact curves were plotted to assess the predictive power of the model.

## **GSEA** functional enrichment analysis

GSEA functional enrichment analysis (GO and KEGG) was performed on biomarkers to screen significantly enriched pathways according to SIZE > 20 and NOM.p.val < 0.05. The pathway entries in which multiple biomarkers were involved together were presented.

## Construction of a ceRNA regulatory network

The miRWalk database was used to predict miRNAs targeting biomarkers (Score = 0.95, Position = 3UTR, bindingp  $\geq$  1). Meanwhile, the miRNet database was utilized for targeting miRNA prediction for biomarkers. The common miRNAs (co-miRNAs) were obtained by taking the intersection of the predicted miRNAs from the two databases. Next, the miRNet database and the ENCORI database were utilized to predict target lncRNAs targeting the co-miRNAs, respectively. The common lncRNAs (co-lncRNAs) were obtained by taking the intersection of the predicted lncRNAs from the two databases. Lastly, using Cytoscape (v 3.6.1) (27), the lncRNA-miRNA-mRNA network was constructed based on those miRNAs and lncRNAs.

## **Construction of biomarkers-drug interaction network**

The drugs targeting the biomarkers were predicted through the DGIDB database (https://dgidb.org). The results of the predictions were presented by the network and the Sankey diagram. Finally, the ceRNA network was combined with the drug prediction network to plot the lncRNA-miRNA-biomarker-drug network.

#### **Expression level validation**

The expression of biomarkers was extracted from the GSE89632 and GSE126848 datasets respectively, and their expression was compared and plotted by the Wilcoxon test method.

#### Animals

Male C57BL6 mice, aged 8 weeks, were used in this experiment. For each strain, 15 mice were grouped as control and 15 were used as a model, raised in an animal cage with saw dust bedding. Conditions of the mice house are as follows:  $23 \pm 2^{\circ}$ C,  $55 \pm 5\%$  relative humidity, 12 h light/dark cycle. All mice were fed with standard chow (26% kcal from fat, Shoobree, Jiangsu, China) for 1 week to acclimatize before the experiment started. Then, the control mice continued to be fed a standard diet, while the experimental group was changed to a high-fat diet (HFD), bought from Research diets (D12451), containing 45 kcal% fat. After 10 weeks of feeding, all mice were put to death. Mice were fasted overnight before the liver was harvested at the end of the experiment. All animal procedures were performed in accordance with the Guidelines for Care and Use of Laboratory Animals of Harbin Medical University and approved by the Animal Ethics Committee of Animal Welfare and Ethics Committee.

#### **Quantitative Real-time PCR**

Total RNA was purified by liver tissues of normal mouse models (HC group) and NAFLD mouse models (SS group) in TRIzol reagent (Invitrogen, Carlsbad, CA, USA). Complementary DNA (cDNA) was obtained from 1  $\mu$ g of total RNA using a High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA). Gene expression was quantified using the Applied Biosystems 7500 real-time PCR system. SYBR-Green Master Mix (Applied Biosystems, Foster City, CA, USA) and mRNA-specific primers were used (Table 1). Relative gene expression was calculated using the 2<sup>- $\Delta$ Ct</sup> formula standardized by the expression of housekeeping gene Actb.

## Statistical analysis

All bioinformatics analyses were undertaken in R language. At the same time, the Wilcoxon test was employed to contrast the data from different groups. All statistical tests were calculated using GraphPad Prism (5.0 version, La Jolla, CA, USA) more than three times. Data were expressed as mean  $\pm$  SEM. The significance of single comparisons was performed using t-test and Normality and Lognormality Tests. P-values less than 0.05 were considered statistically significant.

## Results

## Screening of DE-PMRGs and PPI analysis

In total, 3079 DEGs1 between the SS and HC groups

were gained, including 1393 up-regulated genes and 1686 down-regulated genes (Figure 1a-1b). Meanwhile, a total of 4516 DEGs2 between the SS and HC groups were gained, including 2324 up-regulated genes and 2192 down-regulated genes (Figure 1c-1d). In all, 241 common DEGs with the same expression trend between DEGs1 and DEGs2 were screened and then analyzed by Venn to obtain 17 DE-PMRGs (SULT1E1, NR0B2, SERPINE1, SOD2, CXCR4, LDLR, ALOX5AP, ANXA1, LCN2, NNMT, CXCL8, RRM2, SAA1, SPHK1, JUN, ANXA2, and CYP21A2) (Figure 1e-1f). A PPI network comprising 12 nodes and 19 edges was created (Figure 1g). Of these, only NR0B2 was up-regulated in expression in SS.

#### **Functional enrichment of DE-PMRGs**

The results of the enrichment analysis indicated that the DE-PMRGs implicated 66 GO-BP entries, 3 GO-CC entries, 8 GO-MF entries, 4 REACTOME and 4 KEGG



**Figure 1.** Differential expression analysis and PPI network. (a,b) The volcano map (a) and heat map (b) of DEGs between SS and HC samples in the GSE89632 dataset. (c,d) The volcano map (c) and heat map (d) of DEGs between SS and HC samples in the GSE126848 dataset. (e) The Venn diagram of common DEGs in two datasets. (f) The Venn diagram of DE-PMRGs obtained by overlapping DEG and PMRGs. (g) The PPI network of DE-PMRGs. PPI, protein-protein interaction; DEGs, differentially expressed genes; SS, simple steatosis; HC, health control; PMRGs, propionate metabolism-related genes; DE-PMRGs, differentially expressed PMRGs.

Table 1. Oligonucleotide sequence of primers used to determine the expression levels of mice genes by RT-QPCR.

Gene	Forward primer (5'-3')	Reverse primer (5'-3')
β-Actin	AGCACAATGAAGATCAAGATCATTGCTCC	ACTCGTCATACTCCTGCTTGCTGAT
ANXA1	TCCTCATCTTCGCAGAGTGTT	GGCAAAGAAAGCTGGAGTGC
CXCR4	TGTTGCCATGGAACCGATCA	TGGTGGGCAGGAAGATCCTA
LDLR	ACGGAGGTGACCAACAATAGAAT	TGATGGTGTCGTAGGACAAGTTAG
JUN	GGCTAGAGGAAAAAGTGAAAACCT	ATGACTTTCTGCTTAAGCTGTGC
NNMT	CGTGCAATCAAGCAGGAACC	CAGGGGAAGGCTGGAAAACT



**Figure 2.** Functional enrichment analysis for DE-PMRGs. (a,b) The GO terms were enriched in DE-PMRGs. (c,d) The KEGG pathways and REACTOME pathways were enriched in DE-PMRGs. GO, Gene Ontology; BP, biological progress; CC, cellular component; MF, molecular function; KEGG, Kyoto Encyclopedia of Genes and Genomes

pathways. The GO terms annotation indicated that DE-PMRGs primarily participated in vasculature development, inflammatory response, etc. (Figure 2a-2b). KEGG enrichment results included lipid and atherosclerosis, IL-17 signaling pathways, etc. (Figure 2c-2d). In addition, REACTOME enrichment results included signaling by interleukins, cytokine signaling in the immune system, etc. (Figure 2c-2d).

## Immune cell infiltration in SS and HC groups

The Proportion diagram indicated the proportion of the 20 immune cells in the SS and HC groups (Figure 3a-3b). The highest levels in both SS and HC groups were in resting memory CD4 T cells, while the second highest levels were in Macrophages M2 and naive B cells, respectively (Figure 3a-3b). In total, 8 immune cells (naive B cells, gamma delta T cells, Macrophages M2 and so on) were significantly different between the SS and HC groups (Figure 3c).

#### Identification of key module genes

To probe the genes associated with differential immune cells, we performed a WGCNA analysis. Sample clustering results revealed that there were two outlier samples (GSM2385767 and GSM2385782) (Figure 4a). Next, the sample clusters and phenotypic trait heatmap were constructed (Figure 4b). With a soft threshold equal to 6, R2 = 0.87 (red line), and average connectivity close to 0, which suggested that the interactions between genes maximally conform to a scale-free distribution (Figure 4c). A total of 12 modules were obtained by the dynamic tree-cut algorithm (Figure 4d). Of these, MEblue (naive B cells: Cor = -0.61, P = 2e-05; gamma delta T cells: Cor = 0.61, P = 2e-05; macrophages M2: Cor = 0.77, P = 3e-09; resting dendritic cells: Cor = 0.45, P = 0.003; activated dendritic cells: Cor = -0.54, P = 2e-04; resting mast cells: Cor = 0.72, P = 1e-07; activated mast cells: Cor = -0.72, P =1e-07; Neutrophils: Cor = -0.72, P = 6e-08) showed the highest correlation with differential immune cells overall (Figure 4e). Therefore this module was considered as key module. Finally, the 3442 genes in this key module were defined as key module genes for subsequent analysis.



**Figure 3.** Immune infiltration analysis. (a,b) The proportion of 20 immune cells in the HC (a) and SS (b) groups. (c) Discrepancies of fraction of immune cells between SS and HC samples.



**Figure 4.** Identification of differential immune cells-related key genes. (a) The samples in the GSE89632 dataset were clustered to remove the outlier. (b) The heat map of sample clustering and traits. (c) Selection of the optimal soft-thresholding (power). (d) Hierarchical clustering of genes and module identification. (e) Heatmap of the relationships between gene modules and differential immune cells.



**Figure 5.** Identification of biomarkers for SS. (a) The Venn diagram of common genes obtained by overlapping DE-PMRGs and differential immune cells-related key module genes. (b,c) Performance analysis of the three machine learning models. (d) The importance of 15 common genes in the optimal model.

## Screening for biomarkers

In total, 15 common genes were screened out by overlapping DE-PMRGs and key module genes (Figure 5a). The machine learning model was constructed by three machine learning algorithms based on 15 common genes. Performance analysis of the three models showed that the RF model had the lowest residual and thus was considered the optimal model (Figure 5b-5c). The top five genes from the optimal model (JUN, LDLR, CXCR4, NNMT, and ANXA1) were used as biomarkers for the subsequent analysis (Figure 5d).

## Construction of the nomogram

The nomogram on the basis of 5 biomarkers was utilized to predict the risk of developing SS in patients (Figure 6a). As illustrated by the calibration curve, the precision of the nomogram was excellent (Figure 6b). The decision curve analysis further illustrated that the benefit of the nomogram model was higher with the addition of biomarkers than with individual biomarker features (Figure 6c). In the influence curve, most of the predicted results (red curve) coincide with the actual curve (blue curve), indicating that the nomogram had an accurate prediction ability (Figure 6d).

## **GSEA** analysis of biomarkers

The functional enrichment results revealed that all biomarkers were enriched in the nucleoside triphosphate metabolic process, pyruvate metabolism, etc. (Figure 7a-7d).

## The ceRNA regulatory network

Using two databases, based on the biomarkers, we predicted 962 and 58 target miRNAs, respectively. Finally, 11 co-miRNAs were obtained by taking the intersection of the predicted targeting miRNAs from the two data-



**Figure 6.** Creation of the nomogram for SS patients. (a) The nomogram was established based on five biomarkers. (b) The calibration curve of the nomogram. (c) The decision curve of the nomogram. (d) The clinical influence curve of the nomogram.



Figure 7. GSEA enrichment analysis of biomarkers. (a-c) The GO terms enriched in biomarkers. (d) The KEGG pathways enriched in biomarkers.

bases (Figure 8a). Similarly, we acquired 149 co-lncR-NAs (Figure 8b). The lncRNA-miRNA-mRNA network was constructed according to the predicted mRNA-miR-NA and miRNA-lncRNA regulatory relationships. The network contained 5 biomarkers, 11 miRNAs (hsa-miR-152-3p, hsa-miR-185-5p and so on), and 149 lncRNAs (BLACAT1, C1orf220, etc.) (Figure 8c). The specific mRNA-miRNA pairs were LDLR-hsa-miR-152-3p, etc., and the miRNA-lncRNA pairs were MIR29B2CHG-hsalet-7b-5p, etc. The Sankey diagram indicated that LDLR had the most relational pairs (Figure 8d).

## Prediction of therapeutic agents of biomarkers

Through the DGIDB database, 5 biomarkers (JUN, LDLR, CXCR4, NNMT, and ANXA1) were found that targeted by therapeutic drugs (Figure 9a). The network included 44 drugs (holacanthone, sergeolide and so on) for JUN, 21 drugs (corticotropin, pravastatin and so on) for LDLR, 20 drugs (plerixafor, bkt140 and so on) for CXCR4, one drug (niacin) for NNMT, and 20 drugs (alclometasone, prednisone and so on) for ANXA1 (Figure 9b).

#### **Expression of biomarkers**

The results of the expression analysis showed that the expression of biomarkers in the training set was significantly higher in the HC group than in the SS group (Figure 10a). The expression trends of biomarkers in the GSE126848 dataset were the same as those in the training set (Figure 10b).

#### mRNA expression of five biomarkers in the samples

The results of the mRNA expression analysis showed that the expression of five biomarkers was significantly higher in the HC group than in the SS group (Figure 10c).

## Discussion

Nonalcoholic fatty liver disease affects approximately one-quarter of the global population due to its increased morbidity, mortality, economic burden and health care costs (28). Propionate, a major product of dietary fiber fermentation in the colon, is thought to lower serum cholesterol levels, and there has been increasing evidence for an important role of innate immunity in the development of hepatic steatosis in the last decade (29-31). Therefore based on the current study, we conducted a comprehensive bioinformatics analysis to identify immune and propionate-related diagnostic biomarkers associated with NAFLD progression.

Immunological analysis showed differences in  $\gamma\delta$ -T cells, M2 macrophages, dendritic cells, mast cells and neutrophils between the disease and healthy control groups. Dendritic cells are the major antigen-presenting cells of the liver (32), dendritic cells in NAFLD patients can respond to lipopolysaccharide stimulation, by secreting inflammatory cytokines that promote disease progression (33).  $\gamma\delta$ -T cells bridge the innate immune system and adaptive immunity and are more abundant in the liver than in the blood, accounting for 15-25% of total hepatic T cells (34). In NAFLD,  $\gamma\delta$ -T cells are recruited to the liver and exacerbate the progression of NAFLD by regulating CD4+ T cells and increasing the expression of IL-17 (35, 36). This is consistent with our results, in which NAFLD patients



**Figure 8.** Construction of the ceRNA network for biomarkers. (a) The Venn diagram of common miRNAs was obtained by overlapping the predicted miRNAs. (b) The Venn diagram of common lncRNAs obtained by overlapping the predicted lncRNAs. (c) The network was established based on the lncRNAs, miRNAs, and mRNAs. (d) The Sankey diagram of interaction among lncRNAs, miRNAs, and biomarkers. ceRNA, competing endogenous RNA; miRNA, microRNA; lncRNA, long non-coding RNA.



**Figure 9.** Drug prediction for SS. (a) The drugs targeting five biomarkers in the DGIDB database. (b) The network of biomarkers and drugs.



**Figure 10.** The expression of biomarkers and validation between SS and HC groups by qRT-PCR. (a,b) The expression of five biomarkers in the training set (a) and GSE126848 dataset (b). (c) Validation of biomarkers by qRT-PCR. \*P<0.05; \*\* P<0.01; \*\*\*\*P<0.0001.

in our study did have higher levels of infiltration of resting dendritic cells and  $\gamma\delta$ -T cells than the normal group. M2 macrophages have a weak antigen-presenting capacity, and they play an important role in immune regulation by secreting suppressive cytokines such as IL-10 and transforming growth factor- $\beta$  to down-regulate immune responses (37). Macrophage polarization switch from M1 to M2 phenotype may determine the regression of inflammation in NAFLD (38). The downregulation of M2 macrophages in the disease group may be due to a spontaneous anti-inflammatory response of the organism. Mast cells are generally present in low numbers throughout the body, and it has been shown that mast cells are significantly higher in stage 3-4 nonalcoholic steatohepatitis tissue compared to normal liver tissue, and that the number of mast cells correlates with the stage of liver fibrosis (39). In addition Lindsey Kennedy et al. found that mast cells can promote the development of steatosis biliary senescence and inflammation through the miR-144-3p/ALDH1A3 signaling pathway during the progression of NAFLD (40). Neutrophils can secrete IL-8 and tumor necrotic factor (TNF), among which elevated TNF- $\alpha$  concentrations are thought to be significantly associated with increased risk of NAFLD (41). TNF- $\alpha$  can stimulate the production of C-reactive protein (CRP) in the liver by activating intracellular kinases (41). A 2020 study showed an independent positive correlation between CRP and NAFLD, and even suggested that it could be used as a surrogate diagnostic marker for the severity of NAFLD disease (42). Similarly, IL-8 showed higher levels in nonalcoholic steatohepatitis. In our study, however, neutrophils were down-regulated in the disease group, which may be due to the higher levels of anti-inflammatory effects of the samples. Therefore, the role of neutrophils in NAFLD needs to be further explored. Some studies have shown that propionate can play a key role in regulating CD8<sup>+</sup> T cell activation by inhibiting IL-12 secretion from dendritic cells (43). In addition, the metabolites of propionate can affect the differentiation or function of T cells, macrophages and dendritic cells, and the abnormality of these cells can cause the occurrence and development of NAFLD (44,45). The changes in the expression of propionate-related biomarkers may modulate the inflammatory response in NAFLD.

We obtained five biomarkers associated with the diagnosis of NAFLD through a series of analyses, and the expression of JUN, LDLR, CXCR4, NNMT, and ANXA1 was significantly lower in the disease group than in the healthy control group, which may play a role in inhibiting the progression of NAFLD. Activator protein-1 (AP-1) is a dimeric transcription factor consisting of JUN, transcriptional activator, etc. For the liver, JUN deficiency does not impair organ homeostasis, but AP-1 is critical for the liver's response to acute stress (46-48). It has been demonstrated that the ROS/JNK/AP-1 signaling pathway has been shown to play an important role in HUA-mediated fat accumulation in the liver (49). And expression of AP-1 transcription factor c-Jun correlates with progression from NAFLD to non-alcoholic steatohepatitis in humans and mice (50). The effect between the two needs to be verified by further in vitro and in vivo experiments. The low-density lipoprotein receptor (LDLR), expressed primarily in the liver, helps remove approximately 70% of circulating LDL by endocytosing cholesterol-rich LDL (51), and it plays an important role in the regulation of plasma and intracellular cholesterol homeostasis (52). Dysregulation of LDLR expression causes abnormal accumulation of lipids in cells and tissues such as hepatocytes, renal tubular cells and podocytes (52). Dysregulation of the LDLR pathway may be a cause of accelerated lipid disorder-mediated target organ damage and thus NAFLD (52). As seen in our results, LDLR expression was higher in healthy controls, meaning that LDLR pathway dysregulation was present in NAFLD samples. The c-x-c chemokine receptor type 4 (CXCR4) is also known as a fusion protein or CD184 (53). Chemokines are involved in homeostatic or inflammatory regulation and can mediate pathophysiological changes in disease progression by binding to the corresponding receptors (54). The CXCL12/CXCR4 pathway is involved in the recruitment of CD4+ T cells in nonalcoholic steatohepatitis in both mice and humans (55). It has also been shown that CD4+ T cells are critical in promoting hepatic steatosis-fibrosis transition (35). However, it has also been shown that CXCL12/CXCR4 can improve fibrosis levels in many organs, including the heart, liver, lungs and kidneys (56). The effect of CXCL12 / CXCR4 on fibrosis has two sides, which needs to be further verified by experiments. Nicotinamide-N-methyltransferase (NNMT) is a methylating enzyme for nicotine (vitamin B3) using S-adenosylmethionine as the methyl donor (57). As early as 2018, it was found that polymorphisms in NNMT may be a genetic risk factor for the development of NAFLD, and individuals with wild-type NNMT have a lower risk of developing NAFLD (58). The membrane-linked protein A1 (ANXA1) is an important effector in the regression of inflammation (59), it has been shown to impair the recruitment of neutrophils to sites of inflammation (60). Gadipudi et al. found no effect on hepatic steatosis when treating nonalcoholic steatohepatitis with recombinant human ANXA1 (hr ANXA1), but it reduced liver inflammation and fibrosis (59). This means that NNMT and ANXA1 favor a decreased chance of prevalence or reduced inflammation in NAFLD.

Functional enrichment analysis showed that biomarkers were associated with AGE-RAGE signaling pathway, TNF signaling pathway, Toll-like receptor signaling pathway, IL-17 signaling pathway and other pathways in diabetic complications. AGE/RAGE activation increases oxidative stress and triggers a series of inflammatory responses, angiogenesis and fibrosis (61). Moreover, AGE/ RAGE activation can also activate the NF-kappa B signaling pathway, which is associated with inflammation and fatty liver during liver injury (62, 63). It has been shown that TNF- $\alpha$  reverses LDLR inhibition mediated by lowdensity lipoprotein (LDL) loading and up-regulates the LDLR pathway, thereby increasing the uptake of natural LDL by human glomerular thylakoid cells (64). Then it is possible that it can up-regulate the LDLR pathway through the TNF signaling pathway and thus have an effect on NA-FLD. Toll-like receptor 4 (TLR4) is the upstream signal of the Th17-IL-17 axis (65), and recent studies have shown that TLR4 mediates inflammation in hepatic parenchymal and non-parenchymal cells during the early stages of NA-FLD and is involved in inflammatory events in liver tissue (66). Toll-like receptors are activated to degrade I-kappa B, while free NF-kappa B can enter the nucleus and promote the transcription of cytokines such as TNF-α. NF-kappa B and TNF signaling pathways are important inflammatory pathways in the process of cholestatic liver injury (67, 68).



Figure 9. Schematic representation of the biomarkers associated with immune and propionate metabolism in nonalcoholic fatty liver disease.

IL-17 rises during human obesity and the IL-17 axis has also been associated with liver injury, and IL-17 deficiency has been shown to have a protective effect against hepatitis (69). IL-17 can also induce the release of neutrophil chemokines (70). From the previous discussion we can learn that neutrophils can release cytokines such as TNF, which have an effect on inflammation in NAFLD, and elevated TNF inevitably affects the TNF signaling pathway. Therefore, these pathways may together play an important role in the development of NAFLD.

In addition, for the identified immune and propionate-related biomarkers, we constructed related upstream regulatory networks, such as the lncRNA-miRNA-mRNA network, which provide directions and a basis for further study of the mechanism of action of these genes in NA-FLD.

The present study has some limitations. First, our study was conducted on the basis of existing public database data, and the results obtained from the analysis need to be validated by additional clinical samples and clinical application studies. Secondly, the mechanism of action of the identified immune and propionate metabolism-related disease diagnostic biomarkers needs further experimental investigation.

## Conclusions

In summary, this study identified five diagnostic biomarkers of NAFLD associated with immunity and propionate metabolism, based on immune cells that differed between the disease group and healthy controls, as well as differentially expressed genes (Figure 11). We also analyzed the biomarker pathways, and we hypothesized that these biomarkers are more strongly associated with immune-related pathways and would affect NAFLD through the AGE-RAGE signaling pathway, TNF signaling pathway and Toll-like receptor signaling pathway in diabetic complications. It provides an important reference for future diagnosis, mechanism research and treatment of the disease.

## **Conflicts of Interest**

The authors declare no conflict of interest.

## Ethics approval and consent to participate

The animal experiments were approved by the local ethics committee of Harbin Medical University.

## Data Availability Statement

The GSE89632 and GSE126848 datasets analyzed in this study were sourced from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/).

## Funding

This work was supported by the Natural Science Foundation of Heilongjiang Province (ZD2022H004), the Outstanding Youth Program of Heilongjiang Natural Science Foundation (JQ2019H002).

## **Author Contributions**

Conception and design, S.Q. and Y.L. (Yanze Li); Data curation, S.Q., Y.L. (Yanze Li) and Y.L. (Ying Liu); Project administration, Y.L. (Yanze Li) and Y.L. (Ying Liu); Validation, H.H. and X.C.; Acquisition of data, S.Q., Y.D., B.L., B.T., S.H., Y.L. (Yu Lei), W.M., P.L., J.Z.; Analysis and interpretation, S.Q. and Y.L. (Yanze Li); Writing, review and/or revision of the manuscript, S.Q., Y.L. (Yanze Li) and Y.L. (Ying Liu); Supervision, Y.L. (Yanze Li) and Y.L. (Ying Liu); Supervision, Y.L. (Yanze Li) and Y.L. (Ying Liu). The final manuscript was read and approved by all authors. All authors agree with the final contributions and order. All authors have read and agreed to the published version of the manuscript.

## Acknowledgments

Thanks to Yanze Li for her guidance in the experiment and writing process.

## References

- Marjot T, Moolla A, Cobbold JF, Hodson L, Tomlinson JW. Nonalcoholic Fatty Liver Disease in Adults: Current Concepts in Etiology Outcomes, and Management. Endocr Rev 2020; 41(1): bnz009.
- Zheng Y, Shi H, Zhou Y, Wang A, Kang D, Kang L. Effects of endoplasmic reticulum stress, liver function, insulin resistance and vascular endothelial function in patients with nonalcoholic fatty liver disease. Cell Mol Biol 2021; 67(5): 210-217.
- Arab JP, Arrese M, Trauner M. Recent Insights into the Pathogenesis of Nonalcoholic Fatty Liver Disease. Annu Rev Pathol-Mech 2018; 13: 321-350.
- Hu R, Liu S, Shen W, et al. Study on the Inhibitory Effects of Naringenin-Loaded Nanostructured Lipid Carriers Against Nonalcoholic Fatty Liver Disease. J Biomed Nanotechnol 2021; 17(5): 942-951.
- Polyzos SA, Kountouras J, Mantzoros CS. Obesity and nonalcoholic fatty liver disease: From pathophysiology to therapeutics. Metabolism 2019; 92: 82-97.
- Le MH, Yeo YH, Li X, et al. 2019 Global NAFLD Prevalence: A Systematic Review and Meta-analysis. Clin Gastroenterol H 2022; 20(12): 2809-2817.
- Piazzolla VA, Mangia A. Noninvasive Diagnosis of NAFLD and NASH. Cells-Basel 2020; 9(4): 1005.
- Battisti UM, Monjas L, Akladios F, et al. Exploration of Novel Urolithin C Derivatives as Non-Competitive Inhibitors of Liver Pyruvate Kinase. Pharmaceuticals-Base 2023; 16(5): 668.
- Ye Q, Zou B, Yeo YH, et al. Global prevalence, incidence, and outcomes of non-obese or lean non-alcoholic fatty liver disease: a systematic review and meta-analysis. Lancet Gastroenterol 2020; 5(8): 739-752.
- Konturek PC, Harsch IA, Konturek K, et al. Gut(-)Liver Axis: How Do Gut Bacteria Influence the Liver? Med Sci (Basel) 2018; 6(3): 79.
- Teunis C, Nieuwdorp M, Hanssen N. Interactions between Tryptophan Metabolism, the Gut Microbiome and the Immune System as Potential Drivers of Non-Alcoholic Fatty Liver Disease (NAFLD) and Metabolic Diseases. Metabolites 2022; 12(6): 514.
- 12. Lopez-Escalera S, Lund ML, Hermes G, Choi BS, Sakamoto K,

Wellejus A. In Vitro Screening for Probiotic Properties of Lactobacillus and Bifidobacterium Strains in Assays Relevant for Non-Alcoholic Fatty Liver Disease Prevention. Nutrients 2023; 15(10): 2361.

- Zhang S, Zhao J, Xie F, et al. Dietary fiber-derived short-chain fatty acids: A potential therapeutic target to alleviate obesity-related nonalcoholic fatty liver disease. Obes Rev 2021; 22(11): e13316.
- Hosseini E, Grootaert C, Verstraete W, Van de Wiele T. Propionate as a health-promoting microbial metabolite in the human gut. Nutr Rev 2011; 69(5): 245-258.
- Kumar SA, Ward LC, Brown L. Inulin oligofructose attenuates metabolic syndrome in high-carbohydrate, high-fat diet-fed rats. Brit J Nutr 2016; 116(9): 1502-1511.
- 16. Haghikia A, Zimmermann F, Schumann P, et al. Propionate attenuates atherosclerosis by immune-dependent regulation of intestinal cholesterol metabolism. Eur Heart J 2022; 43(6): 518-533.
- Arrese M, Cabrera D, Kalergis AM, Feldstein AE. Innate Immunity and Inflammation in NAFLD/NASH. Digest Dis Sci 2016; 61(5): 1294-1303.
- Andel M, Polak J, Kraml P, Dlouhy P, Stich V. [Chronic mild inflammation links obesity, metabolic syndrome, atherosclerosis and diabetes]. Vnitr Lek 2009; 55(7-8): 659-665.
- Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43(7): e47.
- 20. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014; 15(12): 550.
- Ito K, Murphy D. Application of ggplot2 to Pharmacometric Graphics. Cpt-Pharmacomet Syst 2013; 2(10): e79.
- Zeng S, Xie L, Mao G, et al. Exploring the Mechanism of the Baishao Luoshi Formula against Poststroke Spasticity by Network Pharmacology and Experimental Validation. Curr Comput-Aid Drug 2022; 18(7): 480-492.
- Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015; 12(5): 453-457.
- 24. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. Bmc Bioinformatics 2008; 9: 559.
- 25. Mahmoudian M, Venalainen MS, Klen R, Elo LL. Stable Iterative Variable Selection. Bioinformatics 2021; 37(24): 4810-4817.
- Tatsumi K, Igarashi N, Mengxue X. Prediction of plant-level tomato biomass and yield using machine learning with unmanned aerial vehicle imagery. Plant Methods 2021; 17(1): 77.
- Su G, Morris JH, Demchak B, Bader GD. Biological network exploration with Cytoscape 3. Curr Protoc Bioinformatics 2014; 47: 8-13.
- Bhat N, Mani A. Dysregulation of Lipid and Glucose Metabolism in Nonalcoholic Fatty Liver Disease. Nutrients 2023; 15(10): 2323.
- 29. Deng ZB, Liu Y, Liu C, et al. Immature myeloid cells induced by a high-fat diet contribute to liver inflammation. Hepatology 2009; 50(5): 1412-1420.
- Xu H, Li H, Woo SL, et al. Myeloid cell-specific disruption of Period1 and Period2 exacerbates diet-induced inflammation and insulin resistance. J Biol Chem 2014; 289(23): 16374-16388.
- Luo X, Li H, Ma L, et al. Expression of STING Is Increased in Liver Tissues From Patients With NAFLD and Promotes Macrophage-Mediated Hepatic Inflammation and Fibrosis in Mice. Gastroenterology 2018; 155(6): 1971-1984.
- Lleo A, Invernizzi P. Apotopes and innate immune system: novel players in the primary biliary cirrhosis scenario. Digest Liver Dis 2013; 45(8): 630-636.

- Rana D, Duseja A, Dhiman RK, Chawla Y, Arora SK. Maturation defective myeloid dendritic cells in nonalcoholic fatty liver disease patients release inflammatory cytokines in response to endotoxin. Hepatol Int 2013; 7(2): 562-569.
- 34. Gao B, Jeong WI, Tian Z. Liver: An organ with predominant innate immunity. Hepatology 2008; 47(2): 729-736.
- Her Z, Tan J, Lim YS, et al. CD4(+) T Cells Mediate the Development of Liver Fibrosis in High Fat Diet-Induced NAFLD in Humanized Mice. Front Immunol 2020; 11: 580968.
- Torres-Hernandez A, Wang W, Nikiforov Y, et al. gammadelta T Cells Promote Steatohepatitis by Orchestrating Innate and Adaptive Immune Programming. Hepatology 2020; 71(2): 477-494.
- 37. Wynn TA, Vannella KM. Macrophages in Tissue Repair, Regeneration, and Fibrosis. Immunity 2016; 44(3): 450-462.
- Zou Y, Chen W, Chen J, You Z. Vascular Endothelial Growth Factor Receptor 1 Facilitates the Effect of Macrophages on Human Umbilical Vein Endothelial Cells Migration by Regulating the M1 Polarization. J Biomed Nanotechnol 2022; 18(12): 2839-2845.
- Lombardo J, Broadwater D, Collins R, Cebe K, Brady R, Harrison S. Hepatic mast cell concentration directly correlates to stage of fibrosis in NASH. Hum Pathol 2019; 86: 129-135.
- Kennedy L, Meadows V, Sybenga A, et al. Mast Cells Promote Nonalcoholic Fatty Liver Disease Phenotypes and Microvesicular Steatosis in Mice Fed a Western Diet. Hepatology 2021; 74(1): 164-182.
- 41. Duan Y, Pan X, Luo J, et al. Association of Inflammatory Cytokines With Non-Alcoholic Fatty Liver Disease. Front Immunol 2022; 13: 880298.
- 42. Kumar R, Porwal YC, Dev N, Kumar P, Chakravarthy S, Kumawat A. Association of high-sensitivity C-reactive protein (hs-CRP) with non-alcoholic fatty liver disease (NAFLD) in Asian Indians: A cross-sectional study. J Fam Med Prim Care 2020; 9(1): 390-394.
- Nastasi C, Fredholm S, Willerslev-Olsen A, et al. Butyrate and propionate inhibit antigen-specific CD8(+) T cell activation by suppressing IL-12 production by antigen-presenting cells. Sci Rep-Uk 2017; 7(1): 14516.
- 44. Sanchez HN, Moroney JB, Gan H, et al. B cell-intrinsic epigenetic modulation of antibody responses by dietary fiber-derived short-chain fatty acids. Nat Commun 2020; 11(1): 60.
- Dupraz L, Magniez A, Rolhion N, et al. Gut microbiota-derived short-chain fatty acids regulate IL-17 production by mouse and human intestinal gammadelta T cells. Cell Rep 2021; 36(1): 109332.
- Min L, Ji Y, Bakiri L, et al. Liver cancer initiation is controlled by AP-1 through SIRT6-dependent inhibition of survivin. Nat Cell Biol 2012; 14(11): 1203-1211.
- Fuest M, Willim K, MacNelly S, et al. The transcription factor c-Jun protects against sustained hepatic endoplasmic reticulum stress thereby promoting hepatocyte survival. Hepatology 2012; 55(2): 408-418.
- 48. Machida K, Tsukamoto H, Liu JC, et al. c-Jun mediates hepatitis C virus hepatocarcinogenesis through signal transducer and activator of transcription 3 and nitric oxide-dependent impairment of oxidative DNA repair. Hepatology 2010; 52(2): 480-492.
- Xie D, Zhao H, Lu J, et al. High uric acid induces liver fat accumulation via ROS/JNK/AP-1 signaling. Am J Physiol-Endoc M 2021; 320(6): E1032-E1043.
- Schulien I, Hockenjos B, Schmitt-Graeff A, 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(9): 1688-1699.
- 51. He Y, Rodrigues RM, Wang X, et al. Neutrophil-to-hepatocyte communication via LDLR-dependent miR-223-enriched extra-

cellular vesicle transfer ameliorates nonalcoholic steatohepatitis. J Clin Invest 2021; 131(3): e141513.

- Zhang Y, Ma KL, Ruan XZ, Liu BC. Dysregulation of the Low-Density Lipoprotein Receptor Pathway Is Involved in Lipid Disorder-Mediated Organ Injury. Int J Biol Sci 2016; 12(5): 569-579.
- 53. Zhang C, Li J, Han Y, Jiang J. A meta-analysis for CXCR4 as a prognostic marker and potential drug target in non-small cell lung cancer. Drug Des Devel Ther 2015; 9: 3267-3278.
- Huynh C, Dingemanse J, Meyer ZSH, Sidharta PN. Relevance of the CXCR4/CXCR7-CXCL12 axis and its effect in pathophysiological conditions. Pharmacol Res 2020; 161: 105092.
- Boujedidi H, Robert O, Bignon A, et al. CXCR4 dysfunction in non-alcoholic steatohepatitis in mice and patients. Clin Sci 2015; 128(4): 257-267.
- 56. Wu X, Qian L, Zhao H, et al. CXCL12/CXCR4: An amazing challenge and opportunity in the fight against fibrosis. Ageing Res Rev 2023; 83: 101809.
- Martin JL, McMillan FM. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Curr Opin Struc Biol 2002; 12(6): 783-793.
- 58. Hasan EM, Abd AAR, Sabry D, et al. Genetic Variants in nicotinamide-N-methyltransferase (NNMT) gene are related to the stage of non-alcoholic fatty liver disease diagnosed by controlled attenuation parameter (CAP)-fibroscan. J Gastrointest Liver 2018; 27(3): 265-272.
- 59. Gadipudi LL, Ramavath NN, Provera A, et al. Annexin A1 treatment prevents the evolution to fibrosis of experimental nonalcoholic steatohepatitis. Clin Sci 2022; 136(9): 643-656.
- Sugimoto MA, Vago JP, Teixeira MM, Sousa LP. Annexin A1 and the Resolution of Inflammation: Modulation of Neutrophil Recruitment, Apoptosis, and Clearance. J Immunol Res 2016; 2016: 8239258.
- 61. Ott C, Jacobs K, Haucke E, Navarrete SA, Grune T, Simm A. Role

of advanced glycation end products in cellular signaling. Redox Biol 2014; 2: 411-429.

- Zhang T, Hu J, Wang X, et al. MicroRNA-378 promotes hepatic inflammation and fibrosis via modulation of the NF-kappaB-TN-Falpha pathway. J Hepatol 2019; 70(1): 87-96.
- Ma P, Ding YS, Xuan LL, et al. Anti-inflammatory effect of a resveratrol derivative 3,4,5-trimethoxy-4',5'-dihydroxy-trans-stilbene (WL-09-5) via ROS-mediated NF-kappaB pathway. J Asian Nat Prod Res 2016; 18(10): 1004-1013.
- 64. Ruan XZ, Varghese Z, Powis SH, Moorhead JF. Dysregulation of LDL receptor under the influence of inflammatory cytokines: a new pathway for foam cell formation. Kidney Int 2001; 60(5): 1716-1725.
- 65. Mu HH, Nourian MM, Jiang HH, Tran JW, Cole BC. Mycoplasma superantigen initiates a TLR4-dependent Th17 cascade that enhances arthritis after blocking B7-1 in Mycoplasma arthritidisinfected mice. Cell Microbiol 2014; 16(6): 896-911.
- 66. Kiziltas S. Toll-like receptors in pathophysiology of liver diseases. World J Hepatol 2016; 8(32): 1354-1369.
- Li Z, Chen D, Jia Y, et al. Methane-Rich Saline Counteracts Cholestasis-Induced Liver Damage via Regulating the TLR4/NFkappaB/NLRP3 Inflammasome Pathway. Oxid Med Cell Longev 2019; 2019: 6565283.
- 68. Wang J, Zhu R, Sun D, et al. Intracellular Uptake of Curcumin-Loaded Solid Lipid Nanoparticles Exhibit Anti-Inflammatory Activities Superior to Those of Curcumin Through the NF-kappaB Signaling Pathway. J Biomed Nanotechnol 2015; 11(3): 403-415.
- Harley IT, Stankiewicz TE, Giles DA, et al. IL-17 signaling accelerates the progression of nonalcoholic fatty liver disease in mice. Hepatology 2014; 59(5): 1830-1839.
- Lafdil F, Miller AM, Ki SH, Gao B. Th17 cells and their associated cytokines in liver diseases. Cell Mol Immunol 2010; 7(4): 250-254.