Original Article
Screening of non-alcoholic steatohepatitis (NASH)-related datasets and identification of NASH-related genes

Ming-Jiang Liu, Hu Jin, Yu-Bing Chen, Jing-Jing Yu, Zhen-Ya Guo, Song-Qing He, Yong-Lian Zeng

Department of Hepatobiliary Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, Guangxi, China

Received December 10, 2020; Accepted February 24, 2021; Epub May 15, 2021; Published May 30, 2021

Abstract: Non-alcoholic steatohepatitis (NASH) is a common liver disease in the western world. The mechanisms behind NASH formation are poorly understood, but there may be multiple targets considering the disease's multifactorial nature. To explore the genes related to the pathogenesis of NASH, we downloaded clinical data and gene expression of NASH patients from the Gene Expression Omnibus database (GEO). We identified 281 genes with a common expression in two NASH-related datasets (GSE89632 and GSE83452), suggesting that they may be related to NASH. Further study showed that Angptl4, Foxo1, and Ttc39B might be essential for NASH progression, and these have been poorly studied. Therefore, we explored their roles in NASH. Our data show that these genes participate in the development of NASH through lipid metabolism. This suggests that the three genes can be used as therapeutic targets in NASH.

Keywords: Non-alcoholic steatohepatitis, Angptl4, Foxo1, Ttc39B

Introduction

Non-alcoholic fatty liver disease (NAFLD) is one of the most common liver diseases globally; with a rising incidence, the attention to the disease increases [1]. It comprises non-alcoholic fatty liver (NAFL) and non-alcoholic steatohepatitis (NASH), and NASH has been recognized as a chronic liver disease in the western world. At present, the diagnosis of NASH mainly relies on imaging and biopsy. There are insufficient molecular diagnostic criteria and no effective therapies available. However, NASH is closely associated with lipid metabolism [2], and lipotoxicity drives progressive hepatic inflammation and fibrosis in patients with NAFLD [3]. NASH is more severe than NAFLD and can progress to liver cirrhosis and cancer [4-6]. Although NASH’s pathogenesis is unclear, lipid accumulation may play an important role [7, 8].

Genome-wide studies have recently opened a new era for NASH research, and some molecular mechanisms have been investigated via bioinformatics to assess the pathogenesis and discover therapeutic targets [9]. Nevertheless, it is still a hot spot that the studies are about the pathogenesis and treatment strategy of NASH.

In this study, we identify 281 NASH-related genes by bioinformatic analysis of the NASH database in the Gene Expression Omnibus database (GEO) [10]. We show that Angptl4, Foxo1, and Ttc39B are predicted to participate in the regulation of lipid metabolism, and deficiencies in these genes may play an essential role in mitigating NASH. Finally, we demonstrate these three genes can serve as reliable biomarkers for the clinical management of patients with NASH.

Materials and methods

Identification of differentially expressed genes (DEGs) in NASH

The gene data from the GSE89632 and GSE83452 datasets were extracted from the GEO website (http://www.ncbi.nlm.nih.gov/geo). Ge-
Angptl4, Foxo1 and Ttc39B and NASH

o2R (https://www.ncbi.nlm.nih.gov/geo/geo-2r/) was used to identify the DEGs between the NASH and the control group.

Volcano Plots were constructed using NetworkAnalyst (https://www.networkanalyst.ca/). The selection criteria for NASH datasets were to compare the NASH group with the normal group P < 0.05. After the screening, we identified two sets of DEGs. Venn (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to acquire the co-expressed genes between the two datasets. Then, these genes were used for further analysis.

Functional enrichment analyses

Gene Ontology (GO) [11] classification was performed for the co-expressed genes using the online analysis tool Metascape (http://metascape.org/). P-values < 0.05 were defined as significant.

Construction of protein-protein interaction (PPI) network

STRING 11.0 (http://string-db.org/) was used to construct a PPI network and explore the interaction between common genes. The minimum required interaction score was considered high confidence (0.500) as the criteria for statistical significance.

Cell culture and treatment

AML12 cells were provided by the Stem Cell Bank of the Chinese Academy of Sciences. Cells were cultured in DMEM/F12 containing 10% fetal bovine serum, 1% Insulin-Transferrin-Selenium (ITS), 0.1 μM dexamethasone solution, and 1% penicillin/streptomycin. The AML12 cells were transfected with small interfering RNA (siRNA) coding for Angptl4, Foxo1, and Ttc39B (RiboBio, Guangzhou, China) using Lipofectamine 2000 (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's instructions. The si-RNAs were synthesized by RiboBio (Guangzhou, China), and the sequences are listed in Table 1. After transfection for 24 hours, the cells were treated with a palmitic acid (PA: 0.1 mM; Sigma-Aldrich) and oleic acid (OA: 0.2 mM; Sigma-Aldrich) mixture for 24 hours to establish an in vitro model of lipid accumulation in AML12 cells. The lipid accumulation in cells was observed by Oil Red O staining. Images were taken with a microscope (IX-71; Olympus, Tokyo, Japan). The cells were cultured at 37°C and 5% CO₂.

Animals and treatment

Male adult ob/ob (leptin-deficient) mice were purchased from the Model Animal Research Center of Nanjing University, and C57BL/6J mice (8 weeks) from the Hunan SJA Laboratory Animal Co, Ltd. Mice were housed at the Animal Facilities of Guangxi Medical University. Ten C57BL/6J mice were split into two weight-matched groups and were fed a normal diet (ND) or a high-fat diet (HFD). A NASH model was established using the C57BL/6J mice by feeding them with HFD (protein, 20%; fat, 60%; carbohydrates, 20%; TP23400, Trophic, Nantong, Jiangsu, China) for 24 weeks. Those fed with ND (protein, 18%; fat, 10%; carbohydrates, 72%; TP23402, Trophic, Nantong, Jiangsu, China) served as controls. The ob/ob and wild-type (WT, C57BL/6) mice were fed with ND for 24 weeks, and each group had free access to water and food. Mice were kept at 21-23°C in 12 hours of light-dark cycles. After 24 weeks, the mice were sacrificed, and tissue samples were collected. All samples were stored at -80°C. The animal experiments were approved by the Animal Care and Use Committee of Guangxi Medical University, China.

Quantitative real-time PCR

TRizol (15596026, Invitrogen, USA) was used to isolate total RNA from liver tissue. RNA was reverse-transcribed into cDNA with the RevertAid First Strand cDNA Synthesis kit (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer’s instructions. RT-PCR was performed using SYBR Green mix (Bio-Rad, California, USA) on a CFX96 Touch Real-Time system (Bio-Rad, California, USA). All primers used in this study are listed in Table 1. GAPDH was used as an internal control.

Histopathologic and biochemical analyses

The liver tissues were treated with 10% neutral formalin for 24 h, after which the tissues were dehydrated and embedded in paraffin. 4 μm thick tissue sections were prepared, and hematoxylin and eosin (H&E) staining was performed. Fresh liver tissues were frozen, cut into 8 μm slides, and stained with Oil Red O. The frozen
sections were fixed with 10% Formaldehyde calcium, rinsed with water, and soaked in isopropanol for 10 seconds. Next, Oil Red O was added for minimally 10 min, after which slides were washed with water and counterstained with hematoxylin. Histopathologic alterations in the liver tissues were observed using a NanoZoomer S60 C13210 series (S60 C13210, Hamamatsu Photonics K.K., Japan). The levels of alanine aminotransferase (ALT) and aspartate transaminase (AST) were measured with an autoanalyzer (Antech Diagnostics, CA, USA).

**Measurement of triglyceride levels in AML12 cell and liver homogenates**

The triglyceride (TG) levels in cell and liver tissue homogenates were measured using standard kits Triglyceride assay kit (Nanjing Jiancheng Bioengineering Institute, A110-1, Nanjing, China) following the manufacturer's protocols.

**Serum cytokine**

TNF-α and IL-6 levels in serum were measured using commercially available ELISA kits (Cusabio, Wuhan, China) per the manufacturer's instructions.

**Metabolic assays**

After the mice were fasted for 6 h, the fasting blood glucose (FBG) was examined using a glucometer (One Touch Ultra Easy, Life Scan, Wayne, USA), and fasting insulin (FINS) levels were measured using ELISA (10-1247-01, Mercodia, Sweden). Homeostatic Model Asessment of Insulin Resistance (HOMA-IR) was calculated as follows: FBG (mmol/L) × FINS (mIU/L)/22.5. A glucose tolerance test (GTT) was performed by intraperitoneal injection of 1 g/kg glucose (Kelun, Sichuan, China) in mice. Mice were injected with 0.75 U/kg of insulin (I8830, Solarbio, Beijing, China) for an insulin tolerance test (ITT). Six time points (0, 15, 30, 60, 90, and 120 min) were selected to determine the blood glucose level after intraperitoneal injections.

**Constructing the competing endogenous RNA (ceRNA) network**

The theoretical basis of the IncRNA-miRNA-mRNA ceRNA network was that IncRNAs could regulate mRNA activity through direct interactions with miRNAs [12]. MiRBase [13] (http://www.mirbase.org/) was used to identify the miRNA-targeted mRNAs. The ability of IncRNAs to sequester and bind miRNAs was predicted using LncBase Predicted v.2 [14]. (http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=LncBasev2). The ceRNA network was plotted with Cytoscape v3.6.0 [15].

**Constructing a transcription factor-related ceRNA network**

We chose 2 kbp upstream and 100 bp downstream of the transcription start site to screen for potential transcription factors of the three selected genes by applying Animal TFDB3.0 (https://bigd.big.ac.cn/databasecommons/database/id/8) and predicting the upstream transcription factors of mRNAs besides the ceRNA network [16]. Scores above 25 and P values < 0.01 were used to select significant transcription factors.

**Statistical analysis**

The results were presented as the mean ± SD. GraphPad Prism 7 software (GraphPad Software) was used to conduct data analysis. For multiple comparisons, statistical evaluation was conducted using ANOVA, followed by posthoc Tukey comparison. The two-tailed
Student’s t-test was used to compare two variables. The threshold for a significant difference was $P < 0.05$.

**Results**

**Bioinformatic analysis of two NASH-related datasets**

The mRNA expression levels of target genes involved in NASH were analyzed by mining the GEO database. Two NASH-related datasets (GSE89632 and GSE83452) were selected (Figure 1A, 1B). A total of 4,357 and 986 DEGs achieved $P < 0.05$. Using the Venn diagram online tool, 281 intersecting common genes of the two NASH-related datasets were obtained (Figure 1C). A PPI network was constructed to distinguish the target genes from common genes. The proteins are represented as nodes, and the interactions between the proteins as lines in the PPI network. Most of these common genes were related to lipid metabolism, as shown in Figure 1D and 1E. Lipid metabolism plays a critical role in NASH. Some of these lipid metabolism-related genes were previously studied and published [17, 18]. The GO enrichment analyses of 281 common genes showed that these common genes were specifically enriched in the liver tissue and adipocytes (Figure 1F). The visualization of the top 20 clusters was shown in Figure 1G. Differences in the GO enrichment analyses were notably focused on the enrichment of lipid metabolism (e.g., steroid metabolic processes, plasma lipoprotein assembly, remodeling and clearance, regulation of lipase activity, fatty acid derivative biosynthetic processes, regulation of fat cell differentiation, and triacylglycerol biosynthesis). Differences were also observed in inflammatory and other signaling pathways. We detected Fox1, Angptl4, and Ttc39B as three important genes involved in lipid metabolism among these genes, whose function in NASH remains unknown. These were selected in the present study for further investigation.

**Upregulation of the mRNA expression of three genes in NASH in the ob/ob mouse model**

To verify these NASH-related genes, a classical ob/ob mouse model of NASH was constructed. We confirmed, using H&E and Oil Red O staining, that the liver of ob/ob mouse had severe lipid deposition and experienced NASH. On the contrary, wild-type mice had normal livers (Figure 2A). The serum and liver TG levels in ob/ob mice increased after 24 weeks of ND (normal diet) induction (Figure 2B and 2C). The mRNA expression of Angptl4, Foxo1, and Ttc39B in the ob/ob mice was higher than that in the WT mice (Figure 2D), implying that the expression of the three genes might be correlated with NASH.

**Angptl4, Foxo1, and Ttc39B are correlated with NASH in the C57BL/6J mouse model of NASH**

The NASH model of wild type mice was generated by subjecting mice to an HFD for 24 weeks to explore further the differences among the three genes in NASH in vivo. ITT and GTT confirmed the glucose metabolism of these mice. Compared with the ND-treated group, the HFD-treated group showed a markedly increased area under the curve obtained from GTT and ITT assays (Figure 3A and 3B). The HFD-induced mice showed higher fasting glucose serum levels (Figure 3C), insulin levels (Figure 3D), and HOMA-IR (Figure 3E) than the ND-induced mice, suggestive of insulin resistance in the HFD-induced mice.

After 24 weeks of HFD consumption, the degree of lipid accumulation was more severe in the HFD-induced group than in the ND-induced group, as evidenced by H&E (Figure 4Aa, 4Ab) and Oil Red O staining (Figure 4Ac, 4Ad). As shown in Figure 4B and 4C, the TG content in the serum ($P < 0.001$) and liver ($P < 0.001$) of the ND-induced mice were significantly lower than those in the HFD-induced mice. The liver function in the HFD group exhibited markedly higher serum ALT ($P < 0.05$) and AST ($P < 0.05$) levels than those in the ND group as shown in Figure 4D and 4E. The IL-6 ($P < 0.05$) and TNF-α ($P < 0.05$) levels were higher in the HFD-induced mice than in the ND-induced mice, indicating more severe inflammation in the HFD-induced mice as shown in Figure 4F and 4G. Thus, HFD induction for 24 weeks created the characteristics of NASH. The expression of Angptl4, Foxo1, and Ttc39B was examined in this NASH model. As shown in Figure 4H, the mRNA expression of these three genes was higher in the HFD-induced mice than in the ND-induced mice (Angptl4: $P < 0.01$; Foxo1: $P < 0.05$; Ttc39B: $P < 0.05$).
Angptl4, Foxo1 and Ttc39B and NASH
Figure 1. Bioinformatic analysis of two NASH-related datasets. A, B. Volcano Plot indicating the DEGs (red, upregulated genes; blue, downregulated genes) in the Normal and NASH groups. C. Venn diagram of the intersection of common genes identified by DEGs from the two NAFLD datasets. DEGs defined from the pairwise comparisons were required to satisfy two selection criteria: a fold change of >1.5 and a corresponding adjusted p-value < 0.05. D. A PPI network was constructed with the common 281 genes. E. Three relative genes (Angptl4, Foxo1, and Ttc39B) were in the PPI network. F. These common genes were specifically enriched in liver tissue and adipocytes. G. GO analysis of common genes were shown in Metascape, which only visualized the top 20 clusters.
Angptl4, Foxo1, and Ttc39B are involved in NASH by affecting lipid metabolism

An in vitro model of lipid accumulation in AML12 cells was established to investigate the function of the genes Angptl4, Foxo1, and Ttc39B. A palmitic acid (PA) and oleic acid (OA) mixture was added to the AML12 cells for 24 hours. The expression levels of these genes were higher than those in the control group (Figure 5A). Knockdown of the three genes in AML12 cells supplemented with PA and OA mixture mitigated the lipid accumulation compared with the control group. The intracellular TGs content was consistent with that of the Oil Red O staining (Figure 5B-D). The expression levels of genes related to fatty acid synthesis and elimination were detected by quantitative RT-PCR. The expression of genes related to SREBP-1c, FASN, and ACC were significantly inhibited when the three genes were knocked down, whereas the expression of genes responsible for PPAR-α and CPT-1 was much higher in siRNA-treated cells than in the si-NC cells after PA and OA treatment for 24 hours (Figure 5E and 5F).

These in vitro data suggest that the decreased expression of the three genes may play a protective role in NASH pathogenesis and may serve as targets for treating NASH.
The ceRNA network

A ceRNA network was constructed to predict the upstream miRNAs and LncRNAs of the three target genes. This study was constructed based on the miRNA, IncRNA, and mRNA expression profiles of NASH patients to improve the reliability of the ceRNA network. Three target mRNAs, 20 miRNAs, and 124 LncRNAs were applied to define the ceRNA network datasets of NASH from the GEO database. Lipid metabolism-related biological functions were obtained through GO enrichment analyses. Three genes, Angptl4, Foxo1, and Ttc39B, which functions remain unclear but are related to lipid metabolism, were selected to further verify the practicability and reliability of the datasets. Angptl4, Foxo1, and Ttc39B were identified to have high expression in NASH-induced mice and OA and PA-treated cells. In (Figure 6). The key molecules in the network should be considered the next key research goals, as these may serve as diagnostic and therapeutic targets.

Transcription factor-related network construction and analysis

A network of interactions was constructed to predict the upstream transcription factors of the three selected target genes (Figure 7). IRF5, GTF3C2, and GLYR1 were the common transcription factors of Ttc39B and Angptl4 in the network. Angptl4 and Foxo1 had common transcription factors, which were FOXA2, MYC, and HDAC2. Angptl4, Foxo1 and Ttc39B may have similar roles in biological function. They may influence lipid metabolism and NASH by regulating these three genes.

Discussion

NASH is a global problem with no available effective therapies. NAFLD is characterized by excessive lipid accumulation in hepatocytes, and NASH is the liver manifestation of metabolic syndrome [19, 20]. Excess accumulated lipid results in lipotoxicity and causes liver inflammation [6, 21, 22]. Therefore, exploring and targeting lipid metabolism mechanisms might serve as an effective way to prevent and treat this disease [23].

In the present study, 281 common DEGs were found in two datasets of NASH from the GEO database. Lipid metabolism-related biological functions were obtained through GO enrichment analyses. Three genes, Angptl4, Foxo1, and Ttc39B, which functions remain unclear but are related to lipid metabolism, were selected to further verify the practicability and reliability of the datasets. Angptl4, Foxo1, and Ttc39B were identified to have high expression in NASH-induced mice and OA and PA-treated cells. In
Figure 4. Hepatic steatosis, inflammation, and the mRNA expression of Angptl4, Foxo1 and Ttc39B in C57BL/6J mice. A. Microscopic views of livers from the normal diet (ND) or high-fat diet (HFD) mice at 24 weeks of the diet. Representative liver sections stained with H&E (a, b) and Oil Red O (c, d) are shown. Original magnification, Scale bar: 100 μm. B, C. Triglycerides (TG) levels in the liver and serum samples of C57BL/6J mice at 24 weeks after HFD treatment. D-G. Serum ALT, AST, IL-6, and TNF-α levels in mice in each group were measured by ELISA (n=5). H. Relative mRNA levels of Angptl4, Foxo1, and Ttc39B were detected in C57BL/6J mice (fed for 24 weeks). *P < 0.05; **P < 0.01, ***P < 0.001.
the cell experiments, these three genes were knocked down by siRNA and lipid accumulation was alleviated, and similar results were found in the in vivo studies. Therefore, our data suggest that these genes have essential roles in NASH.

Our results are consistent with other studies that showed that these genes influence lipid metabolism. Angptl4 is primarily expressed in liver and adipose tissue and is also a secreted protein and adipokine. Lipoprotein lipase (LPL) can increase the metabolism of triglycerides in

Figure 5. The expression and function of Angptl4, Foxo1, and Ttc39B in vitro. A. The expression of Angptl4, Foxo1, and Ttc39B was higher than the control group after PA OA treatment for 24 hours. B. The results of Oil Red O staining on the AML12 cells transfected with siRNA of Angptl4, Foxo1, or Ttc39B after 24 hours and P-AOA stimulation for 24 hours. Scale bar: 100 μm. C. The expressions of Angptl4, Foxo1, and Ttc39B after being transfected with representative siRNA after 24 hours. D. AML12 cells were transfected with representative siRNA after 24 hours, with added palmitic acid (PA) and oleic acid (OA) treatment for 24 hours, and intracellular triglycerides were examined. E, F. mRNA expression of genes associated with fatty acid metabolism in the AML12 cells after siRNA transfection and PA and OA treatment.

Figure 6. The constructed LncRNA-miRNA-mRNA CeRNA network. The green hexagons represent LncRNAs, the blue ellipses represent miRNAs, and the red diamonds represent mRNAs.
Angptl4, Foxo1 and Ttc39B and NASH

which was associated with the PPARα/Foxo1 signaling pathway [35]. In addition, many drugs that treat non-alcoholic fatty liver disease also work by modulating Foxo1 [36, 37].

Ttc39B is a protein-coding gene involved in a variety of pathophysiologic functions, as it affects serum high-density lipoprotein cholesterol levels and is associated with susceptibility to coronary heart disease, ischemic stroke, and gallbladder disease [38-41]. Ttc39B regulates HDL metabolism, which may be associated with the promotion of ubiquitination and degradation of the oxysterol receptor LXR [42]. Hsieh et al. established a NAFLD mouse model and found that Ttc39B regulates lipid deposition in vivo [43]. In summary, these data suggest that Ttc39B is one of the genes associated with lipoprotein metabolism.

Figure 7. The transcription factor-related RNA network. The yellow ellipses represent transcription factors, and the red diamonds represent mRNAs.

In addition, we predicted the upstream miRNAs and LncRNAs of the three genes of interest. The interaction between the RNA molecules in the network may play an important regulatory role in the function of the three genes and even affect lipid metabolism and NASH processes. We predicted the upstream transcription factors of these three genes and showed that some common transcription factors could regulate the function of the genes. Thus, these RNAs and transcription factors may be therapeutic targets.

Conclusions

In summary, we identified the genes (281) that play critical roles in the process of NASH. We validated three genes to play an important role in NASH. Our data suggest that these three genes can be used as targets for the diagnosis or treatment of NASH.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (817716-
References


Angptl4, Foxo1 and Ttc39B and NASH


