Bioinformatics analysis explores potential hub genes in nonalcoholic fatty liver disease

Chutian Wu
Jinan University First Affiliated Hospital
Yun Zhou
First Affiliated Hospital of Gannan Medical University
Min Wang
Jinan University First Affiliated Hospital
Guolin Dai
Jinan University First Affiliated Hospital
Xiongxiu Liu
Jinan University First Affiliated Hospital
Leizhen Lai
Jinan University First Affiliated Hospital
Shaohui Tang (tangshaohui206@163.com)
The First Affiliated Hospital, Jinan University, Guangzhou, Guangdong, China https://orcid.org/0000-0002-1859-0876

Research article

Keywords: Nonalcoholic fatty liver disease, nonalcoholic steatohepatitis, differentially expressed genes, hepatocellular carcinoma, bioinformatics analysis

Posted Date: March 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-929980/v1

Abstract

Background: Nonalcoholic fatty liver disease (NAFLD) is now recognized as the most prevalent chronic liver disease worldwide. However, the dysregulated gene expression for NAFLD are still poorly understood.

Material and methods: we analyzed two public datasets (GSE48452 and GSE89632) to identify differentially expressed genes (DEGs) in NAFLD. Then, we performed a series of bioinformatics analysis to explore potential hub genes in NAFLD.

Results: This study included 26 simple steatosis (SS), 34 nonalcoholic steatohepatitis (NASH), and13 healthy controls (HC). We observed 6 up- and 19 down-regulated genes in SS, and 13 up- and 19 down-regulated genes in NASH compared with HC. Meanwhile, the overlapping pathways between SS and NASH were PI3K-Akt signaling pathway and pathways in cancer. Then, we screened out 10 hub genes by weighted Gene Co-Expression Network Analysis (WGCNA) and protein-protein interaction (PPI) networks. Eventually, we found that CYP7A1/GINS2/PDLIM3 were associated with the prognosis of hepatocellular carcinoma (HCC) in TCGA database.

Conclusion: Although further validation is still needed, we provide useful and novel information to explore the potential candidate genes for NAFLD prognosis and therapeutic options.

Introduction

Nonalcoholic fatty liver disease (NAFLD) is now recognized as the most prevalent chronic liver disease worldwide with a prevalence ranging from 13% in Africa to 42% in southeast Asia, and it may become the major cause of end-stage liver diseases by 2025[1–3]. NAFLD represents a spectrum of disease severity, ranging from simple steatosis (SS) termed as nonalcoholic fatty liver (NAFL) to nonalcoholic steatohepatitis (NASH), cirrhosis, and hepatocellular carcinoma (HCC)[4]. It has been well-recognized that obesity, insulin resistance, and type 2 diabetes mellitus are the strongest risk factors for NAFLD[5]. The cause of NAFLD is multifactorial including genetic and environmental factors[6]. However, the precise molecular mechanisms for NAFLD are still poorly understood, especially in dysregulated gene expression.

Lots of genome-wide association studies have indicated that PNPLA3, HNF1A, NCAN, GCKR, MBOTAT, FADS1, PPAR, TNF, and TM6SF2 are important genetic and epigenetic modifiers played important roles in the pathogenesis and progression of NAFLD[7]. Meanwhile, some bioinformatics researches offer new ideas for exploring potential targets of NAFLD. Zeng et al.[8] found that AKR1B10 and SPP1 were related to immune cell infiltrations and associated with the progression of NAFLD. Liu et al.[9] reported that TOP2A, NHP2L1, PCNA, CHEK1, ACACA, CCS, ACACB had a great impact in the progression of NAFLD and were associated with HCC progression. Although many studies have devoted to explore the pathogenesis and progression of NAFLD, there is still no effective drugs for the treatment of NAFLD except for lifestyle changes[10].

Therefore, we analyzed two public datasets to identify differentially expressed genes (DEGs) among healthy controls (HC), SS and NASH. Then, Weighted Gene Co-Expression Network Analysis (WGCNA) and protein-protein interaction (PPI) networks were performed to explore the impact of DEGs on NAFLD. This study aimed to screen potential genes for the NAFLD development.

Results

Identification of DEGs in the NAFLD patients

Firstly, we identified the DEGs among HC, SS, and NASH in GSE48452 and GSE89632 datasets, respectively (Fig. 2A-B **and Table S1-S2**). Then, we sought for the overlapping DEGs between the two datasets. We observed 6 up- and 19 down-regulated genes in SS compared with HC (Fig. 2C). We also found 13 up- and 19 down-regulated genes in NASH compared with HC (Fig. 2D).

Go And Kegg Pathway Enrichment Analysis

To explore the potential roles of DEGs among HC, SS, and NASH, GO and KEGG pathway enrichment analysis were performed. The up-regulated genes between HC and SS were too few to allow identification of GO and KEGG pathway enrichment analysis, and the up-regulated genes between HC and NASH failed to enrich pathway in KEGG.

GO analysis showed that the down-regulated genes between HC and SS were mainly involved in biological processes (BP) associated with the mesenchyme morphogenesis, organic acid transmembrane transport, smooth muscle cell proliferation, response to wounding, and regulation of MAPK cascade (Fig. 3A **and Table S3**). KEGG analysis indicated that the down-regulated genes between HC and SS primarily enriched in TGF-beta signaling pathway, MAPK signaling pathway, MicroRNAs in cancer, PI3K-Akt signaling pathway, and pathways in cancer (Fig. 3B **and Table S4**).

The DEGs between HC and NASH were mainly involved in biological processes (BP) associated with fatty acid biosynthetic process, positive regulation of T cell proliferation, extracellular matrix organization, cell-cell adhesion via plasma-membrane adhesion molecules, mesenchyme development, and transmembrane receptor protein tyrosine kinase signaling pathway (Fig. 3C **and Table S5**). KEGG analysis indicated that the DEGs between HC and NASH were primarily enriched in Jak-STAT signaling pathway, PI3K-Akt signaling pathway, and pathways in cancer (Fig. 3D **and Table S6**).

Identification Of Key Modules By Wgcna

WGCNA was performed to identify key modules related to clinical traits by using GSE89632 dataset. The power of $\beta = 5$ (scale-free $R^2 = 0.89$) was selected as the soft thresholding parameter to construct a scale-free network (Fig. 4A). A total of 24 modules were identified (Fig. 4B). Similar module clustering was

constructed by using dynamic hybrid cutting (threshold = 0.05). The results in Fig. 4C showed that the greenyellow module was the highest positive module correlated to NAFLD activity score (NAS, $R^2 = 0.79$, $p = 9e^{-10}$) and steatosis ($R^2 = 0.63$, $p = 1e^{-5}$). In addition, the midnightblue module was highly negative correlated to NAS ($R^2 = 0.64$, $p = 7e^{-6}$), and the brown module was highly negative correlated to steatosis ($R^2 = 0.61$, $p = 2e^{-5}$). Figure 4D-E showed the positive and negative modules.

In the module-trait analysis, we intersected the trait-related genes highly associated with NAS and steatosis and 45 DEGs generated from expression difference analysis, and finally extracted 25 trait-expression-related genes for the following analysis (**Table S7-S8**).

Identification of Hub Genes and Construction of Protein-Protein Interaction (PPI) Network

Subsequently, we construct a PPI network with 25 trait-expression-related genes in Metascape database. Then, 15 filtered genes were identified (Fig. 5A) and later imported into CytoHubba plugin to explore the hub genes by “MCC” methods. The results showed that MYC, TGFB3, ADAMTS1, THBS1, RASD1, PCDH20 (Down-regulated genes), MAMDC4, CYP7A1, GINS2, and PDLIM3 (Up-regulated genes) were top 10 hub genes (Fig. 5B).

Hub Genes in NAFLD were Associated with Hepatocellular Carcinoma (HCC) Prognosis

Then, we explored the possible relationship between hub genes and hepatocellular carcinoma (HCC). We found that CYP7A1, GINS2, and PDLIM3 were significantly up-regulated, and MYC, MAMDC4, ADAMTS1, THBS1, and RASD1 were significantly down-regulated in HCC tumor samples compared to normal samples using the TCGA dataset (Fig. 6A). Moreover, we found that the 8 genes above were enriched in tumor-related pathways, such as apoptosis, cell cycle, and EMT (Fig. 6B). Subsequently, we performed survival analysis in the genes above. As demonstrated in Fig. 6C, CYP7A1-high patients showed higher overall survival (OS) rates compared to CYP7A1-low patients, but had no effects on disease free survival rate (DFS). What’s more, compared to GINS2- and PDLIM3-high patients, the OS rates were higher in low expression patients. In addition, GINS2-low patients showed higher DFS rate compared to GINS2-high patients (Fig. 6D-E). In HPA database, the expression of CYP7A1/GINS2/PDLIM3 was also abnormally elevated in HCC. (Fig. 6F).

Discussion

Currently, the pathogenesis of NAFLD is still unclear and the therapeutic treatments are also limited. In the present study, we identified 45 intersected DEGs between HC-SS group and HC-NASH group, and respectively performed GO and KEGG pathway enrichment analysis to explore the potential effects of these DEGs in NAFLD. The results showed that the GO enrichments were involved in fatty acid metabolism, mesenchyme, extracellular matrix, cell adhesion, and inflammatory and immune response, which also played important roles in tumorigenesis. KEGG analysis showed that the DEGs were primarily enriched in TGF-beta signaling pathway, PI3K-Akt signaling pathway, pathways in cancer, MicroRNAs in cancer, MAPK signaling pathway, and Jak-STAT signaling pathway. Both the results of GO and KEGG

analysis all pointed to tumorigenesis. Meanwhile, the overlapping pathways between SS and NASH were PI3K-Akt signaling pathway and pathways in cancer, which suggesting that the two pathways could be an important therapeutic target for NAFLD. Our findings were also consistent with previous reports[9, 11].

WGCNA is a well-established method for studying biological networks and diseases[12]. Due to NAS and steatosis were the two main pathologic indicators in the estimation of NAFLD, we tried to find out the DEGs related to the NAS and steatosis. We totally identify 25 DEGs related to the NAS and steatosis, and PPI network analysis was performed to explore the hub genes in the pathogenesis and progression of NAFLD. Eventually, we determined 10 hub genes (Down-regulated genes: MYC, TGFB3, ADAMTS1, THBS1, RASD1, PCDH20; Up-regulated genes: MAMDC4, CYP7A1, GINS2, and PDLIM3) related to NAS and steatosis.

HCC is the fourth-leading cause of cancer death worldwide, and the morbidity of NAFLD-related HCC is predicted to increase dramatically by 2030, with increases of 82%, 117%, and 122% from 2016 in China, France, and the USA, respectively[3, 13]. Therefore, we explore whether these ten hub genes were associated with the progression in HCC in TCGA database. We found that CYP7A1, GINS2, and PDLIM3 were significantly up-regulated, and MYC, MAMDC4, ADAMTS1, THBS1, and RASD1 were significantly down-regulated in HCC tumor samples compared to normal samples. Surprisingly, we also found that CYP7A1/GINS2/PDLIM3 were correlated with HCC prognosis.

CYP7A1, catalyzing the first and rate-limiting step in the classic bile acid synthesis pathway, has been shown to be involved in the lipid metabolism[14]. Deficiency of CYP7A1 caused by homozygous deletion mutations can inhibit the production of bile acids, leading to the accumulation of cholesterol in liver, reducing LDL receptors and elevating LDL cholesterol[15]. However, CYP7A1 was up-regulated in SS and NASH group compared with HC group in our study. Previous studies shown that CYP7A1 and its associated cholesterol processes were adversely regulated in NAFLD[16], and glucose stimulates CYP7A1 transcription in human hepatocytes[17]. Therefore, up-regulating CYP7A1 in NAFLD may be the consequence rather than cause of disease[18]. In addition, increased CYP7A1 expression and bile acid synthesis ameliorated hepatic inflammation and fibrosis, which proved its effects of anti-tumor[19].

GINS2, a member of GINS family, plays a crucial role in DNA duplication and is highly expressed in various types of cancer[20, 21]. However, very little research can be found about GINS2 in the liver, especially in NAFLD. Previous bioinformatics studies indicated that GINS2 might be the hub genes in the development of NASH to HCC, and predicted poor prognosis in HCC, but there was no further experiment to verify its effects on NAFLD[22, 23].

PDLIM3, highly expressed in skeletal and cardiac muscle, has been suggested to play a pivotal role in myocyte stability, signal transduction, and mechanical signaling, especially in growth and remodeling processes[24]. Interestingly, PDLIM3 was firstly screened out for a new hub gene in the pathogenesis of NAFLD and was associated with the prognosis of HCC. PDLIM3 was highly related to epithelial-mesenchymal transition (EMT) in GSCALite database, which might partially reveal its effects in the pathogenesis in NAFLD and HCC. More future studies are needed to gain more insights about PDLIM3.

In conclusion, we analyzed two public datasets to identify DEGs among HC, SS and NASH. GO and KEGG pathway analysis revealed that the pathogenesis and progression of NAFLD was highly associated with tumorigenesis. Finally, we screened out 10 hub genes related to NAS and steatosis, and three of them were correlated with HCC prognosis. Although further validation is still needed, we provide useful and novel information to explore the potential candidate genes for NAFLD prognosis and therapeutic options.

Material And Methods

Data Retrieving and Processing

The gene expression profiles of GSE48452[25] and GSE89632[26] were downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo). To prevent the effects of overweight in the evaluation, healthy obesity with body mass index (BMI) over 24kg/m^2 were excluded from the HC group. Besides, due to NAFLD commonly happened to the obese population, NALFD patients with BMI less than 24kg/m^2 were also excluded from the experimental group. What's more, individuals with bariatric surgery or severely missing data at baseline were also ruled out. Finally, 9 SS samples 17 NASH samples, and 5 HC samples in the GSE48452, and 17 SS samples, 17 NASH samples and 8 HC samples in the GSE89632 were included in this study. HCC data were obtained from The Cancer Genome Atlas (TCGA) database, including 374 HCC samples and 50 normal samples.

For the analysis of DEGs, we used the GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/) to generate the R script, which used two R packages (GEOquery and limma). The threshold for the DEGs was set as *p*-value < 0.05 and $|\log_2$ fold change (FC) =| ≥ 1. Venn diagrams of overlaps represented the intersection between the two datasets. Figure 1 illustrated the overall research design.

Gene Ontology (GO) Analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Enrichment Analysis

To identify the function of DEGs, GO and KEGG analysis were performed using Metascape (metascape.org) database. GO analysis is primarily divided into three parts, molecular function (MF), biological process (BP), and cellular component (CC). We determined that results were statistically significant at a level of less than 0.05 using a *p*-value.

Weighted Gene Co-expression Network Analysis (Wgcna)

Considering that GSE89632 had more comprehensive and complete data, we used GSE89632 to detect modules highly correlated with NAFLD, and WGCNA was performed using R package 'WGCNA' and carried out on all genes. The scale-free topology of the networks was assessed for various values of the β shrinkage parameter, and we chose β = 5 based on scale-free topology criterion. Finally, the dynamic tree cut algorithm was applied to the dendrogram for module identification with the mini-size of module gene

numbers set as 50 and similar modules were merged following a height cutoff of 0.05. In the module-trait analysis, GS value > 0.3 and MM value > 0.55 were defined as a threshold[8].

Protein-protein Interaction (Ppi) Network Construction

Metascape (metascape.org) database was used to construct a protein-protein interaction (PPI) network. We deleted disconnected nodes in the network. Then, the Cytoscape software (v3.8.2) was utilized to visualize the PPI network. We used CytoHubba plugin to indentify the hub genes through “MCC” calculation methods[27].

Relationship between Hub Gens in NAFLD and Hepatocellular Carcinoma (HCC) Prognosis

The pathway activity was acquired from GSCALite database (http://bioinfo.life.hust.edu.cn/web/GSCALite/), the survival analysis from GEPIA (http://gepia.cancer-pku.cn/), and the immunohistochemical pictures were collected from the HPA database (https://www.proteinatlas.org/).

Statistical analysis

Statistical analysis was performed using R software (Version 4.1.0). Statistical comparisons between groups of normalized data were performed using the t-test or Mann-Whitney U-test according to the test condition. A difference with $p < 0.05$ was considered significant.

Abbreviations

NAFLD
Nonalcoholic fatty liver disease
NAFL
Nonalcoholic fatty liver
DEGs
Differentially expressed genes
SS
Simple steatosis
NASH
Nonalcoholic steatohepatitis
HC
Healthy controls
HCC
Hepatocellular carcinoma
GEO
Gene expression omnibus

BMI
Body mass index
TCGA
The Cancer Genome Atlas
GO
Gene ontology
KEGG
Kyoto encyclopedia of genes and genomes
MF
Molecular function
BP
Biological process
CC
Cellular component
WGCNA
Weighted gene co-expression network analysis
PPI
Protein-protein interaction
OS
Overall survival
DFS
Disease free survival rate

Declarations

Data Availability Statement

Publicly available datasets were analyzed in this study. This data

can be found here: GEO data base, accession number: GSE48452 and GSE89632.

Acknowledgements

The authors appreciate study investigators and staff who participated in this study.

Funding Sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contributions

Chutian Wu, Yun Zhou, and Min Wang contributed equally to this paper. Chutian Wu, Yun Zhou, and Min Wang analyzed the study data, helped draft the manuscript, made critical revisions of the manuscript. Guolin Dai, Xiongxiu Liu, and Leizhen Lai assisted with data collection and the analysis. Shaohui Tang supervised the research and edited the manuscript. All authors contributed to the article and approved the submitted version.

Ethics Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Younossi ZM. Non-alcoholic fatty liver disease - A global public health perspective. J HEPATOL. 2019;70(3):531–44.
2. Zarrinpar A, Gupta S, Maurya MR, Subramaniam S, Loomba R. Serum microRNAs explain discordance of non-alcoholic fatty liver disease in monozygotic and dizygotic twins: a prospective study. GUT. 2016;65(9):1546–54.
3. Huang DQ, El-Serag HB, Loomba R. Global epidemiology of NAFLD-related HCC: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol. 2021;18(4):223–38.
4. Natarajan SK, Ingham SA, Mohr AM, Wehrkamp CJ, Ray A, Roy S, Cazanave SC, Phillippi MA, Mott JL. Saturated free fatty acids induce cholangiocyte lipoapoptosis. HEPATOLOGY. 2014;60(6):1942–56.
5. Chen Y, Tian Z. Roles of Hepatic Innate and Innate-Like Lymphocytes in Nonalcoholic Steatohepatitis. FRONT IMMUNOL. 2020;11:1500.
6. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. HEPATOLOGY. 2016;64(1):73–84.
7. Choudhary NS, Duseja A. Genetic and epigenetic disease modifiers: non-alcoholic fatty liver disease (NAFLD) and alcoholic liver disease (ALD). Transl Gastroenterol Hepatol. 2021;6:2.
8. 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.

9. Liu J, Lin B, Chen Z, Deng M, Wang Y, Wang J, Chen L, Zhang Z, Xiao X, Chen C, et al. Identification of key pathways and genes in nonalcoholic fatty liver disease using bioinformatics analysis. ARCH MED SCI. 2020;16(2):374–85.
10. Leoni S, Tovoli F, Napoli L, Serio I, Ferri S, Bolondi L. Current guidelines for the management of non-alcoholic fatty liver disease: A systematic review with comparative analysis. World J Gastroenterol. 2018;24(30):3361–73.
11. Wang H, Liu Y, Wang D, Xu Y, Dong R, Yang Y, Lv Q, Chen X, Zhang Z. **The Upstream Pathway of mTOR-Mediated Autophagy in Liver Diseases**. *CELLS-BASEL* 2019, 8(12).
12. Rasmussen AH, Kogelman L, Kristensen DM, Chalmer MA, Olesen J, Hansen TF. Functional gene networks reveal distinct mechanisms segregating in migraine families. BRAIN. 2020;143(10):2945–56.
13. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. Nat Rev Gastroenterol Hepatol. 2019;16(10):589–604.
14. Wang Y, Gunewardena S, Li F, Matye DJ, Chen C, Chao X, Jung T, Zhang Y, Czerwinski M, Ni HM, et al. An FGF15/19-TFEB regulatory loop controls hepatic cholesterol and bile acid homeostasis. NAT COMMUN. 2020;11(1):3612.
15. Pullinger CR, Eng C, Salen G, Shefer S, Batta AK, Erickson SK, Verhagen A, Rivera CR, Mulvihill SJ, Malloy MJ, et al. Human cholesterol 7alpha-hydroxylase (CYP7A1) deficiency has a hypercholesterolemic phenotype. J CLIN INVEST. 2002;110(1):109–17.
16. Wruck W, Adjaye J. Meta-analysis reveals up-regulation of cholesterol processes in non-alcoholic and down-regulation in alcoholic fatty liver disease. World J Hepatol. 2017;9(8):443–54.
17. Chiang J, Ferrell JM. Up to date on cholesterol 7 alpha-hydroxylase (CYP7A1) in bile acid synthesis. Liver Res. 2020;4(2):47–63.
18. Jia X, Zhai T. Integrated Analysis of Multiple Microarray Studies to Identify Novel Gene Signatures in Non-alcoholic Fatty Liver Disease. Front Endocrinol (Lausanne). 2019;10:599.
19. Liu H, Pathak P, Boehme S, Chiang JL. Cholesterol 7alpha-hydroxylase protects the liver from inflammation and fibrosis by maintaining cholesterol homeostasis. J LIPID RES. 2016;57(10):1831–44.
20. Kubota Y, Takase Y, Komori Y, Hashimoto Y, Arata T, Kamimura Y, Araki H, Takisawa H. A novel ring-like complex of Xenopus proteins essential for the initiation of DNA replication. Genes Dev. 2003;17(9):1141–52.
21. Tian W, Yang X, Yang H, Zhou B. GINS2 Functions as a Key Gene in Lung Adenocarcinoma by WGCNA Co-Expression Network Analysis. Onco Targets Ther. 2020;13:6735–46.
22. Zhang D, Liu J, Xie T, Jiang Q, Ding L, Zhu J, Ye Q. Oleate acid-stimulated HMMR expression by CEBPalpha is associated with nonalcoholic steatohepatitis and hepatocellular carcinoma. INT J BIOL SCI. 2020;16(15):2812–27.

23. Lian YF, Li SS, Huang YL, Wei H, Chen DM, Wang JL, Huang YH. **Up-regulated and interrelated expressions of GINS subunits predict poor prognosis in hepatocellular carcinoma**. *Biosci Rep* 2018, 38(6).
24. Zheng M, Cheng H, Banerjee I, Chen J. ALP/Enigma PDZ-LIM domain proteins in the heart. J MOL CELL BIOL. 2010;2(2):96–102.
25. Ahrens M, Ammerpohl O, von Schonfels W, Kolarova J, Bens S, Itzel T, Teufel A, Herrmann A, Brosch M, Hinrichsen H, et al. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. CELL METAB. 2013;18(2):296–302.
26. Arendt BM, Comelli EM, Ma DW, Lou W, Teterina A, Kim T, Fung SK, Wong DK, McGilvray I, Fischer SE, et al. Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. HEPATOLOGY. 2015;61(5):1565–78.
27. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC SYST BIOL. 2014;8(Suppl 4):11.

Figures

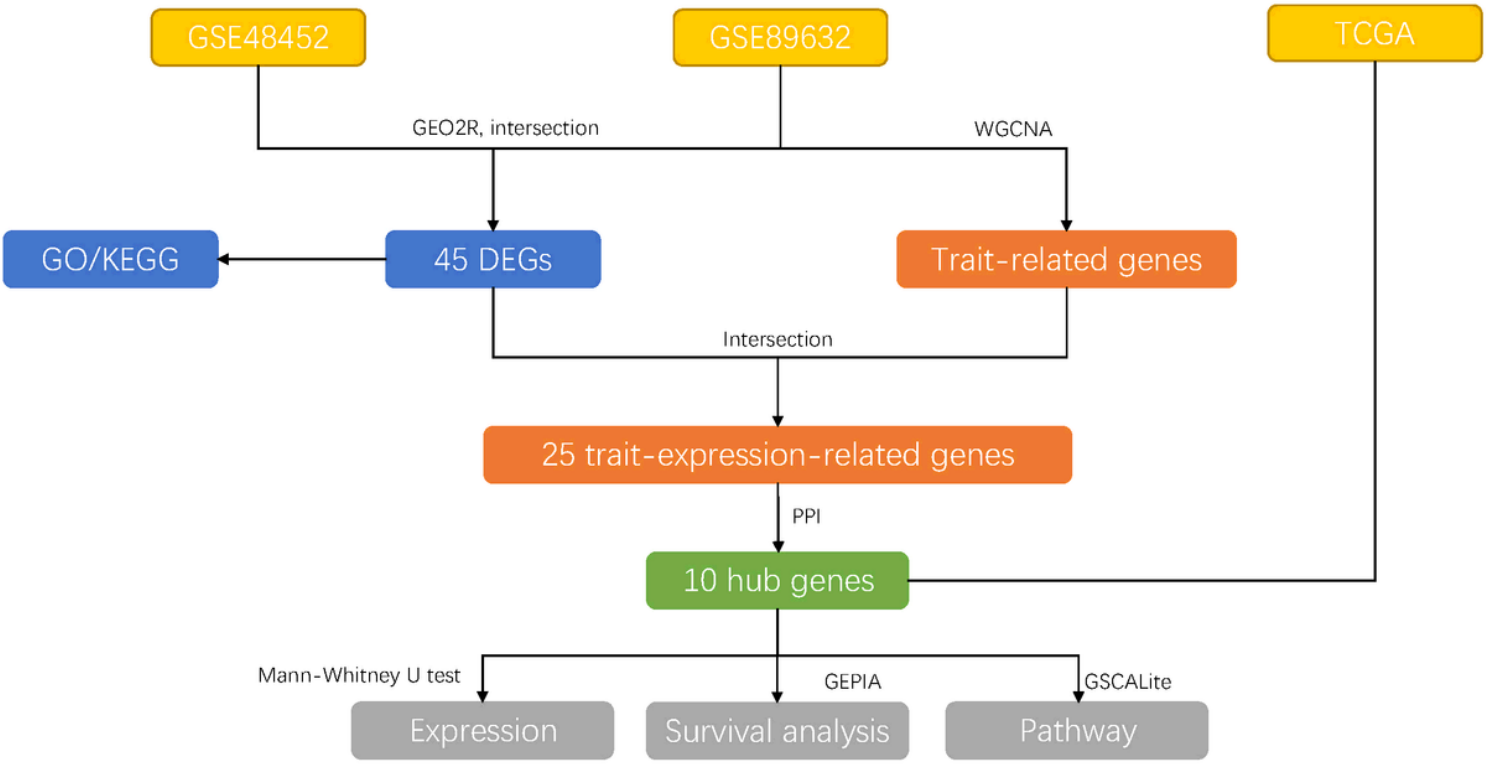

Figure 1

The overall research designs.

Figure 2

Identification of differentially expressed genes (DEGs) among HC, SS, and NASH. (**A**) Heatmap of overlapping DEGs in GSE48452; (**B**) Heatmap of overlapping DEGs in GSE89632; (**C**) Venn diagrams displayed the overlapping DEGs of up- and down- regulated genes between HC and SS; (**D**) Venn diagrams displayed the overlapping DEGs of up- and down- regulated genes between HC and NASH.

Figure 3

GO and KEGG pathway enrichment analysis. (**A**) GO analysis of DEGs among HC, SS, and NASH; (**B**) KEGG analysis of down-regulated DEGs between HC and SS; (**C**) KEGG analysis of down-regulated DEGs between HC and NASH.

Figure 4

WGCNA to identify trait-related modules and genes. (**A**) Calculating soft-thresholding power; Left: scale-free fit indices using different soft-thresholding powers; Right: mean connectivity using different soft-thresholding powers; (**B**) The dendrogram clustered by Dynamic Tree Cut algorithm; (**C**) The heatmap profiling the correlations between module eigengenes and the clinical characteristics; (**D**) Scatter plot of gene significance for NAS and steatosis (Up-regulated); (**E**) Scatter plot of gene significance for NAS and steatosis (Down-regulated).

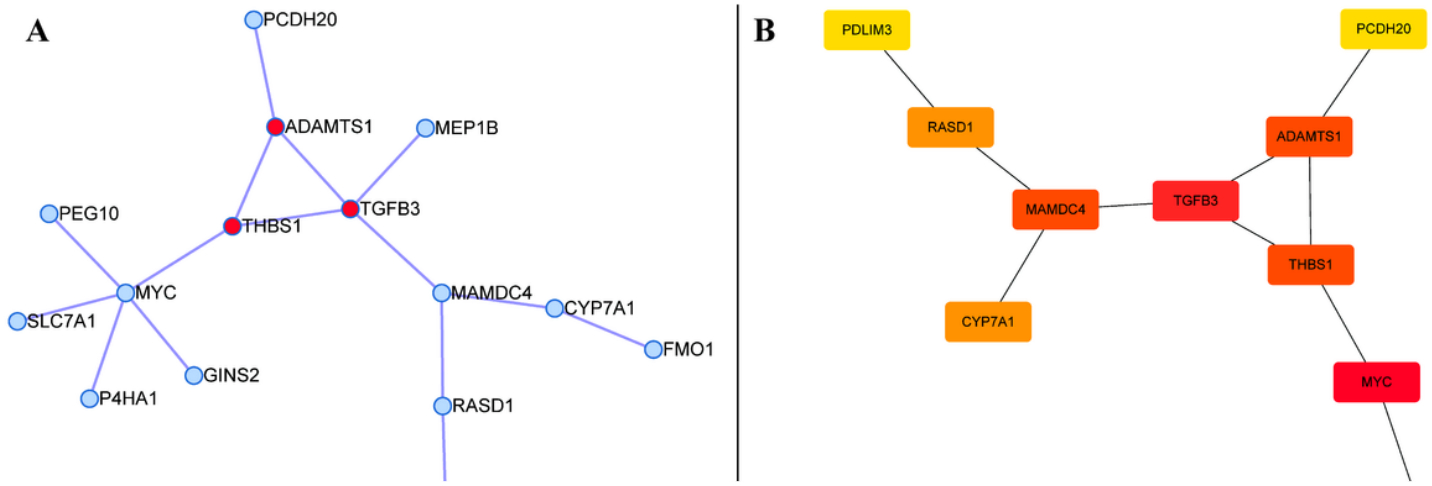

Figure 5

Construction of protein-protein interaction (PPI) networks for 25 trait-expression-related genes in Metascape database. (**A**) PPI network of 15 trait-expression-related genes. Red node represented the

Molecular Complex Detection (MCODE) algorithm applied to identify densely connected network components. (**B**) Results of CytoHubba plugin; the color changed from yellow to red was indicative of the rank of protein, and the deeper the red staining, the higher rank of protein was.

Figure 6

Expression and survival analysis of the NAFLD's hub genes in hepatocellular carcinoma (HCC). (**A**) Hub genes in NAFLD were dysregulated in hepatocellular carcinoma (ns, not significant; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$); (**B**) Enriched pathways of 10 hub genes in TCGA database. (**C-E**) Survival plots of CYP7A1, GINS2, and PDLIM3; (**F**) Protein expression of GINS2 and PDLIM3 between normal patients and HCC patients in HPA database.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.csv
- TableS2.csv
- TableS3.docx
- TableS4.docx
- TableS5.docx
- TableS6.docx
- TableS7.csv
- TableS8.csv