frontiers
in Endocrinology

ORIGINAL RESEARCH
published: 30 August 2019
doi: 10.3389/fendo.2019.00599

Integrated Analysis of Multiple Microarray Studies to Identify Novel Gene Signatures in Non-alcoholic Fatty Liver Disease

Xi Jia[1]* *and Tianyu Zhai*[2]

[1] Department of Endocrinology, Jinshan Hospital, Fudan University, Shanghai, China, [2] Department of Endocrinology and Metabolism, Zhongshan Hospital, Fudan University, Shanghai, China

OPEN ACCESS

Edited by:
Antonio Brunetti,
Università degli studi Magna Græcia di Catanzaro, Italy

Reviewed by:
Consolato Sergi,
University of Alberta Hospital, Canada
Matthew Brook,
University of Edinburgh, United Kingdom
Matthew Sinton,
University of Edinburgh, United Kingdom, in collaboration with reviewer MB

***Correspondence:**
Xi Jia
jiaxi777@hotmail.com

Specialty section:
This article was submitted to Systems Endocrinology, a section of the journal Frontiers in Endocrinology

Received: *15 May 2019*
Accepted: *14 August 2019*
Published: *30 August 2019*

Citation:
Jia X and Zhai T (2019) Integrated Analysis of Multiple Microarray Studies to Identify Novel Gene Signatures in Non-alcoholic Fatty Liver Disease. Front. Endocrinol. 10:599.
doi: 10.3389/fendo.2019.00599

Background: Non-alcoholic fatty liver disease (NAFLD) is a well-known cause of liver dysfunction and has become a common chronic liver disease in many countries. However, the intrinsic molecular mechanisms underlying the pathogenesis of NAFLD have not yet been fully elucidated.

Methods: We obtained the gene expression datasets of NAFLD through the Gene Expression Omnibus (GEO) database. Subsequently, robust rank aggregation (RRA) method was used to identify differentially expressed genes (DEGs) between NAFLD patients and controls. Gene functional annotation and PPI network analysis were performed to explore the potential function of the DEGs. Finally, we used a sequencing dataset GSE126848 to validate our results.

Results: In this study, GSE48452, GSE66676, GSE72756, GSE63067, GSE89632, and GSE107231 were included, including 125 NAFLD patients and 116 control patients. The RRA integrated analysis determined 96 significant DEGs (50 up-regulated and 46 down-regulated) and the most significant gene aberrantly expressed in NAFLD was ENO3 (*P*-value = 7.17E-05), followed by CYP7A1 (*P*-value = 9.04E-05), and P4HA1 (*P*-value = 1.67E-04). Carboxylic acid metabolic process (GO:0019752; *P*-value = 1.39E-03) was the most significantly enriched for biological process in GO (gene ontology) analysis. KEGG pathway enrichment analysis showed that steroid hormone biosynthesis (hsa00140; *P*-value = 6.68E-03) and PPAR signaling pathway (hsa03320; *P*-value = 9.95E-03) were significantly enriched. Based on the results of the PPI and the results of the RRA, we finally defined the four most critical genes as the hub genes, including ENO3, CYP7A1, P4HA1, and CYP1A1.

Conclusions: Our integrated analysis identified novel gene signatures and will contribute to the understanding of comprehensive molecular changes in NAFLD.

Keywords: non-alcoholic fatty liver disease, microarray, differentially expressed genes, robust rank aggregation, integrated analysis

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is a well-known cause of liver dysfunction over the world and is closely correlated with obesity, hyperglycemia, and hyperlipidemia (1, 2). NAFLD represents a spectrum of liver disorders that includes simple steatosis and non-alcoholic steatohepatitis (NASH), progressing to cirrhosis, and even hepatocellular carcinoma (HCC) (3). To date, NAFLD has become a common chronic liver disease in many countries, affecting almost 30% of the general population (3, 4), and is currently the third most common indication for liver transplantation in the USA (5). Some studies have attempted to elucidate the effective or causal effects of oxidative stress, adipocytokine production/release, intra-hepatic insulin resistance, fat accumulation, and innate immune system activation in the pathogenesis of NAFLD (3). However, the intrinsic molecular mechanisms underlying the pathogenesis of NAFLD have not yet been fully elucidated, and more research is needed to provide deeper understanding and to explore more advantageous therapeutic targets.

Identifying gene-specific expression patterns has been useful in the understanding of pathogenic mechanisms or therapeutic assessment for multiple diseases, NAFLD included (6, 7). In the past few years, microarray technology has been widely used for gene expression profiling in liver tissue from NAFLD patients or experimental animals. However, there are some inconsistencies in those microarray studies, such as different analysis platforms, data outliers, sample sizes, and sources. The robust rank aggregation (RRA) approach has been used to select differentially expressed mRNA profiles based on multiple datasets in various diseases, such as cancer and autoimmune disease. To the best of our knowledge, previous researches of NAFLD have not used the RRA method to identified differentially expressed genes (DEGs), which facilitated this study.

Thus, we conducted a gene expression meta-analysis between NAFLD liver tissues and control liver tissues using integrated bioinformatics methods. In addition, based on the result of this analysis, hub genes identification in the DEGs and gene enrichment and pathway annotation analysis were also performed.

MATERIALS AND METHODS

Microarray Datasets of NAFLD

We obtained the gene expression datasets of NAFLD through the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/) (8). We systematically searched the microarray studies by using the terms: "Fatty liver," "Non-alcoholic," "Gene expression," "Homo sapiens," and "Microarray." Datasets were included according to the following eligibility criteria: (1) Containing at least 10 total samples; (2) Containing at least five cases and at least five controls; (3) Raw data or gene expression profiling by array were available in GEO.

Datasets Analyses

First, we downloaded the gene expression matrix and related annotation document for each array dataset from GEO database, and used corresponding annotation document to map the microarray probes to gene symbols. If multiple probes mapped to the same symbol, the mean value was adopted. The 6 NAFLD expression microarray datasets were all standardized by quantiles. The DEGs were determined between NAFLD tissues and normal control liver tissues in each microarray by using the "limma" (linear models for microarray data) R package. The $|\log2$ fold change (FC)$| > 0.5$ and P-value < 0.05 were regarded as the cut-off criteria to determine DEGs.

RRA Analysis

To minimize the inconsistencies and to integrate the results from several microarray studies, RRA method was adopted to identify robust DEGs, which is an effective tool to integrate multiple arrays outcomes (9, 10). Before RRA analysis, we obtained up-ranked and down-ranked gene lists of each dataset which were generated by expression fold change between cases and controls. The "Robust Rank Aggregation" R package was used to integrate all the ranked gene lists of datasets. The adjusted P-value in the RRA tool indicate the possibility of ranking high of each gene in the final gene list. Genes with P-value <0.05 and the fold change >0.5 were considered as significant genes.

Functional and Pathway Enrichment Analysis

Database for Annotation, Visualization, and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov/) (11–13) is regarded as the most common functional annotation tool, which was used for gene ontology (GO) functional enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. We uploaded the significant genes in RRA analysis to investigate the potential functions. P-value <0.05 and false discovery rate (FDR) <0.05 was regarded as the cut-off criteria.

Protein-Protein Interaction (PPI) Network Analysis

Hub genes are usually deemed to be functionally critical and highly interconnected with other genes. We uploaded significant genes in RRA analysis to the STRING database (http://www.string-db.org/) and HIPPIE database (http://cbdm.uni-mainz.de/hippie/), then chose confidence >0.4 to perform the PPI network analysis. We synthesized STRING and HIPPIE methods to find the genes with the top 10 connectivity, and matched the results with the top 10 DEGs defined by RRA, and finally obtained the hub gene. The PPI network diagrams of STRING database were plotted by Cytoscape software. In the Cytoscape, each node represents a gene or protein, and the edge between nodes represents the interaction of the molecules.

RRA and Hub Genes Validation Study

Previous studies have found that arrays also have the risk of cross-hybridization, while RNA-seq data is highly replicable, thus providing a more accurate estimate of gene expression than arrays (14, 15). Therefore, we further conducted RRA validation studies by using data from a RNA-seq dataset (GSE 126848,

TABLE 1 | Characteristics of the included microarray datasets.

GSE ID	Participants	Tissues	Analysis type	Platform	Year
GSE48452	32 cases and 41 controls	Liver	Array	GPL11532	2013
GSE63067	11 cases and 7 controls	Liver	Array	GPL570	2014
GSE66676	33 cases and 34 controls	Liver	Array	GPL6244	2017
GSE72756	5 cases and 5 controls	Liver	Array	GPL16956	2015
GSE89632	39 cases and 24 controls	Liver	Array	GPL14951	2016
GSE107231	5 cases and 5 controls	Liver	Array	GPL20115	2017

including 31 NAFLD patients, 14 healthy controls). The RNA-seq reading count data was analyzed using edgeR function to identify DEGs. Genes with *P*-value <0.05 were considered to be significant.

Ethical Declaration

All of the data used in this study were obtained from public databases. This study does not contain any studies associated with animals or humans.

RESULTS

Information of Included Microarrays

According to the previously established inclusion criteria, GSE48452, GSE66676, GSE72756, GSE63067, GSE89632, and GSE107231 were included in this study. There are 125 NAFLD patients (53 patients with non-alcoholic steatohepatitis and 72 patients with simple steatosis) and 116 controls (non-NAFLD patients, including 89 normal subjects and 27 healthy obese subjects) in these six datasets. The detailed information of these datasets was shown in **Table 1**.

Identification of DEGs in NAFLD

To eliminate individual differences between samples, all six of these microarray data sets were first standardized by quantiles. The results are shown in **Supplementary Figure 1**, all samples in each dataset achieved homogeneity which was acceptable. The DEGs were screened out by using the "limma" package in R software according to the cut-off criteria. The volcano plots of the six microarrays were shown in **Figure 1**.

Results in the RRA Integrated Analysis

The RRA method assumes that each gene is randomly ordered in each dataset. The smaller the *P*-value in the RRA results, the higher gene ranks and the credibility of gene differential expression. Ninety-six significant DEGs (50 up-regulated and 46 down-regulated) were determined through the integrated analysis (**Supplementary Table 1**). The heatmap of the top 10 up- and down-regulated genes was shown in **Figure 2**. The top 10 significant gene aberrantly expressed in NAFLD included five up-regulated genes [ENO3 ($P = 7.17$E-05), CYP7A1 ($P = 9.04$E-05), FMO1 ($P = 6.57$E-04), PEG10 ($P = 8.95$E-04), and MAMDC4 ($P = 1.87$E-03)] and five down-regulated genes [P4HA1 (1.67E-04), CYP1A1 ($P = 2.51$E-04), IGFBP2 ($P = 3.27$E-04), SOCS2 ($P = 5.1$E-04), and SHBG ($P = 1.50$E-03)].

Functional Annotation

We uploaded the 96 DEGs to perform the GO (including biological process, molecular function and cellular component) analysis and KEGG analysis. The outcomes revealed that carboxylic acid metabolic process (GO:0019752; *P*-value = 1.39E-03) was the most significantly enriched for biological process, followed by monocarboxylic acid metabolic process (GO:0032787; *P*-value = 1.39E-03), oxidation-reduction process (GO: 0055114; *P*-value = 1.39E-03) and so on (**Table 2**). In terms of the molecular function and cellular component, the results were shown in **Figure 3**. Furthermore, KEGG pathway enrichment analysis showed that steroid hormone biosynthesis (hsa00140; *P*-value = 6.68E-03) and PPAR signaling pathway (hsa03320; *P*-value = 9.95E-03) were significantly enriched, as shown in **Table 3** and **Supplementary Figure 2**.

PPI Network Analysis and Identification of Hub Gene

STRING and HIPPIE database online database was used to perform PPI network analysis of the DEGs and Cytoscape software was adopted to visualize the results. In the PPI analysis, the connections between nodes were visualized to identify the interactions between the proteins encoded by DEGs in NAFLD (**Figure 4**). The gene located in the central node was considered as the hub gene which may play pivotal physiological regulatory role. In the analysis of STRING, the top 10 genes with most connections were CYP1A1, MYC, CYP7A1, IGF1, JUNB, HPGD, FOSB, ACSL1, ENO3, and CYP2A13. In the analysis of HIPPIE, the top 10 genes were MYC, ENO3, IL32, JUNB, PRKCE, P4HA1, NFIL3, TAGLN, KALRN, and AGMAT. Based on the results of the PPI and the results of the RRA, we finally defined the four most critical genes as the hub genes, including ENO3, CYP7A1, P4HA1, and CYP1A1. The visualization for hub genes in STRING database were shown in **Supplementary Figure 3**.

The Validation of RRA and Hub Genes

We validated our RRA results and hub genes using sequencing dataset GSE126848, and found that eight out of the top 10 DEGs found by RRA also had significant differences in the verification dataset. Only SHBG and MAMDC4 did not show significant differences in the validation dataset ($P > 0.05$). In the validation of the top 10 genes in PPI, we found that six of STRING method and eight of HIPPIE method were also identified as DEGs, and all hub genes have been identified as DEGS in validated datasets. The verification results were presented in **Supplementary Table 2**.

DISCUSSION

Globally, non-alcoholic fatty liver disease (NAFLD) is the most common chronic liver disease, including a range of pathologies, from benign hepatic steatosis to non-alcoholic steatohepatitis, cirrhosis, and eventually it may develop into hepatocellular carcinoma (16). NAFLD is recognized as a complex disease and the interaction between the environment and the susceptible

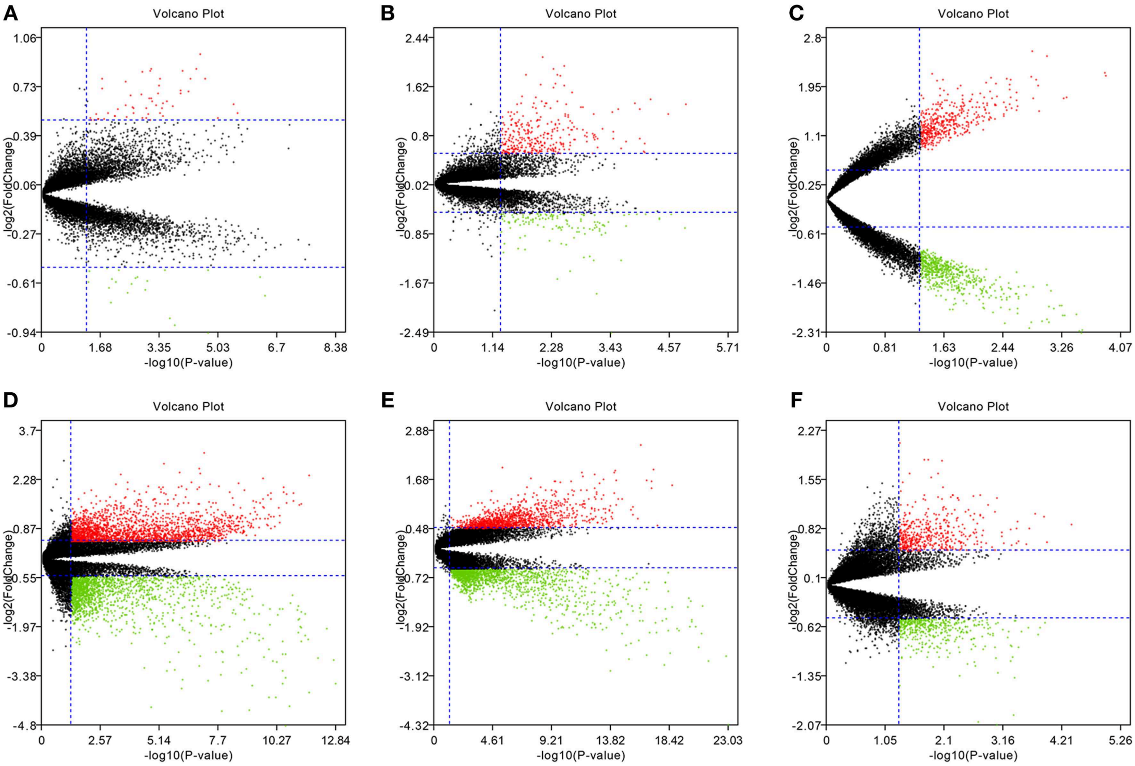
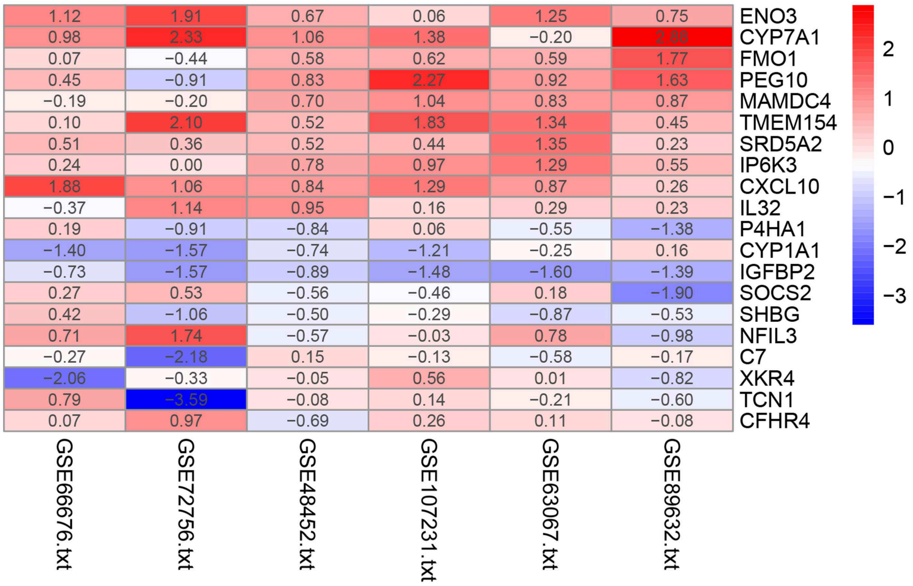


FIGURE 1 | Volcano plots of the six microarrays. Red points represented up-regulated genes, while green points represented down-regulated genes. Black points represented genes with no significant difference. **(A)** GSE48452, **(B)** GSE63067, **(C)** GSE66676, **(D)** GSE72756, **(E)** GSE89632, **(F)** GSE107231.



GSE66676.txt	GSE72756.txt	GSE48452.txt	GSE107231.txt	GSE63067.txt	GSE89632.txt	
1.12	1.91	0.67	0.06	1.25	0.75	ENO3
0.98	2.33	1.06	1.38	−0.20	2.88	CYP7A1
0.07	−0.44	0.58	0.62	0.59	1.77	FMO1
0.45	−0.91	0.83	2.27	0.92	1.63	PEG10
−0.19	−0.20	0.70	1.04	0.83	0.87	MAMDC4
0.10	2.10	0.52	1.83	1.34	0.45	TMEM154
0.51	0.36	0.52	0.44	1.35	0.23	SRD5A2
0.24	0.00	0.78	0.97	1.29	0.55	IP6K3
1.88	1.06	0.84	1.29	0.87	0.26	CXCL10
−0.37	1.14	0.95	0.16	0.29	0.23	IL32
0.19	−0.91	−0.84	0.06	−0.55	−1.38	P4HA1
−1.40	−1.57	−0.74	−1.21	−0.25	0.16	CYP1A1
−0.73	−1.57	−0.89	−1.48	−1.60	−1.39	IGFBP2
0.27	0.53	−0.56	−0.46	0.18	−1.90	SOCS2
0.42	−1.06	−0.50	−0.29	−0.87	−0.53	SHBG
0.71	1.74	−0.57	−0.03	0.78	−0.98	NFIL3
−0.27	−2.18	0.15	−0.13	−0.58	−0.17	C7
−2.06	−0.33	−0.05	0.56	0.01	−0.82	XKR4
0.79	−3.59	−0.08	0.14	−0.21	−0.60	TCN1
0.07	0.97	−0.69	0.26	0.11	−0.08	CFHR4

FIGURE 2 | Heatmap of the top 10 up- and down-regulated genes in the RRA analysis. Red represents high expression of genes in patients with NAFLD, while blue represents low expression of genes in patients with NAFLD.

multi-gene host background determines disease phenotype and progression (17). Therefore, identifying the susceptibility gene of NAFLD is very important for us to study the cause of this disease and to find potential treatments. However, there are still no reports on the use of RRA method to detect DEGs in NAFLD. RRA algorithm, a well-designed tool, has four key features: strong robustness to noise, ability to deal with incomplete ranking, giving significant scores to each element in the result ranking, and high computational efficiency (9). Our study was the first to systematically search and incorporate the microarrays on NAFLD in GEO, which was not available in previous researches. The aim of our study was to identify key genes and their pathways involved in the pathogenesis of NAFLD. In the present study, we included six microarray studies, compared gene expression profiles between NAFLD and controls, adopted the RRA analysis to integrate results with more statistical power. Furthermore, functional annotation and PPI network construction were performed to understand the potential biological function of the DEGs.

TABLE 2 | GO analysis of DEGs associated with NAFLD in the RRA analysis.

Term	Description	Gene count	*P*-value
GO.0019752	Carboxylic acid metabolic process	16	1.39E-03
GO.0032787	Monocarboxylic acid metabolic process	13	1.39E-03
GO.0055114	Oxidation-reduction process	18	1.39E-03
GO.0006082	Organic acid metabolic process	16	6.87E-03
GO.0044281	Small molecule metabolic process	25	6.87E-03
GO.0043436	Oxoacid metabolic process	15	2.12E-02
GO.0042493	Response to drug	10	3.34E-02

GO, gene ontology; DEGs, differentially expressed genes.

In all, 96 DEGs were filtered out across multiple datasets with 50 up-regulated and 46 down-regulated genes. The most significant 10 genes were ENO3, CYP7A1, P4HA1, CYP1A1, IGFBP2, SOCS2, FMO1, PEG10, SHBG, and MAMDC4. Among them, ENO3, CYP7A1, P4HA1, and CYP1A1 were identified as hub genes in PPI network analysis. Enolase 3 (ENO3) encodes the β-subunit of enolase, which is found in skeletal muscle cells in the adult where it may play a role in muscle development and regeneration. Several researches have reported that ENO3 was distributed on various tissues, such as liver, lung, skeletal

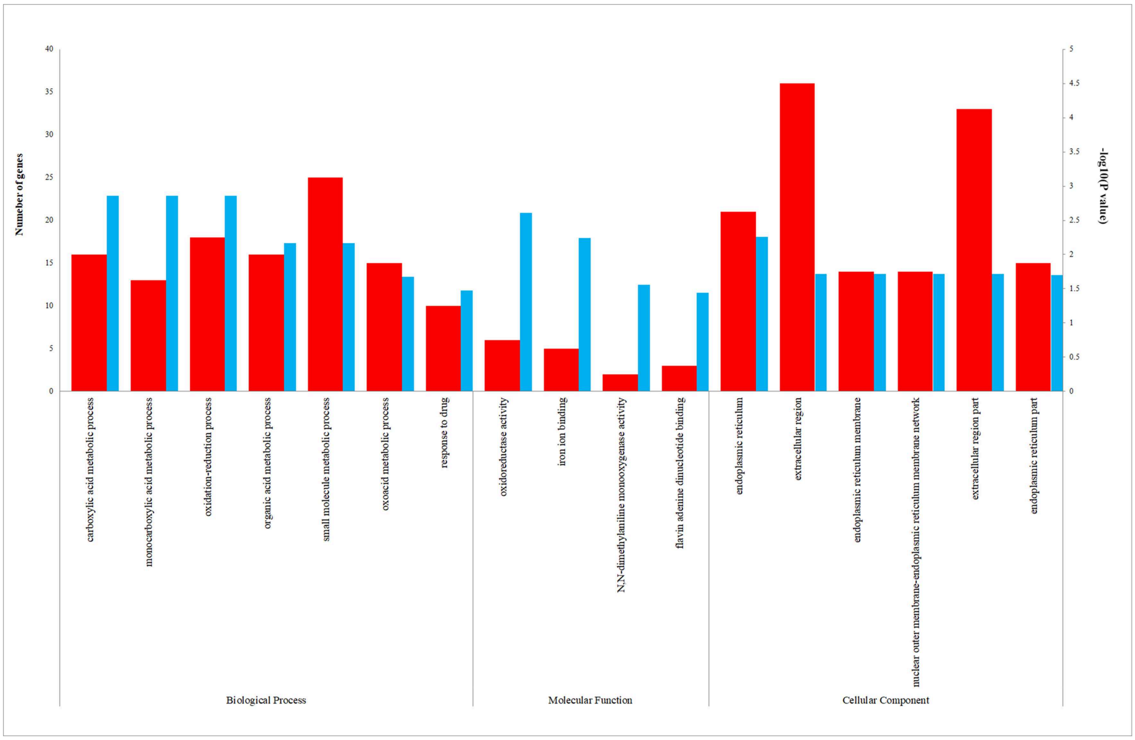


FIGURE 3 | The enriched biological process (BP), molecular function (MF), and cellular component (CC) terms in GO analysis. The red column represents the number of genes enriched in GO term. The blue column represents the $-\log_{10}$(*P*-value) of GO term.

TABLE 3 | KEGG analysis of DEGs associated with NAFLD in the RRA analysis.

Pathway	ID	Count	Genes	*P*-value
Steroid hormone biosynthesis	hsa00140	4	CYP1A1, CYP7A1, HSD17B12, SRD5A2	6.68E-03
PPAR signaling pathway	hsa03320	4	ACSL1, PLIN1, CYP7A1, FABP4	9.95E-03
p53 signaling pathway	hsa04115	3	TP53I3, IGF1, GADD45B	7.36E-02
Bile secretion	hsa04976	3	SLCO1A2, CYP7A1, ABCB4	7.74E-02

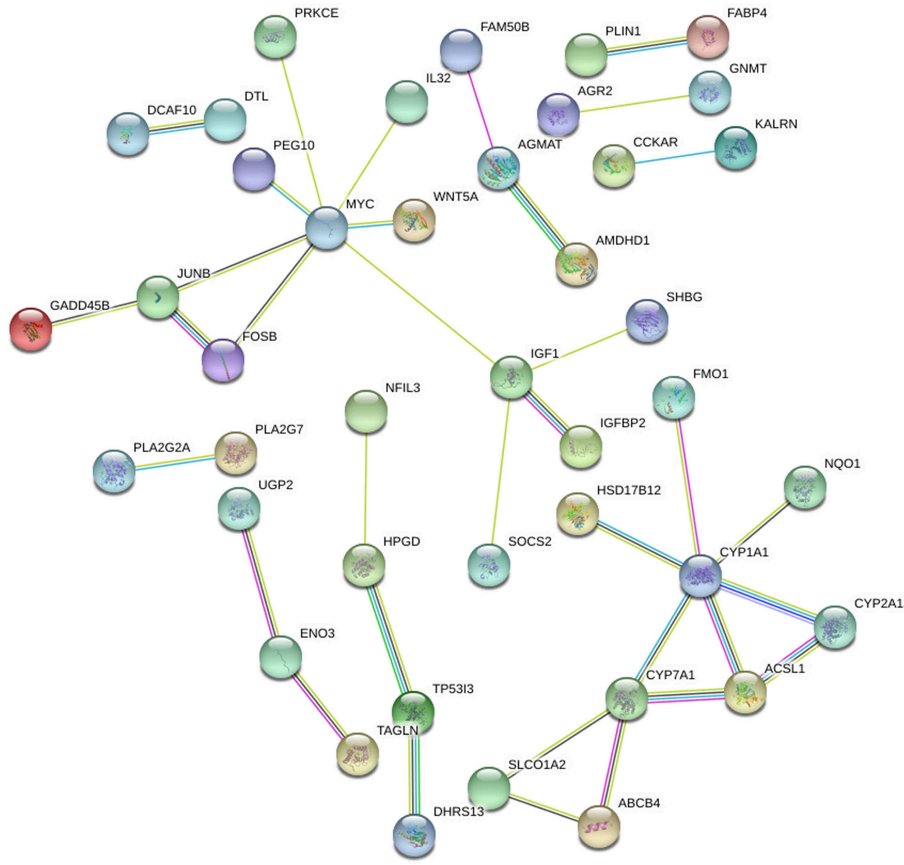


FIGURE 4 | The PPI network of DEGs in NAFLD (STRING). The top 10 genes with most connections included CYP1A1, MYC, CYP7A1, IGF1, JUNB, HPGD, FOSB, ACSL1, ENO3, and CYP2A13.

and heart (18). ENO3 mediates cholesterol ester synthesis, may increase lipid delivery to liver and accelerate hepatic cholesterol ester accumulation (19). In the previous study, Elam et al. used the gene expression profiling and found that ENO3 was significantly higher in livers of morbidly obese women compared with women who had experienced massive weight loss (20). However, the definite function and mechanisms of ENO3 in NAFLD remain unclear.

CYP7A1 and CYP1A1 are both annotated to the pathway of steroid hormone biosynthesis and encode the member of the cytochrome P450 superfamily of enzymes. Cytochrome P450 family seven subfamily A member 1 (CYP7A1) mediates cholesterol metabolism and functions as a rate-limiting enzyme to regulate the process of conversion of cholesterol into bile acids (21, 22). 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 (23). Polymorphism in the promoter of CYP7A1 could affect the synthesis of bile acids and delay the process of lipid responses to the drug fenofibrate (24). CYP7A1 deficiency in mice also showed aberrant changes in the cholesterol and bile acid transformation (25), CYP7A1 transgenic mice represented improved metabolic homeostasis in liver (26, 27). However, our multiple arrays meta-analysis demonstrated that CYP7A1 was up-regulated in NAFLD patients compared with NC. Another research revealed that CYP7A1 and its related cholesterol process were adversely regulated between non-alcoholic fatty liver disease and alcoholic liver disease (28), so it's possible that the increasing CYP7A1 levels in liver tissue are the consequence rather than cause of NAFLD.

Cytochrome P450 family 1 subfamily A member 1 (CYP1A1) encodes a monooxygenase and is widely detected for its ability to activate compounds with carcinogenic properties (29, 30). The enzyme encoded by CYP1A1 catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids, and other lipids. Uno et al. indicated that CYP1A1 has a protective role against NAFLD development through experiments *in vivo* (31). Zhu et al. studied the intersection of differentially expressed mRNAs and miRNA-predicted target genes, and then demonstrated that CYP1A1 plays important roles in NAFLD (32). Elam et al. found that CYP1A1 also differentially expressed in liver tissues of morbidly obese patients (20). Therefore, CYP1A1 has a great possibility to be an important gene of NAFLD, but its specific function needs more experimental evidence. In addition, several studies have reported that CYP1A1 polymorphisms were significant factors for the susceptibility and pathogenesis of cancer (33, 34). The contribution of CYP1A1 to cancer progression may be associated with the balance of pre-amino acid activation/detoxification and extrahepatic metabolism of dietary natural products (29).

Our study found that Prolyl-4-hydroxylase α1 (P4HA1) was a hub gene of NAFLD. P4HA1 is a P4H (also called PHD) isoenzyme and an important rate-limiting enzyme. P4H is the primary cellular oxygen sensor and regulates hypoxia-inducible factor (HIF) proteasomal degradation in an oxygen-dependent manner (16). The involvement of P4H in the pathogenesis of NAFLD has been fully proved (35, 36). In NAFLD, adipose tissue expansion and liver fat accumulation impair local oxygen homeostasis. Hypoxia signaling is also a key mechanism of adipose tissue dysfunction, leading to adipose tissue fibrosis, inflammation, and insulin resistance (37). In this case, the body is prone to tissue hypoxia-induced adaptive response. HIF acts as a major regulator of this hypoxic adaptive response, and is further activated by P4H hydroxylation (36–38). Moreover, Seda et al. also found that P4HA1 was significantly differentially expressed in the liver tissue of patients with NAFLD and liver tissue of the control group (39). In addition, it was also significantly differentially expressed in the liver tissue of morbidly obese patients (20).

In our KEGG enrichment analysis, PPAR signaling pathway was significantly enriched and its aberrant expression genes contained CYP7A1, ACSL1, PLIN1, and FABP4 in the DEGs. Per-oxisome proliferator-activated receptors (PPARs) are ligand-activated transcription factors. The PPAR pathway is widely used and involved in a variety of physiological regulation processes, including regulation of cell differentiation and development, involvement in lipid, protein, and carbohydrate metabolism, and tumorigenesis (40, 41). In recent years, PPARs have been identified to play a significant role in modulation of NAFLD due to its involvement in nutrient metabolism (42, 43). There are three PPAR isoforms: α(α), β/δ(β/δ), and γ(γ), which are differentially expressed in tissues (44). PPARα is mainly found in the liver, although expression is also present in other tissues, while PPARγ is highly expressed in adipose tissue (45, 46). It is worth noting that the expression level of PPARγ in liver tissue was significantly increased in patients with NAFLD and experimental models (47–49). Increased PPARγ activity in mouse liver may result in activation of adipogenic gene expression and increased lipid storage in the liver (43). Our study demonstrated that differential genes in liver tissue of NAFLD and control patients were enriched in the PPAR pathway, further demonstrating the importance of the PPAR pathway in the pathology of NAFLD. Additionally, p53 signaling pathway was reported to associated with insulin resistance, even though it showed borderline significance in KEGG analysis (50).

In summary, by using the RRA method we have successfully provided deeper insight to the comprehensive molecular changes in NAFLD pathogenesis, and identified several potential candidate therapeutic targets, including ENO3, CYP7A1, P4HA1, CYP1A1, IGFBP2, SOCS2, FMO1, PEG10, SHBG, and MAMDC4. Among them, ENO3, CYP7A1, P4HA1, and CYP1A1 were defined as hub genes. In addition, through GO and KEGG pathway analysis, we found that these differential genes were mainly enriched in carboxylic acid metabolic process, monocarboxylic acid metabolic process and oxidation-reduction process, and may be involved in steroid hormone biosynthesis and PPAR signaling pathway. However, the underlying molecular mechanisms have not yet been fully elucidated. In the future, more experiments are needed to verify the changes of gene expression and it is necessary to collect a large number of liver tissues from patients with NAFLD and liver tissues from normal controls for additional functional studies.

DATA AVAILABILITY

The datasets analyzed for this study can be found in the GEO datasets (https://www.ncbi.nlm.nih.gov/gds).

AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct, and intellectual contribution to the work and approved it for publication. Each author acknowledges that he or she participated sufficiently in the work to take public responsibility for its content. All authors of this paper have read and approved the final version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 81402207).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2019.00599/full#supplementary-material

Supplementary Figure 1 | Standardization of the six microarray datasets. The left panel represents the data before standardization, and the right panel represents the standardized data. **(A)** GSE48452, **(B)** GSE63067, **(C)** GSE66676, **(D)** GSE72756, **(E)** GSE89632, **(F)** GSE107231.

Supplementary Figure 2 | The KEGG pathway plots. The DEGs were plotted in red color. **(A)** the bubble chart of KEGG result, **(B)** Steroid hormone biosynthesis, **(C)** PPAR signaling pathway. KEGG pathway enrichment analysis showed that steroid hormone biosynthesis (hsa00140; *P*-value = 6.68E-03) and PPAR signaling pathway (hsa03320; *P*-value = 9.95E-03) were significantly enriched.

Supplementary Figure 3 | Visualization for hub genes by using the Cytoscape software. Nodes with warm color represent the hub genes in PPI analysis.

Supplementary Table 1 | The 96 DEGs identified by RRA method.

Supplementary Table 2 | Verification of RRA and hub genes using dataset GSE126848.

REFERENCES

1. Williams KH, Shackel NA, Gorrell MD, McLennan SV, Twigg SM. Diabetes and nonalcoholic Fatty liver disease: a pathogenic duo. *Endocr Rev.* (2013) 34:84–129. doi: 10.1210/er.2012-1009
2. Hardy T, Oakley F, Anstee QM, Day CP. Nonalcoholic fatty liver disease: pathogenesis and disease spectrum. *Annu Rev Pathol.* (2016) 11:451–96. doi: 10.1146/annurev-pathol-012615-044224
3. Cohen JC, Horton JD, Hobbs HH. Human fatty liver disease: old questions and new insights. *Science.* (2011) 332:1519–23. doi: 10.1126/science.1204265
4. Adams LA, Angulo P, Lindor KD. Nonalcoholic fatty liver disease. *CMAJ.* (2005) 172:899–905. doi: 10.1503/cmaj.045232
5. Charlton MR, Burns JM, Pedersen RA, Watt KD, Heimbach JK, Dierkhising RA. Frequency and outcomes of liver transplantation for nonalcoholic steatohepatitis in the United States. *Gastroenterology.* (2011) 141:1249–53. doi: 10.1053/j.gastro.2011.06.061
6. Murphy SK, Yang H, Moylan CA, Pang H, Dellinger A, Abdelmalek MF, et al. Relationship between methylome and transcriptome in patients with nonalcoholic fatty liver disease. *Gastroenterology.* (2013) 145:1076–87. doi: 10.1053/j.gastro.2013.07.047
7. Moylan CA, Pang H, Dellinger A, Suzuki A, Garrett ME, Guy CD, et al. Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology.* (2014) 59:471–82. doi: 10.1002/hep.26661
8. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.* (2013) 41:D991–5. doi: 10.1093/nar/gks1193
9. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics.* (2012) 28:573–80. doi: 10.1093/bioinformatics/btr709
10. Yan S, Wang W, Gao G, Cheng M, Wang X, Wang Z, et al. Key genes and functional coexpression modules involved in the pathogenesis of systemic lupus erythematosus. *J Cell Physiol.* (2018) 233:8815–25. doi: 10.1002/jcp.26795
11. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Gene Ontol Consortium Nat Genet.* (2000) 25:25–9. doi: 10.1038/75556
12. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* (2003) 4:P3. doi: 10.1186/gb-2003-4-9-r60
13. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* (2012) 40:D109–14. doi: 10.1093/nar/gkr988
14. Zhang W, Yu Y, Hertwig F, Thierry-Mieg J, Zhang W, Thierry-Mieg D, et al. Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biol.* (2015) 16:133. doi: 10.1186/s13059-015-0694-1
15. Peters TJ, French HJ, Bradford ST, Pidsley R, Stirzaker C, Varinli H, et al. Evaluation of cross-platform and interlaboratory concordance via consensus modelling of genomic measurements. *Bioinformatics.* (2019) 35:560–70. doi: 10.1093/bioinformatics/bty675
16. Sinton MC, Hay DC, Drake AJ. Metabolic control of gene transcription in non-alcoholic fatty liver disease: the role of the epigenome. *Clin Epigenet.* (2019) 11:104. doi: 10.1186/s13148-019-0702-5
17. Eslam M, Valenti L, Romeo S. Genetics and epigenetics of NAFLD and NASH: clinical impact. *J Hepatol.* (2018) 68:268–79. doi: 10.1016/j.jhep.2017.09.003
18. Wu J, Zhou D, Deng C, Wu X, Long L, Xiong Y. Characterization of porcine ENO3: genomic and cDNA structure, polymorphism and expression. *Genet Sel Evol.* (2008) 40:563–79. doi: 10.1051/gse:2008015
19. Stahlberg D, Rudling M, Angelin B, Bjorkhem I, Forsell P, Nilsell K, et al. Hepatic cholesterol metabolism in human obesity. *Hepatology.* (1997) 25:1447–50. doi: 10.1002/hep.510250623
20. Elam MB, Cowan GS Jr, Rooney RJ, Hiler ML, Yellaturu CR, Deng X, et al. Hepatic gene expression in morbidly obese women: implications for disease susceptibility. *Obesity.* (2009) 17:1563–73. doi: 10.1038/oby.2009.49
21. Beigneux A, Hofmann AF, Young SG. Human CYP7A1 deficiency: progress and enigmas. *J Clin Invest.* (2002) 110:29–31. doi: 10.1172/JCI0216076
22. Davis RA, Miyake JH, Hui TY, Spann NJ. Regulation of cholesterol-7alpha-hydroxylase: BAREly missing a SHP. *J Lipid Res.* (2002) 43:533–43.
23. Pullinger CR, Eng C, Salen G, Shefer S, Batta AK, Erickson SK, et al. Human cholesterol 7α-hydroxylase (CYP7A1) deficiency has a hypercholesterolemic phenotype. *J Clin Invest.* (2002) 110:109–17. doi: 10.1172/JCI0215387
24. Shen J, Arnett DK, Parnell LD, Lai CQ, Straka RJ, Hopkins PN, et al. The effect of CYP7A1 polymorphisms on lipid responses to fenofibrate. *J Cardiovasc Pharmacol.* (2012) 59:254–9. doi: 10.1097/FJC.0b013e31823de86b
25. Jones RD, Lopez AM, Tong EY, Posey KS, Chuang JC, Repa JJ, et al. Impact of physiological levels of chenodeoxycholic acid supplementation on intestinal and hepatic bile acid and cholesterol metabolism in Cyp7a1-deficient mice. *Steroids.* (2015) 93:87–95. doi: 10.1016/j.steroids.2014.11.002
26. Li T, Owsley E, Matozel M, Hsu P, Novak CM, Chiang JY. Transgenic expression of cholesterol 7alpha-hydroxylase in the liver prevents high-fat diet-induced obesity and insulin resistance in mice. *Hepatology.* (2010) 52:678–90. doi: 10.1002/hep.23721
27. Li T, Matozel M, Boehme S, Kong B, Nilsson LM, Guo G, et al. Overexpression of cholesterol 7alpha-hydroxylase promotes hepatic bile acid synthesis and secretion and maintains cholesterol homeostasis. *Hepatology.* (2011) 53:996–1006. doi: 10.1002/hep.24107
28. 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) 9:443–54. doi: 10.4254/wjh.v9.i8.443
29. Androutsopoulos VP, Tsatsakis AM, Spandidos DA. Cytochrome P450 CYP1A1: wider roles in cancer progression and prevention. *BMC Cancer.* (2009) 9:187. doi: 10.1186/1471-2407-9-187
30. Mescher M, Haarmann-Stemmann T. Modulation of CYP1A1 metabolism: from adverse health effects to chemoprevention and therapeutic options. *Pharmacol Ther.* (2018) 187:71–87. doi: 10.1016/j.pharmthera.2018.02.012
31. Uno S, Nebert DW, Makishima M. Cytochrome P450 1A1 (CYP1A1) protects against nonalcoholic fatty liver disease caused by Western diet containing benzo[a]pyrene in mice. *Food Chem Toxicol.* (2018) 113:73–82. doi: 10.1016/j.fct.2018.01.029
32. Zhu M, Wang Q, Zhou W, Liu T, Yang L, Zheng P, et al. Integrated analysis of hepatic mRNA and miRNA profiles identified molecular networks and potential biomarkers of NAFLD. *Sci Rep.* (2018) 8:7628. doi: 10.1038/s41598-018-25743-8
33. Girdhar Y, Singh N, Behera D, Sharma S. Synergistic association of CYP1A1 polymorphisms with increased susceptibility to squamous cell lung cancer in north Indian smokers. *Int J Biol Markers.* (2016) 31:e402–12. doi: 10.5301/jbm.5000221
34. Liu HX, Li J, Ye BG. Correlation between gene polymorphisms of CYP1A1, GSTP1, ERCC2, XRCC1, and XRCC3 and susceptibility to lung cancer. *Genet Mol Res.* (2016) 15:1–10. doi: 10.4238/gmr15048813
35. Kivirikko KI, Helaakoski T, Tasanen K, Vuori K, Myllyla R, Parkkonen T, et al. Molecular biology of prolyl 4-hydroxylase. *Ann N Y Acad Sci.* (1990) 580:132–42. doi: 10.1111/j.1749-6632.1990.tb17925.x
36. Koivunen P, Kietzmann T. Hypoxia-inducible factor prolyl 4-hydroxylases and metabolism. *Trends Mol Med.* (2018) 24:1021–35. doi: 10.1016/j.molmed.2018.10.004
37. Lefere S, Van Steenkiste C, Verhelst X, Van Vlierberghe H, Devisscher L, Geerts A. Hypoxia-regulated mechanisms in the pathogenesis of obesity and non-alcoholic fatty liver disease. *Cell Mol Life Sci.* (2016) 73:3419–31. doi: 10.1007/s00018-016-2222-1
38. Gorres KL, Raines RT. Prolyl 4-hydroxylase. *Crit Rev Biochem Mol Biol.* (2010) 45:106–24. doi: 10.3109/10409231003627991
39. Seda O, Cahova M, Mikova I, Sedova L, Dankova H, Heczkova M, et al. Hepatic gene expression profiles differentiate steatotic and non-steatotic grafts in liver transplant recipients. *Front Endocrinol.* (2019) 10:270. doi: 10.3389/fendo.2019.00270
40. Costa V, Foti D, Paonessa F, Chiefari E, Palaia L, Brunetti G, et al. The insulin receptor: a new anticancer target for peroxisome proliferator-activated receptor-gamma (PPARgamma) and thiazolidinedione-PPARgamma agonists. *Endocr Relat Cancer.* (2008) 15:325–35. doi: 10.1677/ERC-07-0226
41. Bougarne N, Weyers B, Desmet SJ, Deckers J, Ray DW, Staels B, et al. Molecular actions of PPARalpha in lipid metabolism and inflammation. *Endocr Rev.* (2018) 39:760–802. doi: 10.1210/er.2018-00064
42. Gross B, Pawlak M, Lefebvre P, Staels B. PPARs in obesity-induced T2DM, dyslipidaemia and NAFLD. *Nat Rev Endocrinol.* (2017) 13:36–49. doi: 10.1038/nrendo.2016.135

43. Liss KH, Finck BN. PPARs and nonalcoholic fatty liver disease. *Biochimie.* (2017) 136:65–74. doi: 10.1016/j.biochi.2016.11.009
44. Tailleux A, Wouters K, Staels B. Roles of PPARs in NAFLD: potential therapeutic targets. *Biochim Biophys Acta.* (2012) 1821:809–18. doi: 10.1016/j.bbalip.2011.10.016
45. Motojima K, Passilly P, Peters JM, Gonzalez FJ, Latruffe N. Expression of putative fatty acid transporter genes are regulated by peroxisome proliferator-activated receptor alpha and gamma activators in a tissue- and inducer-specific manner. *J Biol Chem.* (1998) 273:16710–4. doi: 10.1074/jbc.273.27.16710
46. Wahli W, Michalik L. PPARs at the crossroads of lipid signaling and inflammation. *Trends Endocrinol Metab.* (2012) 23:351–63. doi: 10.1016/j.tem.2012.05.001
47. Yu S, Matsusue K, Kashireddy P, Cao WQ, Yeldandi V, Yeldandi AV, et al. Adipocyte-specific gene expression and adipogenic steatosis in the mouse liver due to peroxisome proliferator-activated receptor gamma1 (PPARgamma1) overexpression. *J Biol Chem.* (2003) 278:498–505. doi: 10.1074/jbc.M210062200
48. Nakamuta M, Kohjima M, Morizono S, Kotoh K, Yoshimoto T, Miyagi I, et al. Evaluation of fatty acid metabolism-related gene expression in nonalcoholic fatty liver disease. *Int J Mol Med.* (2005) 16:631–5. doi: 10.3892/ijmm.16.4.631
49. Pettinelli P, Videla LA. Up-regulation of PPAR-gamma mRNA expression in the liver of obese patients: an additional reinforcing lipogenic mechanism to SREBP-1c induction. *J Clin Endocrinol Metab.* (2011) 96:1424–30. doi: 10.1210/jc.2010-2129
50. Zhang X, Duan W, Lee WN, Zhang Y, Xiang F, Liu Q, et al. Overexpression of p53 improves blood glucose control in an insulin resistant diabetic mouse model. *Pancreas.* (2016) 45:1010–7. doi: 10.1097/MPA.0000000000000637

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.