



# Gene expression profiling reveals key genes and pathways related to the development of non-alcoholic fatty liver disease

Ruifeng Wang,\* Xiaobing Wang,\*\* Liwei Zhuang\*

\* Gastroenterology and Hepatology Department, the Forth Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang province, China.

\*\* Gastroenterology and Hepatology Department, the First Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang province, China.

## ABSTRACT

**Background.** This study aims to identify key genes and pathways involved in non-alcoholic fatty liver disease (NAFLD). **Material and methods.** The dataset GSE48452 was downloaded from Gene Expression Omnibus, including 14 control liver samples, 27 healthy obese samples, 14 steatosis samples and 18 nonalcoholic steatohepatitis (NASH) samples. Differentially expressed genes (DEGs) between controls and other samples were screened through LIMMA package. Then pathway enrichment analysis for DEGs was performed by using DAVID, and alterations of enriched pathways were determined. Furthermore, protein-protein interaction (PPI) networks were constructed based on the PPI information from HPRD database, and then, networks were visualized through Cytoscape. Additionally, interactions between microRNAs (miRNAs) and pathways were analyzed via Fisher's exact test. **Results.** A total of 505, 814 and 783 DEGs were identified for healthy obese, steatosis and NASH samples in comparison with controls, respectively. DEGs were enriched in ribosome (*RPL36A*, *RPL14*, etc.), ubiquitin mediated proteolysis (*UBE2A*, *UBA7*, etc.), focal adhesion (*PRKCA*, *EGFR*, *CDC42*, *VEGFA*, etc.), FcγR-mediated phagocytosis (*PRKCA*, *CDC42*, etc.), and so on. The 27 enriched pathways gradually deviated from baseline (namely, controls) along with the changes of obese-steatosis-NASH. In PPI networks, *PRKCA* interacted with *EGFR* and *CDC42*. Besides, hsa-miR-330-3p and hsa-miR-126 modulated focal adhesion through targeting *VEGFA* and *CDC42*. **Conclusions.** The identified DEGs (*PRKCA*, *EGFR*, *CDC42*, *VEGFA*), disturbed pathways (ribosome, ubiquitin mediated proteolysis, focal adhesion, FcγR-mediated phagocytosis, etc.) and miRNAs (hsa-miR-330-3p, hsa-miR-126, etc.) might be closely related to NAFLD progression. These results might contribute to understanding NAFLD mechanism, conducting experimental researches, and designing clinical practices.

**Key words.** Non-Alcoholic Steatohepatitis. Differentially expressed genes. Pathway. Network. microRNA.

## INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is the most common cause of liver disease worldwide with a prevalence of 15-30% in Western populations.<sup>1-4</sup> In patients with NAFLD, nearly one quarter is nonalcoholic steatohepatitis (NASH),<sup>5</sup> which is defined as the presence of hepatic steatosis and inflammation with hepatocyte injury (ballooning) with or without fibrosis.<sup>6</sup> Many conditions, including obesity and location of fat storage, play a role in the pathogenesis of NAFLD, especially NASH.<sup>7</sup> However, the molecular mechanism underlying NAFLD is still vague.

In the past years, NAFLD-related genes and microRNAs (miRNAs) have been discovered. Previous studies have shown elevated levels of C-reactive protein (CRP), interleukin-6 (IL-6) and vascular endothelial growth factor (VEGF) in NASH patients compared with controls.<sup>8,9</sup> The genetic variant I148M (rs738409) in Patatin-like phospholipase domain-containing 3 (PNPLA3) affects the secretion of hepatic very low density lipoproteins (VLDL) and promotes intracellular lipid accumulation in the liver by reducing the lipidation of VLDL, facilitating NAFLD development.<sup>10</sup> Besides, miRNA-10b has been proven to regulate steatosis level by targeting peroxisome proliferator-activated receptors (PPAR)- $\alpha$  expression in

NAFLD.<sup>11</sup> It's reported that apoptosis-related p53 overexpression activated miR-34a/SIRT1/p53 pathway in NAFLD morbid obese patients.<sup>12</sup>

Along with the development of high-throughput technology, microarray analysis and next-generation sequencing have been widely utilized to identify candidate disease-related genes and pathways. Based on RNA sequencing analysis of Shp(-/-) mice, novel gene signatures implicated in NASH have been identified and validated, including tetraspanin 4 (TSPAN4), peptidoglycan recognition protein 2 (PGLYRP2), thrombospondin 1 (THBS1), dual specific phosphatase-4 (DUSP4), etc.<sup>13</sup> Through gene expression profiling and gene set enrichment analysis, Yoneda, *et al.* have found genes involved in NASH, including 27 up-regulated gene sets and 25 down-regulated gene sets.<sup>14</sup> By analyzing gene expression profiles in high- and low-risk NAFLD patients, Moylan, *et al.* has found a 64-gene profile that can differentiate severe NAFLD from mild NAFLD and an independent 20-gene subset that correlates with NAFLD severity.<sup>15</sup> Through array-based DNA methylation and mRNA expression profiling analyses of liver samples from healthy obese patients, steatosis patients, NASH patients, and controls, Ahrens, *et al.*<sup>16</sup> have discovered NAFLD-specific mRNA expression and DNA methylation of nine genes coding for key enzymes in intermediate metabolism (including *PC*, *ACLY* and *PLCG1*) and insulin/insulin-like signaling (including *IGF1*, *IGFBP2* and *PRKCE*). It's also found that NAFLD-associated methylation changes were partially reversible. However, the pathways changed in NAFLD progression, and the miRNAs that regulate these altered pathways are still unclear.

In this study, the microarray dataset GSE48452 up-loaded by Ahrens, *et al.*<sup>16</sup> was further analyzed to screen differentially expressed genes (DEGs) between healthy obese samples, steatosis samples, NASH samples, and controls. Then, pathway enrichment analysis was performed for DEGs, and variation of enriched pathways at different statuses was detected. Furthermore, protein-protein interac-

tion (PPI) networks were constructed. Additionally, miRNAs targeting DEGs enriched in pathways were predicted. These results might be conducive to further experimental studies and provide better understanding of the nosogenesis of NAFLD.

## MATERIAL AND METHODS

### Gene expression data

The gene expression data set GSE48452 was downloaded from Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>), including 14 control liver samples (group C), 27 healthy obese samples (group H), 14 steatosis samples (group S) and 18 NASH samples (group N). It should be noted that only gene expression data were downloaded, and thus the sample size and information of enrolled individuals are different from the previous article of Ahrens, *et al.* In addition, information on sample number, age, body mass index, sex, fat area, inflammation, fibrosis, and non-alcoholic fatty liver disease activity score were collected from GEO (Table 1). Gene expression levels were measured by using [HuGene-1\_1-st] Affymetrix Human Gene 1.1 ST Array [transcript (gene) version] (Affymetrix Inc., Santa Clara, California, USA). For the gene expression experiment, all patients provided written and informed consent, and its protocol was approved by the institutional review board ("Ethikkommission der Medizinischen Fakultät der Universität Kiel," D425/07, A111/99).<sup>16</sup>

### DEGs screening

Linear Models for Microarray Data (LIMMA) package<sup>17</sup> of Bioconductor (available at <http://www.bioconductor.org/packages/release/bioc/html/>) was used to identify genes that were differentially expressed between healthy obese and control samples (H *vs.* C), between steatosis and control samples (S *vs.* C), and between

**Table 1.** Overview for normal controls, healthy obese samples, steatosis samples, and NASH samples.

	Normal controls (C)	Healthy obese (H)	Steatosis (S)	NASH (N)
Number	14	27	14	18
Age	52 (23-77)	45 (33-60)	42 (24-65)	45 (32-58)
BMI	25 (17-31)	41 (29-56)	48 (40-60)	49 (24-70)
Sex (% male)	36	7	29	22
Fat (area in %)	1 (0-3)	2 (0-7)	36 (10-70)	72 (40-90)
Inflammation (0-3)	0 (0-1)	0 (0-1)	0 (0-1)	2 (1-3)
Fibrosis (0-4)	0 (0-2)	0 (0-1)	0 (0-1)	1 (0-4)
NAS (0-8)	0 (0-1)	0 (0-2)	2 (1-3)	5 (4-7)

The median and the interquartile range are provided for all numeric parameters. NASH: nonalcoholic steatohepatitis. BMI: body mass index. NAS: non-alcoholic fatty liver disease activity score.

NASH and control samples (N *vs.* C). Because the DEGs with  $|\log_2 \text{fold change (FC)}| > 1$  were rare,  $|\log_2 \text{FC}| \neq 0$  and  $p\text{-value} < 0.01$  were chosen as the cut-off criteria.

### Pathway enrichment analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched by the three sets of DEGs were revealed by using the Database for Annotation, Visualization and Integrated Discovery (DAVID, available at <http://david.abcc.ncifcrf.gov/>).<sup>18</sup> The  $p\text{-value} < 0.1$  was set as the cut-off criterion.

### Scoring altered pathways

Three sets of significantly enriched pathways (H *vs.* C, S *vs.* C, and N *vs.* C) were combined, and a total of N pathways were obtained. The deviations of every pathway at status H, status S, and status N from status C were calculated based on gene expression levels.<sup>19</sup> Firstly, DEGs from the gene pool of a specific pathway were selected out. Secondly,  $p\text{-value}$  was converted to Z-score (Z) by using inverse cumulative distribution function. Finally, an alteration score of a pathway, namely A (P), was determined for pathway P based on DEG Z-scores and following algorithm:

$$A(P) = \max \frac{1}{\sqrt{t}} \sum_{i=1}^t Z_i \mid 1 < t < k$$

$$A_{\text{corrected}} = \frac{A(P) - \mu_k}{\sigma_k}$$

Several issues should be taken into account. First, not all DEGs involved in a pathway had significant influence on this pathway. For example, changes of downstream genes might be induced by abnormalities of upstream genes. In other words, changes of upstream genes involved in pathway could have more significant influence on pathways. Second, number of genes varies in different pathways. Thus, random perturbation treatment was applied to eliminate the influence of pathway size. During calculation, Z-score of DEGs was sorted in descending order. Bigger Z-score meant higher degree of differential expression. Assuming that pathway P contained k DEGs, when t genes ( $1 < t < k$ ) had maximum average Z-score, these t genes were considered to have the most significant contribution to the pathway P, and then, A (P) was also calculated. In order to eliminate the influence of pathway size, normalization of each A (P) to  $A_{\text{corrected}}(P)$  score was accomplished by correcting background distributions based on the mean and standard deviation which were derived from A(P) scores of 10,000 random sets of genes.

### Construction of protein-protein interaction networks

PPIs were obtained from Human Protein Reference Database (HPRD, available at <http://www.hprd.org/>)<sup>20</sup> via Python program.<sup>21</sup> If a protein normally expressed in human interacted with at least three DEGs, this protein would be selected to construct PPI networks together with DEGs. Then, PPI networks were visualized by using Cytoscape (available at <http://cytoscape.org/>).<sup>22</sup> Moreover, in order to disclose changes of the networks at different statuses, topological properties of these networks were analyzed by using network analysis (a plugin in Cytoscape), such as degree, average shortest path length (ASPL), clustering coefficient (CC) and topological coefficient (TC).

### Prediction of miRNAs associated with significant pathways

The common regulatory relationships between miRNAs and DEGs were abstracted from three databases, including miRecord,<sup>23</sup> miRTarBase<sup>24</sup> and Tarbase 6.0.<sup>25</sup> A total of 5,489 interactions were obtained, containing 482 miRNAs and 2,331 target genes. Then, Fisher's exact test<sup>26</sup> was adopted to identify pairs of miRNA and pathway with significant relevance. Fisher's exact test examines whether the proportion of target genes (namely, target genes of miRNA) in the gene group is same to the proportion of target genes in the entire genome. There are two hypotheses: first, whether the gene belongs to target gene group or not; second, whether the gene belongs to functional gene group (for example, group of pathway genes). Fisher's score was defined as the probability of at least x genes can be identified in functional gene group in the K target genes.

$$p = 1 - \sum_{i=0}^{x-1} \frac{\binom{M}{i} \binom{N-M}{K-i}}{\binom{N}{K}}$$

N represents the number of total genes in the whole genome; M represents the number of genes in a functional genes group; and K represents the number of target genes. Finally, altered pathways at different statuses were identified, as well as the miRNAs regulating the corresponding pathways.

## RESULTS

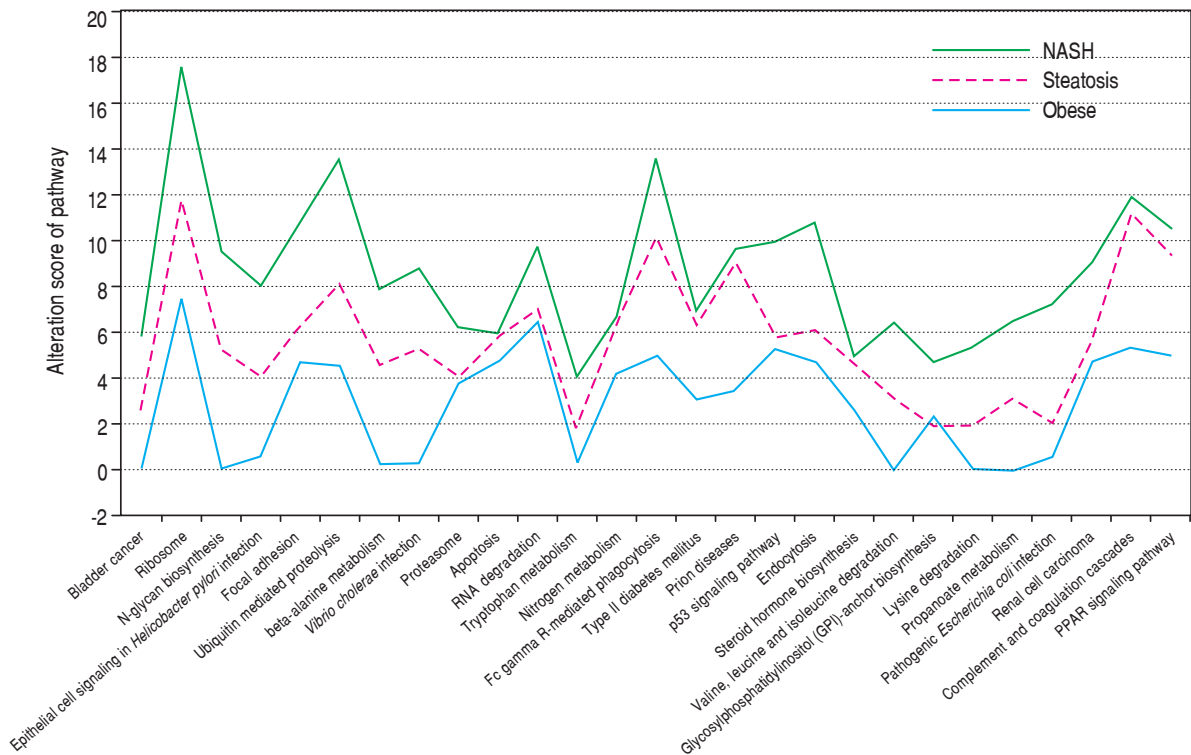
### Identification of DEGs

According to the threshold ( $p\text{-value} < 0.01$  and  $|\log_2 \text{FC}| \neq 0$ ), a total of 505 DEGs (including 205 up-regulated

**Table 2.** Pathway enrichment analysis result for the DEGs between healthy obese, steatosis, NASH, and controls.

Group (term)	Count	p-value	DEGs
<b>Obese</b>			
hsa00410: beta-alanine metabolism	5	0.001	CNDP1, EHHADH, ALDH2, HIBCH, DPYD
hsa00310: lysine degradation	5	0.017	EHHADH, WHSC1L1, ALDH2, ACAT1, BBOX1
hsa05120: epithelial cell signaling in pylori infection	6	0.010	EGFR, CDC42, IL8, ATP6V0D2, ATP6V0B, ATP6V1F
hsa00640: propanoate metabolism	4	0.036	EHHADH, ALDH2, HIBCH, ACAT1
hsa05110: <i>Vibrio cholerae</i> infection	5	0.039	PRKCA, ATP6V0D2, ATP6V0B, SEC61G, ATP6V1F
hsa05130: pathogenic <i>Escherichia coli</i> infection	5	0.041	PRKCA, CDC42, TUBA8, KRT18, ARPC4
hsa00380: tryptophan metabolism	4	0.063	EHHADH, ALDH2, CAT, ACAT1
hsa04666: FcγR-mediated phagocytosis	6	0.066	PRKCA, DNMT3, CDC42, WASF3, ARPC4, CRK
hsa05219: Bladder cancer	4	0.071	EGFR, CCND1, IL8, MYC
hsa00280: valine, leucine and isoleucine degradation	4	0.079	EHHADH, ALDH2, HIBCH, ACAT1
hsa00510: N-glycan biosynthesis	4	0.088	MGAT5B, DPM2, DPAGT1, ALG14
hsa04120: ubiquitin mediated proteolysis	7	0.094	UBE2N, CBLC, UBE2I, FBXO4, ITCH, UBE2L3, UBE2B
<b>Steatosis</b>			
hsa03010: ribosome	27	0.001	RPL17, RPL36A, RPL14, RPL15, RPL27A, RPS27L, RPS25, RPS27, RPL30, MRPL13...
hsa04120: ubiquitin mediated proteolysis	14	0.004	UBE2A, UBA7, SKP1, UBE2B, RBX1, UBE2N, CUL2, UBE2D2, CUL7, UBA2...
hsa03050: proteasome	7	0.012	PSMA1, PSMC6, PSMD12, PSMA6, PSMA4, POMP, PSMD6
hsa05211: renal cell carcinoma	8	0.026	CDC42, CUL2, PAK2, VEGFA, RAP1B, TCEB1, CRK, RBX1
hsa04510: focal adhesion	15	0.041	PRKCA, PARVG, EGFR, FLT4, MYL12B, MYL12A, PPP1CB, KDR, CDC42, VEGFA
hsa04115: p53 signaling pathway	7	0.063	CCNB3, ZMAT3, TSC2, MDM2, RRM2B, CCNG1, GADD45A
hsa04144: endocytosis	13	0.083	EGFR, PLD2, RET, TSG101, ADRBK1, KDR, CDC42...
hsa00563: (GPI)-anchor biosynthesis	4	0.084	PIGG, PIGX, PIGW, PIGA
hsa03018: RNA degradation	6	0.088	WDR61, EXOSC4, LSM6, LSM3, CNOT7, LSM1
<b>NASH</b>			
hsa03010: ribosome	18	0.001	RPL36A, RPL15, RPS8, RPS7, RPL18A, RPS3A, RPL22, RPL9, FAU, RPS13...
hsa04610: complement and coagulation cascades	11	0.001	C8B, FGG, C9, MASP2, C6, C5, TFPI, SERPINC1, C4BPB, C4BPA...
hsa00140: steroid hormone biosynthesis	7	0.018	AKR1C2, HSD17B2, CYP7A1, SRD5A1, SRD5A2, UGT2A3, AKR1C1
hsa04666: FcγR-mediated phagocytosis	10	0.030	PLCG1, LIMK1, CFL2, PIP5K1C, ARPC5, INPP5D, VAV2, PRKCE, PRKCD, SYK
hsa04930: type II diabetes mellitus	6	0.064	SOCS2, PKLR, IKBKB, PRKCE, INSR, PRKCD
hsa04210: apoptosis	5	0.075	GTF2E2, TAF7, GTF2A2, TAF9B, GTF2B
hsa05020: prion diseases	5	0.075	NCAM2, C8B, C9, C6, C5
hsa00910: nitrogen metabolism	4	0.087	CTH, CA14, CA12, HAL
hsa03320: PPAR signaling pathway	7	0.097	ME1, LPL, PLIN1, CYP7A1, SCD, FADS2, FABP4

DEGs: differentially expressed genes. NASH: nonalcoholic steatohepatitis.



**Figure 1.** The line chart displaying the deviation of the 27 pathways at three statuses. Blue line: healthy obese; red line: steatosis; green line: NASH. The 27 biological pathways are in X-axis and alteration score of the pathway is in Y-axis. NASH: nonalcoholic steatohepatitis.

and 300 down-regulated DEGs), 814 DEGs (including 283 up-regulated and 531 down-regulated DEGs), and 783 DEGs (including 392 up-regulated and 391 down-regulated DEGs) were identified between healthy obese and control samples (H *vs.* C), between steatosis and control samples (S *vs.* C), and between NASH and control samples (N *vs.* C), respectively.

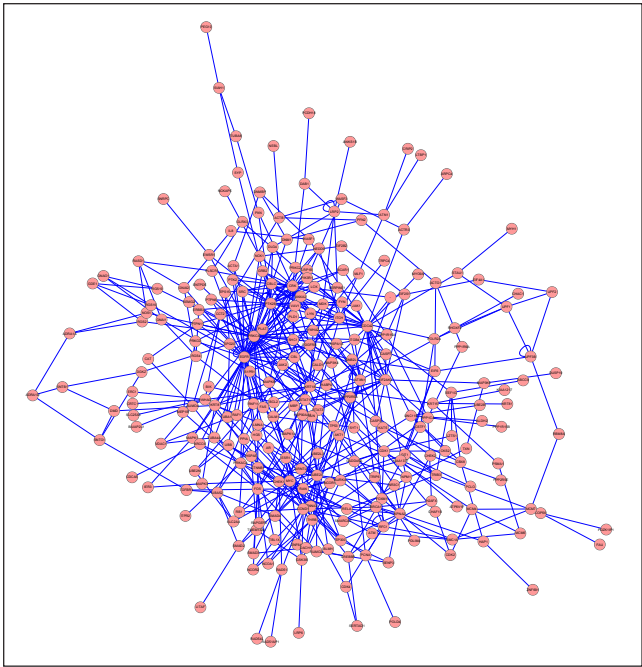
### Pathway enrichment analysis for DEGs

In H *vs.* C contrast group, DEGs were significantly enriched in 6 KEGG pathways, such as FcγR-mediated phagocytosis (*PRKCA*, *CDC42*, etc.), ubiquitin mediated proteolysis (*UBE2N*, *UBE2B*, etc.), and so on. In S *vs.* C contrast group, DEGs were markedly enriched in ribosome (*RPL36A*, *RPL14*, etc.), ubiquitin mediated proteolysis (*UBE2A*, *UBA7*, etc.), focal adhesion (*PRKCA*, *EGFR*, *CDC42*, *VEGFA*, etc.), and so on. Meanwhile, in N *vs.* C contrast group, DEGs were significantly enriched in ribosome (*RPL36A*, *RPL15*, etc.), FcγR-mediated phagocytosis (*PRKCE*, *PRKCD*, etc.), and so on (Table 2).

### Alterations of pathways

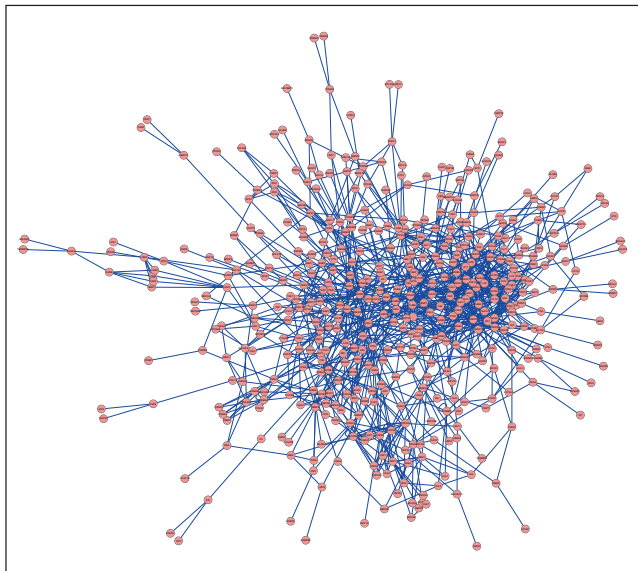
After combining the above three sets of pathways enriched by DEGs, a total of 27 significant pathways were ob-

tained. Taking normal controls as the baseline (Figure 1), alteration score was determined for each pathway. On the

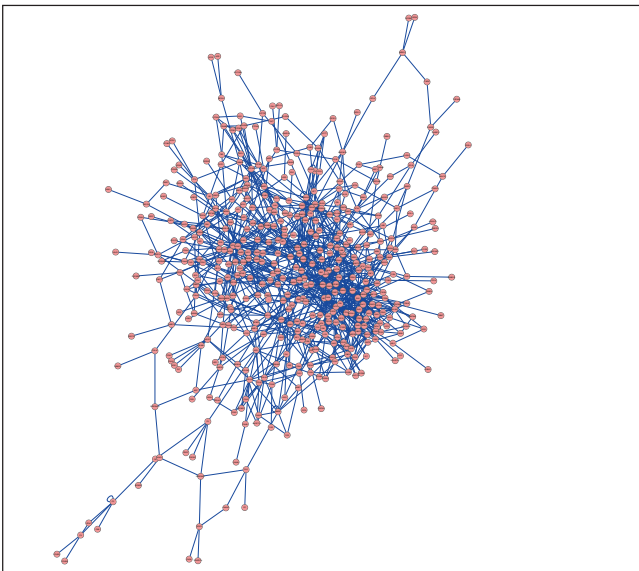


**Figure 2.** The protein-protein interaction network for the differentially expressed genes between healthy obese samples and normal controls.





**Figure 3.** The protein-protein interaction network for the differentially expressed genes between steatosis samples and normal controls.



**Figure 4.** The protein-protein interaction network for the differentially expressed genes between NASH samples and normal controls. NASH: nonalcoholic steatohepatitis.

**Table 3.** The top 10 genes in the PPI networks of obese, steatosis, and NASH.

Group	Gene	Degree	Gene	Degree
Obese	<i>EGFR</i>	49	<i>CDC42</i>	22
	<i>PRKCA</i>	40	<i>KRT18</i>	19
	<i>PTK2B</i>	29	<i>EIF2AK2</i>	18
	<i>CRK</i>	26	<i>MYC</i>	18
	<i>UBE2I</i>	24	<i>ATXN1</i>	17
Steatosis	<i>EGFR</i>	62	<i>PTK2B</i>	29
	<i>PRKCA</i>	47	<i>KDR</i>	27
	<i>CRK</i>	36	<i>CDC42</i>	26
	<i>MDM2</i>	31	<i>PPP2CA</i>	23
	<i>HDAC1</i>	30	<i>PAK2</i>	23
NASH	<i>PRKCD</i>	51	<i>HDAC1</i>	23
	<i>PLCG1</i>	48	<i>PPP2CA</i>	23
	<i>SYK</i>	41	<i>MET</i>	20
	<i>INSR</i>	34	<i>KRT18</i>	20
	<i>KDR</i>	27	<i>INPP5D</i>	20

PPI: protein-protein interaction. NASH: nonalcoholic steatohepatitis.

whole, all of the 27 pathways were gradually deviated from the baseline along with the change of normal-obesity-steatosis-NASH. Some pathways showed significant deviation from the baseline and obvious changes along with the change of normal-obesity-steatosis-NASH, such as ribosome, ubiquitin mediated proteolysis, focal adhesion and FcγR-mediated phagocytosis.

### PPI networks analysis

The PPI network for DEGs in H *vs.* C contrast group consisted of 254 genes and 551 interactions (Figure 2), and

some genes had high degrees, such as *EGFR* (degree = 49) and *PRKCA* (degree = 40) (Table 3). Besides, in the network for DEGs in S *vs.* C contrast group, there were 464 genes and 1,059 interactions (Figure 3), and some genes had high degrees, such as *EGFR* (degree = 62), *PRKCA* (degree = 47) (Table 3). Meanwhile, in these two networks, *PRKCA* interacted with numerous genes, such as *EGFR* and *CDC42*. Furthermore, the network for DEGs in N *vs.* C contrast group contained 471 genes and 995 interactions (Figure 4), and some genes had higher degrees, such as *PRKCD* (degree = 51) and *PLCG1* (degree = 48) (Table 3). Additionally, along with the change of obesi-

ty-steatosis-NASH, the average degrees of PPI network decreased, ASPLs of PPI network increased, CCs and TCs of PPI network exhibited small declines (Table 4).

### miRNAs related to significant pathways

Some miRNAs were identified to regulate DEGs enriched in a set of significant pathways. For instance, hsa-miR-126 and hsa-miR-330-3p regulated focal adhesion via targeting *VEGFA* and *CDC42* (Table 5).

## DISCUSSION

In the present study, gene expression profiles of healthy obese, steatosis and NASH samples were compared to healthy controls, and a total of 505, 814 and 784 DEGs were identified, respectively. According to the pathway deviation analysis, obvious changes of pathways enriched by DEGs along with the change of normal-obesity-steatosis-NASH were observed, such as ribosome (*RPL36A*, *RPL14*, etc.), ubiquitin mediated proteolysis (*UBE2A*, *UBE2B*, etc.), focal adhesion (*PRKCA*, *EGFR*, *CDC42*, *VEGFA*, etc.), and FcγR-mediated phagocytosis (*PRKCA*, *CDC42*, etc.). Reportedly, obesity is a risk factors of NAFLD,<sup>7</sup> and NASH is always developed from steatosis. Therefore, it's supposed that obesity develops into steatosis, steatosis develops into

NASH, and these developments are associated with the deviation of the 27 pathways.

Among these pathways, ubiquitin mediated proteolysis and ribosome were both associated with protein metabolism. In the liver with non-alcoholic steatosis or steatohepatitis, inactivation of the ubiquitin-proteasome pathway mediates cell death caused by oxidative stress.<sup>27</sup> In this study, *UBE2A* (ubiquitin-conjugating enzyme E2A) and *UBA7* (ubiquitin-like modifier activating enzyme 7) were demonstrated to take part in ubiquitin mediated proteolysis. Thus, *UBE2A* and *UBA7* might be involved in NAFLD through ubiquitin mediated proteolysis.

There is no relevant evidence showing ribosome is associated with NASH or NAFLD, while ribosome is closely related to apoptosis,<sup>28</sup> which is a pathological feature of NAFLD and promote the development from simple steatosis to NASH.<sup>29</sup> In the present study, *RPL36A* (ribosomal protein L36α) and *RPL14* (ribosomal protein L14) were found to participate in ribosome pathway. In addition, the expression levels of *RPL36A* were significantly down-regulated in healthy obese samples ( $\log_2$  FC = -0.186), steatosis samples ( $\log_2$  FC = -0.421), and NASH samples ( $\log_2$  FC = -0.340), in comparison with normal controls; the expression levels of *RPL14* were significantly down-regulated in healthy obese samples ( $\log_2$  FC = -0.104), steatosis samples ( $\log_2$  FC = -0.221), and NASH samples ( $\log_2$  FC = -0.187), in comparison with normal controls. The down-regulation of *RPL36A* and *RPL14* might further promote cell death. Therefore, these genes might play a role in NAFLD through ribosome pathway.

Focal adhesion is crucial in liver disease, because the stage of simple steatosis progressing over steatohepatitis to fibrosing steatohepatitis is characterised by adhesion of leucocytes to the sinusoidal endothelium.<sup>30</sup> In this study, *PRKCA* (protein kinase C α), *EGFR* (epidermal growth factor receptor), *CDC42* (cell division cycle 42) and *VEGFA*

**Table 4.** Topological properties of the 3 networks.

Network features	Obese	Steatosis	NASH
Degree	4.338	4.267	3.869
CC	0.059	0.052	0.048
ASPL	3.881	4.251	4.419
TC	0.265	0.256	0.254

CC: clustering coefficient. ASPL: average shortest path length. TC: topological coefficient. NASH: nonalcoholic steatohepatitis. The value is a mean for all genes in each network.

**Table 5.** The top 10 miRNA-target-pathway with the highest significant relevance.

miRNA	KEGG pathway	p-value	DEGs
hsa-miR-7	hsa04144: endocytosis	0.001555	<i>EGFR</i>
hsa-miR-34c-5p	hsa05219: bladder cancer	0.001578	<i>MYC</i>
hsa-miR-106b	hsa04120: ubiquitin mediated proteolysis	0.001630	<i>ITCH</i>
hsa-miR-29b	hsa05211: renal cell carcinoma	0.001702	<i>CDC42</i> , <i>VEGFA</i>
hsa-miR-7	hsa05219: bladder cancer	0.001719	<i>EGFR</i>
hsa-miR-126	hsa04510: focal adhesion	0.001878	<i>VEGFA</i> , <i>CRK</i>
hsa-miR-330-3p	hsa04510: focal adhesion	0.001941	<i>VEGFA</i> , <i>CDC42</i>
hsa-miR-15a	hsa05219: bladder cancer	0.002370	<i>CCND1</i>
hsa-miR-29a	hsa04144: endocytosis	0.002834	<i>CDC42</i> , <i>RET</i>
hsa-miR-34a	hsa05219: bladder cancer	0.003064	<i>CCND1</i>

miRNA: microRNA. KEGG: Kyoto Encyclopedia of Genes and Genomes. DEGs: differentially expressed genes. The genes enriched in pathways were regulated by the corresponding miRNAs.

(vascular endothelial growth factor A) were significantly enriched in focal adhesion. The expression levels of *PRKCA* were significantly up-regulated in healthy obese samples ( $\log_2$  FC = 0.210), steatosis samples ( $\log_2$  FC = 0.253), and NASH samples ( $\log_2$  FC = 0.063), in comparison with normal controls. In the PPI networks for DEGs of H vs. C contrast group and S vs. C contrast group, *PRKCA* had high degrees and interacted with *EGFR* and *CDC42*. The study of Yang, *et al.* has shown that *EGFs* can stimulate proliferation of hepatic stellate cells,<sup>31</sup> which is the primary effector cell, orchestrating the deposition of extracellular matrix (ECM) in fibrotic liver.<sup>32</sup> In this study, the expression levels of *EGFR* were significantly up-regulated in healthy obese samples ( $\log_2$  FC = 0.224), steatosis samples ( $\log_2$  FC = 0.243), and NASH samples ( $\log_2$  FC = 0.059), in comparison with normal controls. Besides, *CDC42* participates in the activation of saturated fatty acid-induced c-Jun N-terminal kinase (JNK) in hepatocytes,<sup>33</sup> which is evident in NASH patients.<sup>34,35</sup> In addition, VEGFs play an important role in the development of liver fibrosis and hepatocarcinogenesis in NASH, coordinating with leptin-mediated neovascularization.<sup>36</sup> In this study, the expression levels of *VEGFA* were significantly up-regulated in healthy obese, steatosis, and NASH samples, in comparison with normal controls. This result was consistent with the study (at serum level) of Coulon, *et al.*<sup>37</sup> Hence, *PRKCA*, *EGFR*, *CDC42*, and *VEGFA* might participate in the development of NAFLD through focal adhesion.

In the progression of NAFLD, phagocytic dysfunction is aggravated.<sup>38</sup> In this study, DEGs *PRKCA*, *CDC42*, *PRKCE* and *PRKCD* were enriched in FcγR-mediated phagocytosis. It should be noted that *PRKCA*, *PRKCE* (protein kinase C ε), and *PRKCD* (protein kinase C δ) belong to the same protein family, and thus share similar bio-functions in liver.<sup>39</sup> As discussed above, *PRKCA*, *CDC42*, *PRKCE*, and *PRKCD* might participate in the development of NAFLD through FcγR-mediated phagocytosis.

In our study, it's also discovered that *CDC42* and *VEGFA* were regulated by hsa-miR-330-3p, and *VEGFA* was regulated by hsa-miR-126, in the pathway of focal adhesion. Reportedly, the expression of hsa-miR-330-3p is altered in NAFLD,<sup>40</sup> and miR-126 is down-regulated in type 2 diabetes,<sup>41</sup> which is commonly associated with NASH.<sup>42</sup> Also, miR-126 can target insulin receptor substrate-1 (IRS-1),<sup>43</sup> which is correlated with the severity of liver damage in patients with NAFLD.<sup>44</sup> Thereby, hsa-miR-330-3p and hsa-miR-126 might play pivotal roles in the development of NAFLD via regulating focal adhesion through *VEGFA* or *CDC42*.

In summary, a number of DEGs (including *PRKCA*, *EGFR*, *CDC42*, *VEGFA*, etc.), disturbed pathways (including ribosome, ubiquitin mediated proteolysis, FcγR-

mediated phagocytosis, focal adhesion, etc.) and several miRNAs (including hsa-miR-330-3p, hsa-miR-126, etc.) were revealed in the development of NAFLD. The results of this study might be conducive to the further understanding of NAFLD mechanism, and provide a theoretical basis for further experimental researches and clinical practices. For example, the key genes and miRNAs identified in this study might serve as therapeutic targets for the treatment of NAFLD after generating pre-clinical evidences.

## DISCLOSURES OF CONFLICT OF INTEREST

The authors declare that they have no conflict of interests.

## AUTHORS' CONTRIBUTIONS

RFW proposed the study. RFW and XW performed research and wrote the first draft. SR collected and analyzed the data. All authors contributed to the design and interpretation of the study and to further drafts. LWZ is the guarantor.

## ACKNOWLEDGEMENTS

This work was financially supported by the Natural Science Foundation of Heilongjiang Province of China (No. D201139).

## REFERENCES

1. Bedogni G, Miglioli L, Masutti F, Tiribelli C, Marchesini G, Bellentani S. Prevalence of and risk factors for nonalcoholic fatty liver disease: the Dionysos nutrition and liver study. *Hepatology* 2005; 42: 44-52.
2. Browning JD, Szczepaniak LS, Dobbins R, Nuremberg P, Horton JD, Cohen JC, Grundy SM, et al. Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology* 2004; 40: 1387-95.
3. Clark JM, Diehl AM. Hepatic steatosis and type 2 diabetes mellitus. *Curr Diab Rep* 2002; 2: 210-5.
4. Vernon G, Baranova A, Younossi ZM. Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. *Aliment Pharmacol Ther* 2011; 34: 274-85.
5. Adams LA LK. Non-alcoholic fatty liver disease. *Ann Epidemiol* 2007; 17: 6.
6. Chalasani N, Younossi Z, Lavine JE, Diehl AM, Brunt EM, Cusi K, Charlton M, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. *Hepatology* 2012; 55: 2005-23.
7. Mirza MS. Obesity, Visceral Fat, and NAFLD: Querying the Role of Adipokines in the Progression of Nonalcoholic Fatty Liver Disease. *ISRN Gastroenterol* 2011; 2011: 592404.



8. Yoneda M, Mawatari H, Fujita K, Iida H, Yonemitsu K, Kato S, Takahashi H, et al. High-sensitivity C-reactive protein is an independent clinical feature of nonalcoholic steatohepatitis (NASH) and also of the severity of fibrosis in NASH. *J Gastroenterol* 2007; 42: 573-82.
9. Tarantino G, Conca P, Pasanisi F, Ariello M, Mastrolia M, Arena A, Tarantino M, et al. Could inflammatory markers help diagnose nonalcoholic steatohepatitis? *Eur J Gastroenterol Hepatol* 2009; 21: 504-11.
10. Pirazzi C, Adiels M, Burza MA, Mancina RM, Levin M, Stahlman M, Taskinen MR, et al. Patatin-like phospholipase domain-containing 3 (PNPLA3) I148M (rs738409) affects hepatic VLDL secretion in humans and in vitro. *J Hepatol* 2012; 57: 1276-82.
11. Zheng L, Lv Gc, Sheng J, Yang Yd. Effect of miRNA-10b in regulating cellular steatosis level by targeting PPAR $\alpha$  expression, a novel mechanism for the pathogenesis of NAFLD. *J Gastroenterol Hepatol* 2010; 25: 156-63.
12. Castro RE, Ferreira D, Afonso MB, Borralho PM, Machado MV, Cortez-Pinto H, Rodrigues CM. miR-34a/SIRT1/p53 is suppressed by ursodeoxycholic acid in the rat liver and activated by disease severity in human non-alcoholic fatty liver disease. *J Hepatol* 2013; 58: 119-25.
13. Smalling RL, Delker DA, Zhang Y, Nieto N, McGuinness MS, Liu S, Friedman SL, et al. Genome-wide transcriptome analysis identifies novel gene signatures implicated in human chronic liver disease. *Am J Physiol Gastrointest Liver Physiol* 2013; 305: G364-G374.
14. Yoneda M, Endo H, Mawatari H, Nozaki Y, Fujita K, Akiyama T, Higurashi T, et al. Gene expression profiling of non-alcoholic steatohepatitis using gene set enrichment analysis. *Hepatol Res* 2008; 38: 1204-12.
15. Moylan CA, Pang H, Dellinger A, Suzuki A, Garrett ME, Guy CD, Murphy SK, et al. Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology* 2014; 59: 471-82.
16. Ahrens M, Ammerpohl O, von Schonfels W, Kolarova J, Bens S, Itzel T, Teufel A, et al. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab* 2013; 18: 296-302.
17. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004; 3.
18. Jiao X, Sherman BT, Huang da W, Stephens R, Baseler MW, Lane HC, Lempicki RA. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics* 2012; 28: 1805-6.
19. Nam H, Lee J, Lee D. Computational identification of altered metabolism using gene expression and metabolic pathways. *Biotechnol Bioeng* 2009; 103: 835-43.
20. Goel R, Muthusamy B, Pandey A, Prasad TS. Human protein reference database and human proteinpedia as discovery resources for molecular biotechnology. *Mol Biotechnol* 2011; 48: 87-95.
21. Sanner MF. Python: a programming language for software integration and development. *J Mol Graph Model* 1999; 17: 57-61.
22. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; 13: 2498-504.
23. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res* 2009; 37: D105-10.
24. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, Tsai WT, et al. miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res* 2011; 39: D163-9.
25. Vergoulis T, Vlachos IS, Alexiou P, Georgakilas G, Maragkakis M, Reczko M, Gerangelos S, et al. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res* 2012; 40: D222-9.
26. Bishara AJ, Hittner JB. Testing the significance of a correlation with non-normal data: comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychol Methods* 2012; 17: 399-417.
27. Joshi-Barve S, Barve SS, Butt W, Klein J, McClain CJ. Inhibition of proteasome function leads to NF- $\kappa$ B-independent IL-8 expression in human hepatocytes. *Hepatology* 2003; 38: 1178-87.
28. Naora H. Involvement of ribosomal proteins in regulating cell growth and apoptosis: translational modulation or recruitment for extraribosomal activity? *Immunol Cell Biol* 1999; 77: 197-205.
29. Feldstein AE, Canbay A, Angulo P, Tanaii M, Burgart LJ, Lindor KD, Gores GJ. Hepatocyte apoptosis and fas expression are prominent features of human nonalcoholic steatohepatitis. *Gastroenterology* 2003; 125: 437-43.
30. McCuskey RS, Ito Y, Robertson GR, McCuskey MK, Perry M, Farrell GC. Hepatic microvascular dysfunction during evolution of dietary steatohepatitis in mice. *Hepatology* 2004; 40: 386-93.
31. Yang C, Zeisberg M, Mosterman B, Sudhakar A, Yerramalla U, Holthaus K, Xu L, et al. Liver fibrosis: insights into migration of hepatic stellate cells in response to extracellular matrix and growth factors. *Gastroenterology* 2003; 124: 147-59.
32. Lee UE, Friedman SL. Mechanisms of hepatic fibrogenesis. *Best Pract Res Clin Gastroenterol* 2011; 25: 195-206.
33. Sharma M, Urano F, Jaeschke A. Cdc42 and Rac1 are major contributors to the saturated fatty acid-stimulated JNK pathway in hepatocytes. *J Hepatol* 2012; 56: 192-8.
34. Czaja MJ. JNK regulation of hepatic manifestations of the metabolic syndrome. *Trends Endocrinol Metab* 2010; 21: 707-13.
35. Ferreira D, Castro R, Machado M, Evangelista T, Silvestre A, Costa A, Coutinho J, et al. Apoptosis and insulin resistance in liver and peripheral tissues of morbidly obese patients is associated with different stages of non-alcoholic fatty liver disease. *Diabetologia* 2011; 54: 1788-98.
36. Kitade M, Yoshiji H, Kojima H, Ikenaka Y, Noguchi R, Kaji K, Yoshii J, et al. Leptin-mediated neovascularization is a prerequisite for progression of nonalcoholic steatohepatitis in rats. *Hepatology* 2006; 44: 983-91.
37. Coulon S, Francque S, Colle I, Verrijken A, Blomme B, Heindryckx F, De Munter S, et al. Evaluation of inflammatory and angiogenic factors in patients with non-alcoholic fatty liver disease. *Cytokine* 2012; 59: 442-9.
38. Cheong H, Lee SS, Lee JS, Kim J, Kim SW, Lee WJ. Phagocytic function of Kupffer cells in mouse nonalcoholic fatty liver disease models: Evaluation with superparamagnetic iron oxide. *J Magn Reson Imaging* 2014.
39. Bezy O, Tran TT, Pihlajamaki J, Suzuki R, Emanuelli B, Winay J, Mori MA, et al. PKC $\delta$  regulates hepatic insulin sensitivity and hepatosteatosis in mice and humans. *J Clin Invest* 2011; 121: 2504-17.
40. Estep M, Armistead D, Hossain N, Elarainy H, Goodman Z, Baranova A, Chandhoke V, et al. Differential expression of miRNAs in the visceral adipose tissue of patients with non-alcoholic fatty liver disease. *Aliment Pharmacol Ther* 2010; 32: 487-97.

41. Zampetaki A, Kiechl S, Drozdov I, Willeit P, Mayr U, Prokopi M, Mayr A, et al. Plasma microRNA profiling reveals loss of endothelial miR-126 and other microRNAs in type 2 diabetes. *Circ Res* 2010; 107: 810-7.
42. Gupte P, Amarapurkar D, Agal S, Baijal R, Kulshrestha P, Pramanik S, Patel N, et al. Non-alcoholic steatohepatitis in type 2 diabetes mellitus. *J Gastroenterol Hepatol* 2004; 19: 854-8.
43. Zhang J, Du YY, Lin YF, Chen YT, Yang L, Wang HJ, Ma D. The cell growth suppressor, mir-126, targets IRS-1. *Biochem Biophys Res Commun* 2008; 377: 136-40.
44. Dongiovanni P, Valenti L, Rametta R, Daly A, Nobili V, Mozzi E, Leathart J, et al. Genetic variants regulating insulin recep-

tor signalling are associated with the severity of liver damage in patients with non-alcoholic fatty liver disease. *Gut* 2010; 59: 267-73.

**Correspondence and reprint request:**

Liwei Zhuang, M.D.

Gastroenterology and Hepatology Department, the Forth  
Affiliated Hospital of Harbin Medical University, Harbin,  
Heilongjiang province, China, 150001.

Tel.: +86-0451-82576955. Fax: +86-0451-82576955

E-mail: zhuanglw123@163.com