

SCIENTIFIC REPORTS



OPEN

Characterization of transcriptional modules related to fibrosing-NAFLD progression

Yi Lou^{1,2}, Guo-Yan Tian¹, Yu Song¹, Yin-Lan Liu¹, Yi-Dan Chen¹, Jun-Ping Shi¹ & Jin Yang¹

Based on the severity of liver fibrosis, low or high-risk profile of developing end-stage liver disease was present in nonalcoholic fatty liver disease (NAFLD). However, the mechanisms inducing transition from mild to advanced NAFLD are still elusive. We performed a system-level study on fibrosing-NAFLD by weighted gene co-expression network analysis (WGCNA) to identify significant modules in the network, and followed by functional and pathway enrichment analyses. Moreover, hub genes in the module were analyzed by network feature selection. As a result, fourteen distinct gene modules were identified, and seven modules showed significant associations with the status of NAFLD. Module preservation analysis confirmed that these modules can also be found in diverse independent datasets. After network feature analysis, the magenta module demonstrated a remarkably correlation with NAFLD fibrosis. The top hub genes with high connectivity or gene significance in the module were ultimately determined, including LUM, THBS2, FBN1 and EFEMP1. These genes were further verified in clinical samples. Finally, the potential regulators of magenta module were characterized. These findings highlighted a module and affiliated genes as playing important roles in the regulation of fibrosis in NAFLD, which may point to potential targets for therapeutic interventions.

Nonalcoholic fatty liver disease (NAFLD) represents a wide spectrum of disorders ranging from simple steatosis (nonalcoholic fatty liver, NAFL), nonalcoholic steatohepatitis (NASH), to cirrhosis or hepatocellular carcinoma (HCC). To date, NAFLD is one of the most common types of liver disease in the world¹.

The diagnosis of NAFLD rests on clinicopathological criteria, requiring both clinical and biopsy-based information. The histological findings are graded as fatty change and necroinflammatory using NAFLD activity score (NAS) scoring system, while fibrosis staging is useful to assess the severity and underlying cause of liver disease². Challenges still lies in a lacking consensus for the classification of fatty liver disease, and absence of a uniform histological definition of NAFLD³.

Correspondingly, the phenotype and outcome of NAFLD is quite heterogeneous. For instance, though most NAFLD patients do not develop clinically significant hepatic disease, some patients can progress to cirrhosis, leading ultimately to HCC⁴. In addition, not all individuals with NASH finally develop cirrhosis or liver cancer⁵. On the contrary, although the severity of NASH generally correlates with the stage of fibrosis, some individuals with advanced fibrosis have relatively little NASH⁶. Moreover, the stage of fibrosis on liver biopsy independently associates with liver-related mortality⁷. When advanced fibrosis is present, absence of NASH is no longer prognostic. For these reasons, recent clinical data focus increasing attention on determining fibrosis, as it is a strong indicator of the risk extent for NAFLD⁸. Recently, our group have shown that fibrosis score is a useful predictor of long-term outcome in NAFLD patients⁹. Mechanisms leading to strongly differing progression of NAFLD, in term of fibrosis, have to be elucidated.

The high-throughput technology such as gene expression profiling has been applied to NAFLD and provides insights into molecular aspects of NAFLD progression⁵. However, gene lists based on differential expression analysis methods are biased against genes with large changes in expression, while lacking the consideration of the relationship between changing genes as a whole. In this sense, biological networks represent valuable tools for understanding system-level properties¹⁰. One network approach, named weighted gene co-expression network analysis (WGCNA) allows for the grouping functionally related genes into modules¹¹. It is believed that modules are stable units since the overall function of a module can remain the same while individual gene expression can

¹Center for Translational Medicine, The Affiliated Hospital of Hangzhou Normal University, Hangzhou, Zhejiang, China. ²Department of occupational medicine, Hangzhou Red Cross Hospital, Hangzhou, Zhejiang, China. Correspondence and requests for materials should be addressed to J.Y. (email: hz_zhiy@163.com)

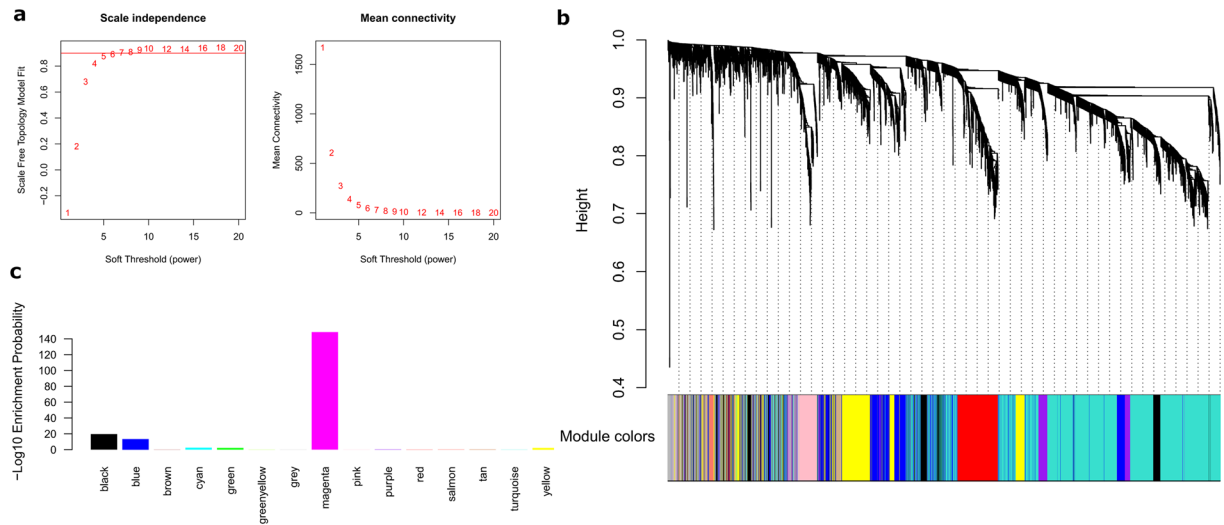


Figure 1. WGCNA network and module detection. (a) Selection of the soft-thresholding powers. The left panel shows the scale-free fit index versus soft-thresholding power. The right panel displays the mean connectivity versus soft-thresholding power. Power 5 was chosen, for which the fit index curve flattens out upon reaching a high value (>0.9). (b) Cluster dendrogram and module assignment for modules from WGCNA. Genes were clustered based on a dissimilarity measure (1-TOM). The branches correspond to modules of highly interconnected groups of genes. Colours in the horizontal bar represent the modules. 7012 transcripts were assigned to one of 15 modules including module grey. (c) Enrichment of DEGs in each module.

be changed or replaced by other genes with similar redundant functions¹². Therefore, functional modules can more effectively reveal consistent differences during NAFLD progression.

In the present study, we applied a WGCNA approach to quantitatively assess the traits of mild or advanced fibrosing-NAFLD. Genome-scale modules of co-expressed genes with clear functional annotations were identified.

Results

Weighted co-expression network construction. We used GSE49541 dataset in this study, since it contained the relatively largest NAFLD samples with clear fibrosis staging. After preprocessing the data, we applied the WGCNA package to compile the network. One outlier sample (GSM789152) was eliminated using hierarchical average linkage clustering (Supplementary File 1, Fig. S1). Keeping to the scale-free topology criterion, $\beta = 5$ was considered in this study (Fig. 1).

Following dynamic tree cut, the hierarchical clustering dendrogram identified 15 distinct gene modules, as shown in Fig. 1b. 348 genes failed to fit within a distinct group and were assigned to the grey module. The grey module was ignored in this study. The size of modules ranged from 38 (cyan module) to 2514 (turquoise module) genes. All attributes of genes and samples were shown in Supplementary File 2, Table S1–S2.

When compared with the mild samples, a total of 1134 differentially expressed genes (DEGs) were screened from the advanced NAFLD samples, including 762 upregulated and 371 downregulated genes. Consistent with the earlier research¹³, WGCNA modeling using DEGs could not match the scale-free feature of the network (Supplementary File 1, Fig. S2). DEG enrichment in each module was shown in Fig. 1c, in which DEG was mostly enriched in magenta module, and followed by black and blue module.

Identification of meta-modules associated with NAFLD severity. Next, we evaluated the relationship between each module and NAFLD status by correlating the eigengenes for each module with the fibrosis trait. Seven modules showed association evidence with $p < 0.05$ (Fig. 2). Among them, five modules (tan, green, yellow, cyan, magenta) were positively correlated with fibrosis, thereafter named fibrosing-NAFLD modules. Two negatively correlated modules (blue, black) named NAFLD modules thereafter.

Module-module relationship, also called meta-module, is the groups of correlated eigengenes with correlation of eigengenes > 0.5 . As shown in Fig. 2c,d, the dendrogram indicates that module magenta is highly correlated with NAFLD fibrosis. While black and blue modules are highly related, their mutual correlations are stronger than their correlations with fibrosis trait.

Stability and preservation of co-expression modules. To test the stability of the identified modules, internal analysis by repeating network construction and module identification on expression data that consists of resampled sets of the original dataset was performed¹⁴. The result proved the robustness of module assignments (Supplementary File 1, Fig. S3).

To ask if the identified modules were common in different datasets, an independent validation was performed. We retrieved 8 datasets relevant to NAFLD. All samples were from human liver tissue. Fibrosing-NAFLD modules (tan, green, yellow, cyan, magenta) were stable across E-MEXP-3291, GSE48452, and GSE59045. To examine

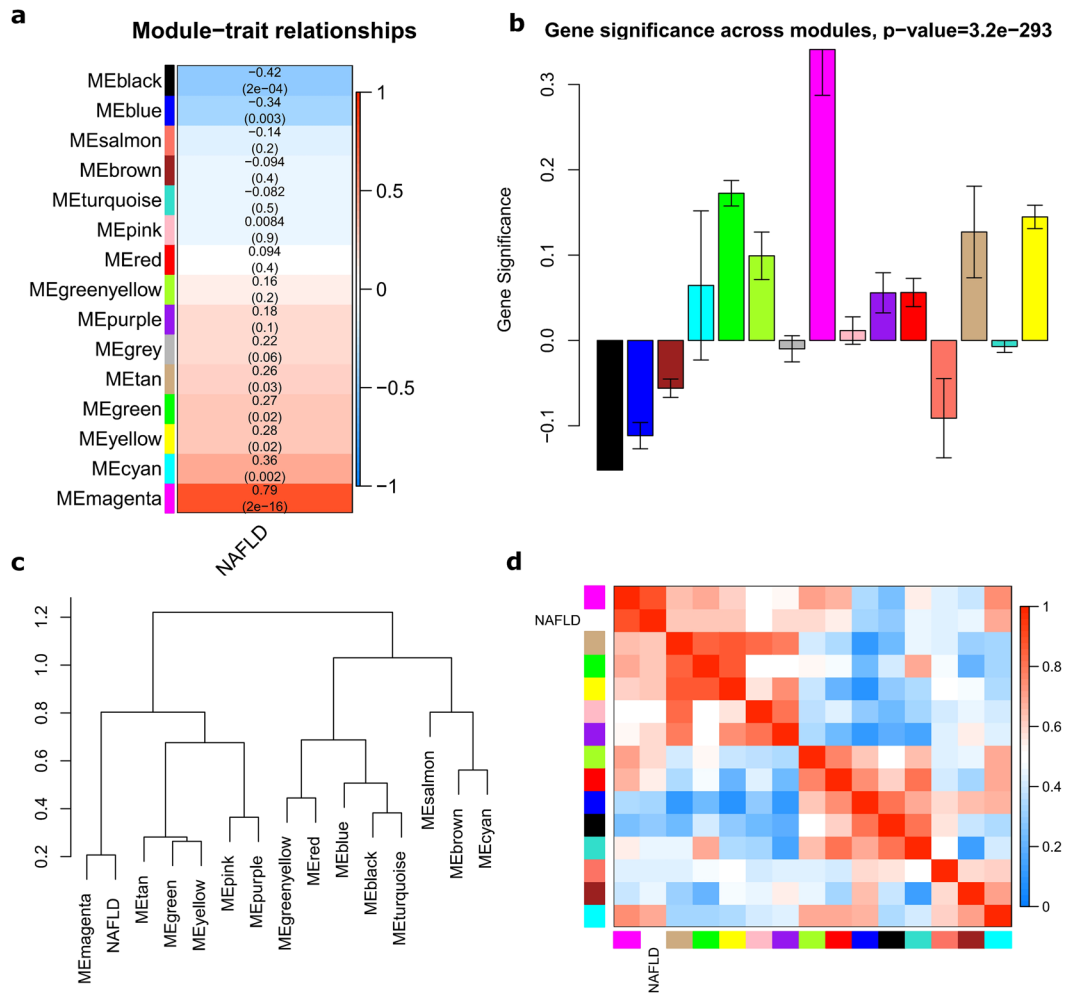


Figure 2. Module-trait and module-module associations of the network. **(a)** Each row corresponds to a module eigengene, column to a trait. Each cell contained the corresponding correlation and p value. The table was color-coded by correlation according to the color legend. The grey module included all the genes that can't be clustered. **(b)** Module significance of each module, which is determined as the average absolute gene significance measure for all genes in a given module. **(c,d)** Eigengene network, including the clustering tree and heatmap, represents the relationships among the modules and the NAFLD trait. Meta-modules are defined as tight clusters of modules. The dendrogram indicates that magenta module and fibrosing-NAFLD trait are highly related. Conversely, blue and black modules are highly related, this meta-module is inversely correlated with fibrosis.

if these identified modules were also presented in fibrosing liver disease such as HBV, GSE84044 dataset containing 124 samples with different stages of fibrosis was used as a specific control (Fig. 3).

In addition, NAFLD modules (blue, black) were preserved in other 5 datasets (GSE17470, GSE24807, GSE37031, GSE46300, and GSE63067) according to the summary preservation statistics, while fibrosing-NAFLD modules showed weak to none evidence for module preservation (Supplementary File 1, Fig. S4).

Functional enrichment analysis of the gene modules of interest. Gene ontology annotation and enrichment analysis were accomplished using DAVID version 6.8 (<https://david-d.ncicrf.gov/>)¹⁵. Top biological processes and KEGG pathway in each module was shown in Table 1.

Globally, top 5 biological processes were enriched in the modules of interest, including small molecule metabolic process (blue, FDR = 1.10E-35), extracellular matrix organization (magenta, FDR = 2.50E-10), gluconeogenesis (blue, FDR = 2.70E-06), extracellular matrix disassembly (magenta, FDR = 7.31E-06), respiratory electron transport chain (black, FDR = 1.30E-04). Top 5 enriched pathways were as follows: hsa01100: Metabolic pathways (FDR = 7.81E-15) in the blue module, hsa01200: Carbon metabolism (FDR = 8.11E-11) in the blue module, hsa04510: Focal adhesion (FDR = 2.91E-08) in the magenta module, hsa04512: ECM-receptor interaction (FDR = 1.91E-06) in the magenta module, and hsa04932: Non-alcoholic fatty liver disease (NAFLD) (FDR = 4.19E-03) in the black module. The complete annotation for each module was provided in Supplementary File 2, Table S3–S4.

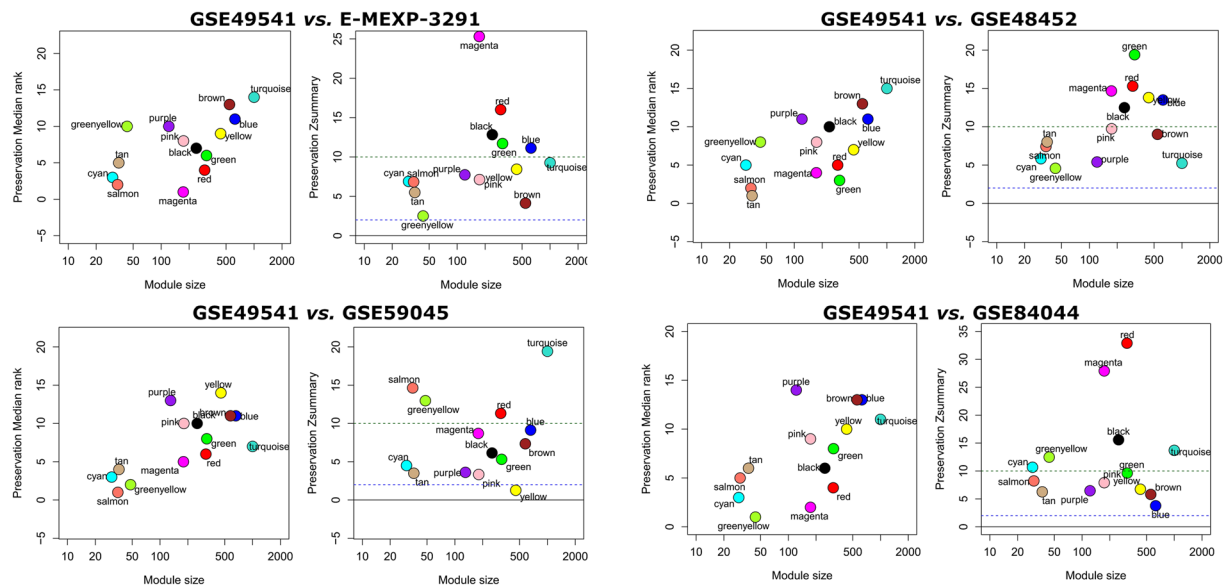


Figure 3. Preservation of GSE49541 network modules in different datasets. Each module is represented by its color-code and name. Left figure shows the composite statistic Preservation median rank. This measure tends to be independent from module size with high median ranks indicating low preservation. Right figure shows preservation Zsummary statistic. The dashed blue and green lines indicate the thresholds $Z = 2$ and $Z = 10$, respectively. $Z_{summary} < 2$ implies no evidence for module preservation, $2 < Z_{summary} < 10$ implies weak to moderate evidence, and $Z_{summary} > 10$ implies strong evidence for module preservation. Fibrosing-NAFLD modules (tan, green, yellow, cyan, magenta) show high preservation statistics summary than expected by random chance using bootstrapping validation procedures.

	module	Category	Term ^a	PValue	FDR
Fibrosing-NAFLD module	magenta	GOTERM_BP	GO:0030198~extracellular matrix organization	1.51E-13	2.50E-10
	cyan	GOTERM_BP	GO:0097284~hepatocyte apoptotic process	2.45E-02	2.73E+01
	yellow	GOTERM_BP	GO:0043687~post-translational protein modification	1.29E-06	2.21E-03
	green	GOTERM_BP	GO:0032486~Rap protein signal transduction	6.16E-04	1.01E+00
	tan	GOTERM_BP	GO:0050790~regulation of catalytic activity	7.24E-03	9.18E+00
NAFLD module	black	GOTERM_BP	GO:0022904~respiratory electron transport chain	6.68E-08	1.09E-04
	blue	GOTERM_BP	GO:0044281~small molecule metabolic process	6.21E-39	1.10E-35
	module	Category	Term	PValue	FDR
Fibrosing-NAFLD module	magenta	KEGG_PATHWAY	hsa04510:Focal adhesion	2.37E-11	2.91E-08
	cyan	KEGG_PATHWAY	hsa03030:DNA replication	8.05E-02	5.08E+01
	yellow	KEGG_PATHWAY	hsa04120:Ubiquitin mediated proteolysis	8.90E-04	1.13E+00
	green	KEGG_PATHWAY	hsa04720:Long-term potentiation	1.73E-02	1.94E+01
	tan	KEGG_PATHWAY	hsa04015:Rap1 signaling pathway	5.78E-02	4.58E+01
NAFLD module	black	KEGG_PATHWAY	hsa04932:Non-alcoholic fatty liver disease (NAFLD)	3.33E-06	4.19E-03
	blue	KEGG_PATHWAY	hsa01100:Metabolic pathways	5.96E-18	7.81E-15

Table 1. Top GO and pathway enrichment in each module.

Network analysis of the gene modules of interest. After viewing the global properties of the interesting networks, we next examined the gene constitution of particular modules based on network unique properties, such as gene significance (GS), module membership (MM) and intramodular connectivity (K.in).

Abstractly speaking, a gene is more meaningful with high GS, MM and K.in¹⁶. Thus, if MM, K.in or GS of the specific module were significantly connected and associated with the NAFLD fibrosis status, it implied that the module serves an more important biological role to NAFLD progression.

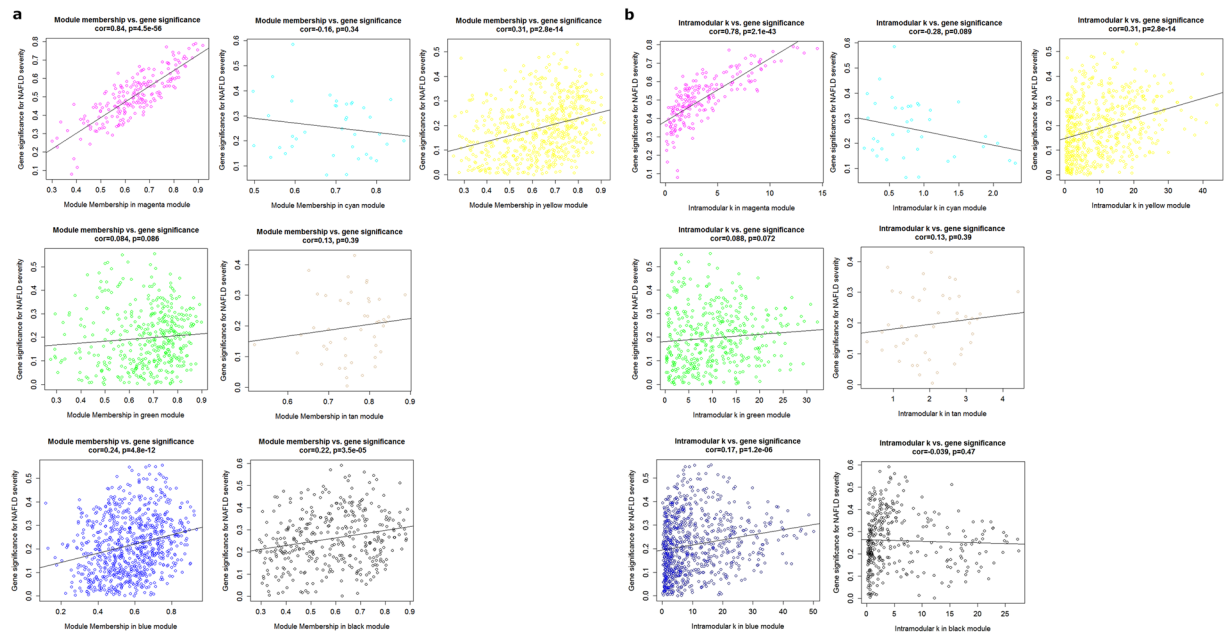


Figure 4. Module features of GS, MM and K.in. **(a)** Modules significantly correlated with NAFLD status (mild versus advanced fibrosis). Each point represents an individual gene within each module, which are plotted by GS on the y-axis and MM on the x-axis. The regression line, correlation value and p-value are shown for each plot. **(b)** Correlation of the K.in (x-axis) and the GS (y-axis).

Of the seven interesting modules, significant correlations were observed between MM and GS in the yellow, blue, black, and magenta modules. We also found a markedly correlation between GS and K.in in the yellow, blue, black, and magenta modules (Fig. 4). Overall, module magenta shows the best as reflected by its strongly positive correlations ($r = 0.84$, $p = 4.5E-56$ in GS vs. MM; $r = 0.78$, $p = 2.1E-43$ in GS vs. K.in). These results indicated that magenta module is heavily involved in NAFLD fibrosis progression.

Characterization of the magenta module content and hub genes. A network view of magenta module, modeled by cytoscape with $TOM \geq 0.1$, is depicted in Fig. 5. The K.in count for each gene ranged from 0.1 to 14.5, with an average of 3.49 ± 2.90 . The GS score for each gene ranged from -0.58 to 0.79 , with an average of 0.34 ± 0.03 . The MM for each gene ranged from -0.68 to 0.92 , with an average of 0.43 ± 0.03 . Using STRING¹⁷ or GeneMANIA¹⁸ database to model the network gave the similar results (Supplementary File 1 Fig. S6).

Focusing on the magenta module, we explored those core genes that had a high significance for NAFLD status, as well as high K.in. Network top interesting genes of the magenta module based on the above two indexes are listed in Table 2. The three top network hub genes (LUM, FBN1, and THBS2) based on K.in and the three top genes (EFEMP1, THBS2, and LUM) ranked on GS were disclosed.

Specifically, several proteomic studies have identified lumican (LUM) is expressed differentially across the progressive stages of NAFLD¹⁹. Upregulated expression of LUM is in association to hepatic fibrosis²⁰. Moreover, in animal studies, LUM is a prerequisite for hepatic fibrosis, which involves collagen fibrillogenesis, and matrix turnover²¹.

To the best of our knowledge, there was nothing directly implicating EFEMP1, FBN1 and THBS2 reported to be associated with severe NAFLD. However, EFEMP1 has been showed decreased expression in HCC tissue²². The following two genes, FBN1 and THBS2, were belong to the cellular adhesion and extracellular matrix constituent. FBN1, has been shown mounting a hepatic progenitor cell response for tissue repair in rat liver²³. THBS2 was found over-represented in patients with vascular liver lesions such as sinusoidal dilatations²⁴.

All these four genes were significantly upregulated in advanced fibrosing-NAFLD (GSE49541). High expression of these genes was also confirmed in advanced NAFLD in other cohorts (E-MEXP-2191, GSE48452 and GSE84044, Fig. 5c). Conversely, these genes were not differentially expressed in the NAFLD dominant datasets (Supplementary File 1, Fig. S5), suggesting EFEMP1, FBN1 and THBS2 maybe the novel candidate biomarkers for fibrosing-NAFLD.

Hub genes were significantly up-regulated in the livers from NAFLD patients and mice. To investigate if hub genes were modified in fibrosing-NAFLD, the production of these genes were further examined in livers from NAFLD animals and patients.

Liver biopsy tissues from NAFLD patients were stained to assess disease severity. Compared with NAFLD fibrosis stage 0–1 patients ($n = 4$), LUM, EFEMP1, FBN1, and THBS2 were remarkably up-regulated in the fibrosis stage 3–4 state patients ($n = 4$) (Fig. 6a,b).

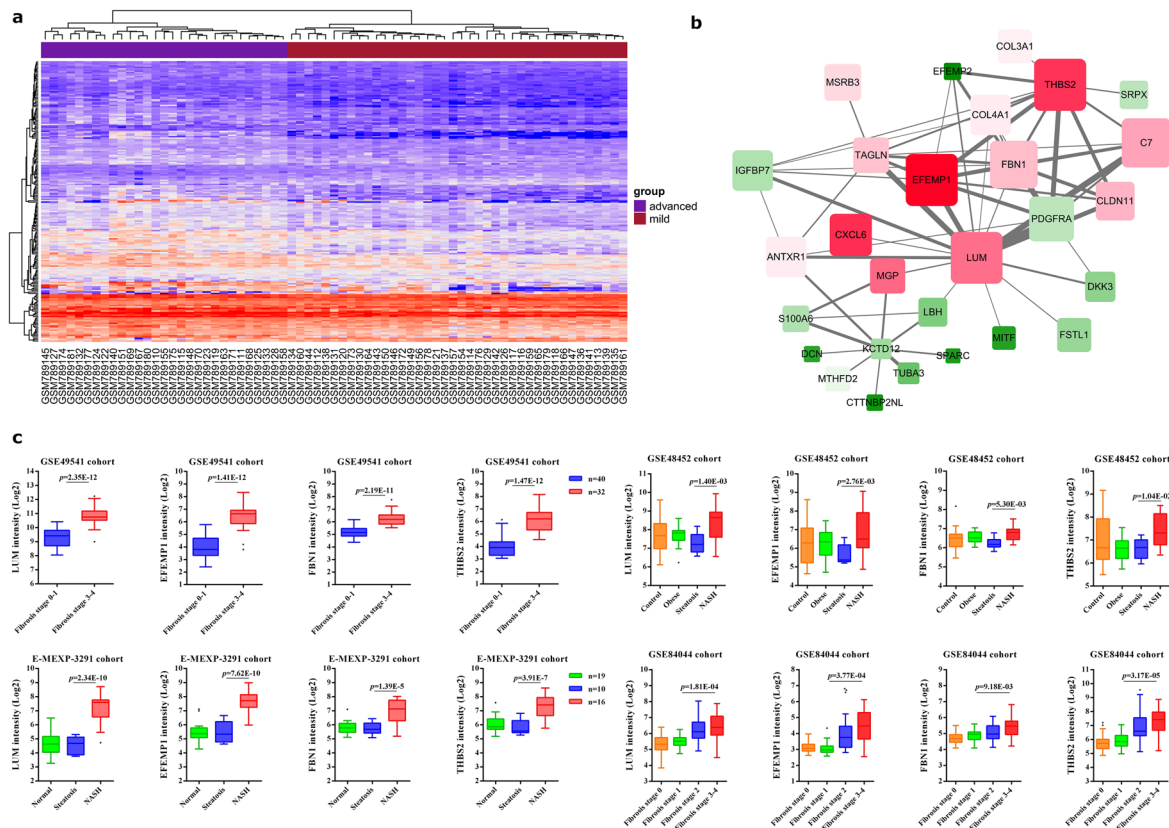


Figure 5. Characterization of the magenta module. **(a)** Gene expression heat-map of module magenta. **(b)** Interaction of gene co-expression patterns in the magenta module. The module was visualized using Cytoscape 3.0 software. The node colors coded from green to red (low to high) indicate the fold change when compared mild with advanced NAFLD state. The node size is proportional to the significance of the expression changes compared to mild NAFLD. **(c)** Four hub genes expression pattern in liver tissues according to GSE49541, E-MEXP-3291, GSE48452 and GSE84044 cohort. Data were shown as box and whisker plot. Limma package was used for statistical analysis.

	Gene	GS	GSRank	k.in	k.in Rank	Potential Transcription factor
Network Hub genes (based on k.in)	LUM	0.779	3	14.511	1	ESR1, BACH1, TRIM28, RUNX2
	FBN1	0.756	6	13.234	2	TP63
	THBS2	0.785	2	12.597	3	ESR1, TP53, BACH1, TP63, RELA
Network top genes (ranked on GS)	EFEMP1	0.790	1	12.224	4	ZNF217, BACH1, TP63
	THBS2	0.785	2	12.597	3	ESR1, TP53, BACH1, TP63, RELA
	LUM	0.779	3	14.511	1	ESR1, BACH1, TRIM28, RUNX2

Table 2. Hub genes in module magenta.

ApoE^{-/-} mice receiving a HFHC diet are the well-established animal models mimicking human NAFLD²⁵. As expected, ApoE^{-/-} mice fed a HFHC diet for 20 weeks evidenced hepatic steatosis, ballooning, hepatic inflammation and increased fibrosis. Similarly, Lum, Efemp1, Fbn1, and Thbs2 were significantly higher in livers from NAFLD mice with severe fibrosis compared to livers from mice with mild fibrosis (n = 3 in each group, Fig. 6c,d). The data *in vivo* above suggests a close relationship between hub genes and NAFLD fibrosis.

Functional organization of the magenta module. Next, function relevance of magenta module was reannotated using DAVID tool. With the cutoff set as FDR < 0.1, focal adhesion, ECM-receptor interaction, and phosphatidylinositol 3-kinase (PI3K)-Akt signaling pathway constitute the main pathways in magenta module. In complete accord with the phenotype, GO term extracellular matrix organization and cell adhesion were significantly enriched.

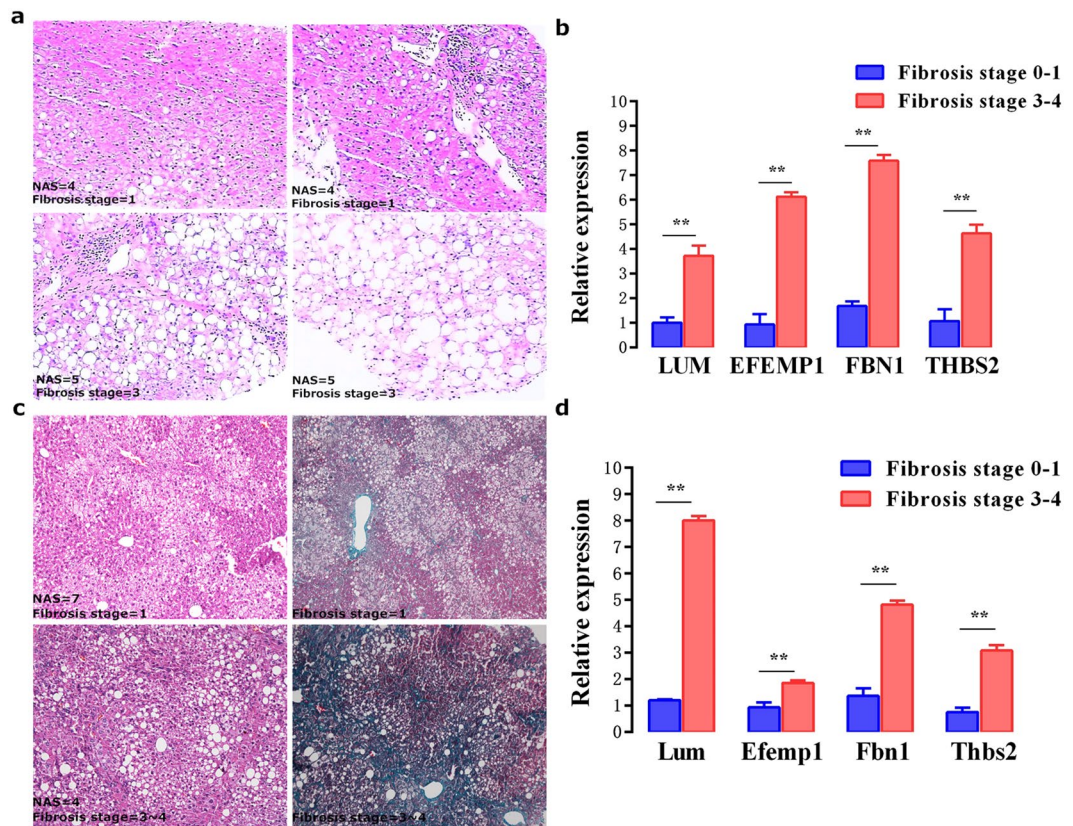


Figure 6. Expression of hub genes in different fibrosis stages of NAFLD. (a,b) The representative HE staining of NAFLD patients with different fibrosis stages were shown. Quantification of hub genes was presented. (c,d) Liver sections were stained with HE in mice fed with HFHC diet at 20 weeks. Masson's trichrome staining was used to detect the accumulated collagen. The hepatic production of hub genes was confirmed and presented. ** $P < 0.01$.

Since co-expressed genes may be co-regulated by the common transcription factors (TFs) and microRNAs, we performed gene-set enrichment analysis using ChEA, Encode, and Targetscan database^{26–28} for magenta module. Significant enrichments of transcription factors were observed for ESR1, SOX2, TP53, etc (Fig. 7; Supplementary File 2, Table S5–S7). Several TFs were reported to be functionally associated with NAFLD. For instance, CpG island methylation of ESR1 was found to be involved in lipid and glucose metabolism, and the progression of fibrosis in mouse feeding with methyl-deficient diets²⁹. SOX2 expression may predict the prognosis of HCC patients³⁰. As a metabolic modulator, growing evidences highlight TP53 a new player in NAFLD pathogenesis³¹. BACH1 gene ablation reduces steatohepatitis in mouse³². Specifically, RELA (p65), is well-known for its triggering inflammatory responses in NAFLD³³. RUNX2 downregulation is involved in cirrhotic liver³⁴. Most recently, TRIM28 was found to modulate the prevalence of obesity in the population³⁵. The other two genes (ZNF217, TP63) enriched in magenta module have not been reported for their relationship to liver.

H3 lysine 27 trimethylation (H3K27me3) was strongly enriched for most of the genes in module magenta. Recent paper has shown that in human NAFLD-associated HCC, level of H3K27me3 was correlated positively in tumors compared with nontumor tissues³⁶.

Finally, the most enriched miRNAs were observed for miR-200b, miR-200c and miR-429. In high fat diet feeding rats, miR-200a, miR-200b and miR-429 were significantly correlated with a severity of NAFLD-specific liver pathomorphological features³⁷. Another study has showed, miR-200a, miR-200b, and miR-200c were up-regulated in NAFLD³⁸.

Discussion

Here, we present a systematic WGCNA of NAFLD with either mild fibrosis or advanced fibrosis. Among the 14 modules discovered in this study, seven modules were significantly associated with disease progression.

As an alternative to traditional differential expression analyses which centered on individual genes distinguishing the status of NAFLD, in an unbiased manner, WGCNA groups co-expressed genes that are biologically integrated on a genome-wide scale. We found several features in term of the relationship between WGCNA and DEG analysis. First, using differential expression genes to model WGCNA network is not suggested, since it completely invalidates the scale-free topology assumption. Second, we found that, globally, DEG genes have less connectivity than non-DEG genes (9.57 ± 12.91 vs. 24.93 ± 37.59 , respectively), and this association does not follow a simple monotonic trend in each module. Third, DEG genes tend not to be hub genes as determined by connectivity in

a	ChEA_Transcription factors	P-value	Adjusted P-value	Z-score	Combined Score
	ESR1_22446102_ChIP-Seq_UTERI_Mouse	5.62E-22	7.07E-20	-2.54	111.80
	SOX2_20726797_ChIP-Seq_SW620_Human	0.00E+00	3.00E-30	-1.64	111.42
	TP53_18474530_ChIP-ChIP_U20S_Human	3.65E-17	1.15E-15	-2.04	70.15
	ZNF217_24962896_ChIP-Seq_MCF7_Human	1.42E-19	8.91E-18	-1.69	66.44
	BACH1_22875853_ChIP-PCR_HELA AND SCP4_Human	4.85E-16	1.31E-14	-1.79	57.16
	TRIM28_21343339_ChIP-Seq_HEK293_Human	3.66E-10	2.71E-09	-2.60	51.38
	RUNX2_24764292_ChIP-Seq_MC3T3_Mouse	8.56E-19	4.61E-17	-1.35	50.88
	TP53_16413492_ChIP-PET_HCT116_Human	1.03E-08	5.62E-08	-2.85	47.52
	TP63_23658742_ChIP-Seq_EP156T_Human	6.89E-23	1.30E-20	-1.00	45.94
	RELA_24523406_ChIP-Seq_FIBROSARCOMA_Human	4.43E-15	9.27E-14	-1.49	44.79
b	ENCODE_Histone_Modifications	P-value	Adjusted P-value	Z-score	Combined Score
	H3K27me3_cardiac_mesoderm_hg19	9.10E-12	2.99E-09	-1.62	31.79
	H3K27me3_GM12878_hg19	1.46E-11	2.99E-09	-1.56	30.60
	H3K27me3_brain_mm9	5.76E-11	7.86E-09	-1.55	28.92
	H3K27me3_testis_mm9	9.08E-09	7.42E-07	-1.73	24.46
	H3K27me3_H1-hESC_hg19	9.08E-09	7.42E-07	-1.65	23.28
	H3K27me3_mononuclear_cell_hg19	2.97E-08	1.74E-06	-1.73	22.93
	H3K27me3_MCF-7_hg19	4.47E-08	2.29E-06	-1.76	22.89
	H3K27me3_erythroblast_mm9	1.32E-08	9.01E-07	-1.59	22.15
	H3K4me1_CH12_LX_mm9	9.52E-08	4.33E-06	-1.63	20.09
	H3K27me3_CD14-positive_monocyte_hg19	2.57E-07	1.05E-05	-1.58	18.15
c	TargetScan_microRNA	P-value	Adjusted P-value	Z-score	Combined Score
	CAGTATT, MIR-200B, MIR-200C, MIR-429	2.29E-02	7.14E-01	-1.87	0.63
	CTGTAT, MIR-381	1.27E-02	7.14E-01	-1.85	0.62
	AACATTC, MIR-409-3P	2.20E-02	7.14E-01	-1.79	0.60
	ATACTGT, MIR-144	1.05E-02	7.14E-01	-1.75	0.59
	ACCAAG, MIR-9	3.57E-02	7.14E-01	-1.71	0.58
	TGTGTGA, MIR-377	3.12E-02	7.14E-01	-1.71	0.58
	TGCCTTA, MIR-124A	3.68E-02	7.14E-01	-1.67	0.56
	TGGTGCT, MIR-29A, MIR-29B, MIR-29C	4.62E-02	7.14E-01	-1.66	0.56
	TACTTGA, MIR-26A, MIR-26B	3.66E-02	7.14E-01	-1.61	0.54
	ATACCTC, MIR-202	5.29E-02	7.14E-01	-1.48	0.50

Figure 7. Potential factors regulating genes in magenta module. (a) Transcription factors. (b) Histone modification markers. (c) Enriched seed and its associated microRNA.

most modules. However, DEG numbers show a significant enrichment in seven important modules (Fig. 1c). Thus, WGCNA and DEG showed efficient mutual complementation when for transcriptome analysis.

Results suggest that the modules identified here are biologically rational. First, fibrosing-NAFLD and NAFLD modules were clearly separated by WGCNA approach. Module preservation was extensively studied among different datasets. Second, most of the identified modules are enriched for specific GO terms and KEGG pathways. For instance, module black, together with module blue, are markedly inverse correlated with NAFLD fibrosis. Both the black and blue modules were enriched in respiratory electron transport chain and small molecule metabolic process, which have already been implicated in oxidative stress, and mitochondrial dysfunction for NAFLD³⁹. KEGG analysis showed that these modules are directly related to NAFLD. Moreover, hub genes in these modules are implicated in NAFLD as reported by literature annotation. As the hub gene of blue module, selenium-binding protein 1 (SELENBP1) has been shown to be downregulated in the liver tissue of HCC patients and the association of its gradual loss with an increased malignant grade⁴⁰. Thus, it is tempting to speculate that the NAFLD modules (black, and blue) dominate the period of NAFLD when fibronesis is not obvious.

Recently, the systems biology approach for NAFLD has been applied in studies by integrating genomic data and metabolic networks. One study showed that metabolism of amino acids, chondroitin and heparan sulphates seem to be involved in the appearance of NASH⁴¹. In our analysis, we observed regulation of cellular amino acid metabolic process in black or blue module, including branched-chain, sulfur, alpha-amino, serine family, aspartate family amino acid metabolic process. Simultaneously, cellular response to amino acid stimulus was also enriched in magenta module with relatively small probability, which suggests the continuous amino acid metabolic stress during NAFLD pathogenesis.

This study highlights the importance of magenta module as a driver of fibrosis based on the meta-module, and further through the network feature (GS, MM and K.in) analysis. Enriched GO terms or pathway are highly concordant. In particular, extracellular matrix (ECM) and PI3K pathway were top core gene sets of the magenta module. The ECM is mainly composed of an intricate interlocking mesh of fibrillar and non-fibrillar collagens, elastic fibers and glycoproteins⁴², which is a highly dynamic structure undergoing controlled remodelling⁴³. The deposition of increased and abnormal ECM is the hallmark of liver fibrosis⁴⁴. Correspondingly, the PI3K signaling pathway has been shown to regulate procedures associated with hepatic stellate cell (HSC) activation such as collagen synthesis and cell proliferation⁴⁵. Inhibition of PI3K signaling in HSCs suppresses ECM deposition,

type I collagen synthesis, and reduce the expression of profibrogenic factors. In a feed-back manner, collagen crosslinking increases stiffness, $\beta 1$ integrin clustering, PI3K signalling and focal adhesion formation to drive the disease progression⁴³.

The progression of NAFLD from mild steatosis up to severe steatohepatitis and even cirrhosis or HCC, varies widely between individual patients. Recently, one study was to assess the histological severity of NAFLD in a cohort with serial biopsy data, and contrary to current dogma, this study suggests that steatosis can directly progress to NASH and clinically significant fibrosis⁴⁶. Since causality between liver fibrosis and the prognosis of the liver disease is well recognized, the intervening measure may have to be adjusted in different subsets of NAFLD patients.

To demonstrate the usefulness of our modules in the development of efficient NAFLD interfere strategies, given magenta module as an example, small compounds derived from the Library of Integrated Network-based Cellular Signatures (LINCS) L1000 platform⁴⁷ affecting the gene expression was provided in Supplementary File 2 Table S8–9. Next, novel potential biomarkers including EFEMP1, THBS1, FBN1 were disclosed in magenta module, after extensive cross-validation. Another noteworthy was the link between the regulators (transcription factors and epigenetic markers) and the co-expression mode of genes in magenta module, which may suggest the regulatory circuit during NAFLD progression. In the future, more experiments are needed to validate these discovery clues.

In summary, this study generated a comprehensive and unbiased snap-shot of the modules as well as genes in fibrosis-NAFLD. In particular, magenta module and genes regulating ECM remodelling during NAFLD progression deserve further attention. An identification of mechanistically linked key module and regulators will aid intervention development.

Methods

Gene expression dataset and processing. Transcription profile of NAFLD was downloaded from the Gene Expression Omnibus (GEO) with accession number GSE49541. The raw data were corrected and normalized using the RMA function of affy package of R 3.2.0 in Bioconductor. This dataset represents two clinically defined pathological groups at the extremes of NAFLD: mild NAFLD ($n = 40$, fibrosis stages 0–1), with little risk of developing severe liver disease; advanced NAFLD ($n = 32$, fibrosis stage 3–4), with significant likelihood of developing liver-related morbidity and mortality. As described earlier, the two groups were matched for gender, age and body mass index⁴⁸. Differential expressed genes (DEG) were considered by using the criterion with Benjamini & Hochberg adjusted p value less than 0.05.

The microarray datasets referenced during the study (E-MEXP-3291, GSE45428, GSE50594 and GSE84044, etc) are available in a public repository from EBI (<http://www.ebi.ac.uk/>) or NCBI GEO. All the other data supporting the findings of this study are available within the article and its Supplementary File 1 Table S1. It is important to emphasize we include only the human NAFLD samples in our study. Full experimental methods and detailed descriptions of these public data sets can be found in the original references.

WGCNA network construction and module detection. After normalization, we removed lowly and nonexpressed genes by selecting probes with a mean expression in the top 50% of all probes. Next, genes with expression variance above average level were selected. Different probes targeting the same gene were collapsed. These steps finally resulted in 7012 genes to infer co-expression networks.

Networks were formed following the protocols of WGCNA⁴⁹. A pairwise Pearson correlation coefficient matrix was first computed, and an adjacency matrix and topological overlap matrix (TOM) were constructed¹¹. TOM is a parameter referring to the interconnection between two genes, and a module is a cluster of genes with high topological overlap. Modules were identified on the dendrogram using the dynamic tree cut algorithm⁵⁰. The module eigengene (ME) is defined as the first principal component of a given module, which can be considered a representative of the gene expression profiles in a module. Module Membership (MM), also known as eigengene-based connectivity (kME), is defined as the correlation between the module eigengene and gene expression values. Genes weakly correlated with all of the MEs ($|kME| < 0.7$) were assigned to none of the modules. Finally, the interesting module network was visualized by Cytoscape⁵¹.

Module preservation implemented in WGCNA was used to detect the conservation of gene pairs between two networks⁵². Briefly, three types of network based module preservation statistics, including density based preservation statistics, connectivity based preservation statistics, and network based statistics have been identified. Then, two composite measures have been defined. Median rank is defined as the mean of median ranks computed for connectivity and density measures of each module. Zsummary is used to assess the significance of observed statistics and is defined as the mean of Z scores computed for density and connectivity measures. We utilize median rank to identify module preservation and Zsummary to assess significance of module preservation via permutation testing 200 times.

Feature vectors in WGCNA network. Gene significance (GS) was defined as the correlation between individual genes and NAFLD trait. The intramodular connectivity (K.in) was calculated as the summation of adjacency performed over all genes in a particular network. If GS and MM are highly correlated, it means that genes are the most important elements of modules and are highly significantly associated with the trait. Generally, the MM is high in relation to k.in, and a higher correlation indicates that a gene is more important to the given module¹¹.

Hub genes tends to be located in the centre of a network, highly connected with other genes and hence of high functional significance. Therefore, a gene with high GS, high MM and high K.in in a module was considered to be a hub gene.

Functional annotation of the modules. Gene ontology (GO) and KEGG pathway enrichment analysis for network modules were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID)¹⁵. In DAVID, an overrepresentation of a term is defined as a modified Fisher's exact P value with an adjustment for multiple tests using Benjamini method.

In addition, we related interesting modules to biological curated gene sets on the basis of Enrichr (<http://amp.pharm.mssm.edu/Enrichr>), which currently contains a large collection of diverse gene set libraries available for analysis⁵³. In Enrichr, apart from the Fisher exact test used in most gene list enrichment analyses, a z-score test statistics capture the deviation from the expected rank by the Fisher exact test. Then, a combined score evaluating the enrichment was computed as follows: $C = \log(p) \cdot Z$. Where C indicates the combined score, p is the p-value computed using the Fisher exact test, and Z is the z-score computed by assessing the deviation from the expected rank.

Ethical considerations. All the experiments protocols involving humans and animals were approved by the Human Ethics Committee of the affiliated hospital of Hangzhou Normal University. Methods were carried out in accordance with the approved guidelines and regulation. Written informed consent was obtained from all participants. Appropriate care was given to all animals included for experiments.

Patients. The criteria for NAFLD were based on those recommended by the Chinese Liver Disease Association. Liver biopsies from NAFLD patients with different stages of fibrosis, identified between Feb 2013 and Oct 2016 in the Department of Liver Diseases, affiliated hospital of Hangzhou Normal University, China was collected and stored at -80°C before analysis. Paraffin embedded liver tissues were used for immunohistochemistry (IHC).

Animal studies. Four-week-old male ApoE^{-/-} mice were purchased from Model Animal Research Center of Nanjing University (Nanjing, China). All mice were bred in a specific pathogen-free facility and maintained in a 12-hour light-dark-cycle at room temperature and fed ad libitum. The mice were divided into two groups and allocated into either a normal Chow-diet (Normal group) or high fat high cholesterol-diet (HFHC group). HFHC diet was from Research Diets, New Brunswick (D12079B; Research Diets New Brunswick, NJ). At the end of the experiments, a part of the liver tissue was fixed with 10% formaldehyde and the remaining liver was snap frozen.

Liver histology. Formalin-fixed liver tissue was processed into 4 μm thick paraffin sections and stained with hematoxylin and eosin (HE) and Masson's staining. Degree of NAFLD activity score (NAS) were scored according to NASH clinical research network (CRN) scoring system.

Quantitative real-time PCR. Hepatic mRNA levels were analyzed by qRT-PCR using a 7900 Real Time PCR System (Applied Biosystems, USA). The RNA was isolated with TRIzol (Invitrogen, USA). cDNA was synthesized using 2 μg of total RNA with PrimeScriptTM Reverse Transcriptase (Takara). Amplification reactions were performed using the SYBR[®] Premix Ex Taq kit (Takara) and 0.2 μM of gene specific primers (Supplementary File 1 Table S2) and PCR products were verified by melting curve analysis. The relative quantification expression was calculated using the delta-delta Ct method with each gene normalized to GAPDH.

References

- Unalp-Arida, A. & Ruhl, C. E. Noninvasive fatty liver markers predict liver disease mortality in the U.S. population. *Hepatology* **63**, 1170–1183, doi:10.1002/hep.28390 (2016).
- Hubscher, S. G. Histological assessment of non-alcoholic fatty liver disease. *Histopathology* **49**, 450–465, doi:10.1111/j.1365-2559.2006.02416.x (2006).
- Hashimoto, E., Tokushige, K. & Ludwig, J. Diagnosis and classification of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis: Current concepts and remaining challenges. *Hepatology research: the official journal of the Japan Society of Hepatology* **45**, 20–28, doi:10.1111/hepr.12333 (2015).
- Pagadala, M. R. & McCullough, A. J. The relevance of liver histology to predicting clinically meaningful outcomes in nonalcoholic steatohepatitis. *Clinics in liver disease* **16**, 487–504, doi:10.1016/j.cld.2012.05.006 (2012).
- Younossi, Z. M. *et al.* Pathologic criteria for nonalcoholic steatohepatitis: interprotocol agreement and ability to predict liver-related mortality. *Hepatology* **53**, 1874–1882, doi:10.1002/hep.24268 (2011).
- Caldwell, S. H. *et al.* NASH and cryptogenic cirrhosis: a histological analysis. *Annals of hepatology* **8**, 346–352 (2009).
- Mendes, F. D., Suzuki, A., Sanderson, S. O., Lindor, K. D. & Angulo, P. Prevalence and indicators of portal hypertension in patients with nonalcoholic fatty liver disease. *Clinical gastroenterology and hepatology: the official clinical practice journal of the American Gastroenterological Association* **10**, 1028–1033 e1022, doi:10.1016/j.cgh.2012.05.008 (2012).
- Haas, J. T., Francque, S. & Staels, B. Pathophysiology and Mechanisms of Nonalcoholic Fatty Liver Disease. *Annual review of physiology* **78**, 181–205, doi:10.1146/annurev-physiol-021115-105331 (2016).
- Xun, Y.-h. *et al.* Non-alcoholic fatty liver disease (NAFLD) fibrosis score predicts 6.6-year overall mortality of Chinese patients with NAFLD. *Clinical and Experimental Pharmacology and Physiology* **41**, 643–649, doi:10.1111/1440-1681.12260 (2014).
- Furlong, L. I. Human diseases through the lens of network biology. *Trends in genetics: TIG* **29**, 150–159, doi:10.1016/j.tig.2012.11.004 (2013).
- Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* **4**, Article17, doi:10.2202/1544-6115.1128 (2005).
- Wang, A. *et al.* Functional modules distinguish human induced pluripotent stem cells from embryonic stem cells. *Stem cells and development* **20**, 1937–1950, doi:10.1089/scd.2010.0574 (2011).
- Yang, Y. *et al.* Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nature communications* **5**, 3231, doi:10.1038/ncomms4231 (2014).
- Langfelder, P. & Horvath, S. Fast R Functions for Robust Correlations and Hierarchical Clustering. *Journal of statistical software* **46** (2012).
- Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44–57, doi:10.1038/nprot.2008.211 (2009).
- Langfelder, P., Mischel, P. S. & Horvath, S. When is hub gene selection better than standard meta-analysis? *PLoS one* **8**, e61505, doi:10.1371/journal.pone.0061505 (2013).

17. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic acids research* **43**, D447–452, doi:10.1093/nar/gku1003 (2015).
18. Montojo, J., Zuberi, K., Rodriguez, H., Bader, G. D. & Morris, Q. GeneMANIA: Fast gene network construction and function prediction for Cytoscape. *F1000Research* **3**, 153, doi:10.12688/f1000research.4572.1 (2014).
19. Nuno-Lambarri, N., Barbero-Becerra, V. J., Uribe, M. & Chavez-Tapia, N. C. Mitochondrial Molecular Pathophysiology of Nonalcoholic Fatty Liver Disease: A Proteomics Approach. *International journal of molecular sciences* **17**, doi:10.3390/ijms17030281 (2016).
20. Bracht, T. *et al.* Analysis of disease-associated protein expression using quantitative proteomics—fibulin-5 is expressed in association with hepatic fibrosis. *Journal of proteome research* **14**, 2278–2286, doi:10.1021/acs.jproteome.5b00053 (2015).
21. Krishnan, A. *et al.* Lumican, an extracellular matrix proteoglycan, is a novel requisite for hepatic fibrosis. *Laboratory investigation; a journal of technical methods and pathology* **92**, 1712–1725, doi:10.1038/labinvest.2012.121 (2012).
22. Nomoto, S. *et al.* Epidermal growth factor-containing fibulin-like extracellular matrix protein 1, EFEMP1, a novel tumor-suppressor gene detected in hepatocellular carcinoma using double combination array analysis. *Annals of surgical oncology* **17**, 923–932, doi:10.1245/s10434-009-0790-0 (2010).
23. Vestentoft, P. S. *et al.* Molecular constituents of the extracellular matrix in rat liver mounting a hepatic progenitor cell response for tissue repair. *Fibrogenesis & tissue repair* **6**, 21, doi:10.1186/1755-1536-6-21 (2013).
24. Agostini, J. *et al.* Identification of molecular pathways involved in oxaliplatin-associated sinusoidal dilatation. *Journal of hepatology* **56**, 869–876, doi:10.1016/j.jhep.2011.10.023 (2012).
25. Jeon, S., Park, Y. J. & Kwon, Y. H. Genistein alleviates the development of nonalcoholic steatohepatitis in ApoE(–/–) mice fed a high-fat diet. *Molecular nutrition & food research* **58**, 830–841, doi:10.1002/mnfr.201300112 (2014).
26. Lachmann, A. *et al.* ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438–2444, doi:10.1093/bioinformatics/btq466 (2010).
27. Mouse, E. C. *et al.* An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome biology* **13**, 418, doi:10.1186/gb-2012-13-8-418 (2012).
28. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**, doi:10.7554/eLife.05005 (2015).
29. Cordero, P., Campion, J., Milagro, F. I. & Martinez, J. A. Transcriptomic and epigenetic changes in early liver steatosis associated to obesity: effect of dietary methyl donor supplementation. *Molecular genetics and metabolism* **110**, 388–395, doi:10.1016/j.ymgme.2013.08.022 (2013).
30. Huang, P. *et al.* Role of Sox2 and Oct4 in predicting survival of hepatocellular carcinoma patients after hepatectomy. *Clinical biochemistry* **44**, 582–589, doi:10.1016/j.clinbiochem.2011.02.012 (2011).
31. Derdak, Z. *et al.* Inhibition of p53 attenuates steatosis and liver injury in a mouse model of non-alcoholic fatty liver disease. *Journal of hepatology* **58**, 785–791, doi:10.1016/j.jhep.2012.11.042 (2013).
32. Inoue, M. *et al.* Bach1 gene ablation reduces steatohepatitis in mouse MCD diet model. *Journal of clinical biochemistry and nutrition* **48**, 161–166, doi:10.3164/jcfn.09-122GFR (2011).
33. Willy, J. A., Young, S. K., Stevens, J. L., Masuoka, H. C. & Wek, R. C. CHOP links endoplasmic reticulum stress to NF- κ B activation in the pathogenesis of nonalcoholic steatohepatitis. *Molecular biology of the cell* **26**, 2190–2204, doi:10.1091/mbc.E15-01-0036 (2015).
34. Hattori, S. *et al.* FR-167653, a selective p38 MAPK inhibitor, exerts salutary effect on liver cirrhosis through downregulation of Runx2. *Laboratory investigation; a journal of technical methods and pathology* **87**, 591–601, doi:10.1038/labinvest.3700539 (2007).
35. Dalgaard, K. *et al.* Trim28 Haploinsufficiency Triggers Bi-stable Epigenetic Obesity. *Cell* **164**, 353–364, doi:10.1016/j.cell.2015.12.025 (2016).
36. Tian, Y. *et al.* Histone Deacetylase HDAC8 Promotes Insulin Resistance and beta-Catenin Activation in NAFLD-Associated Hepatocellular Carcinoma. *Cancer research* **75**, 4803–4816, doi:10.1158/0008-5472.CAN-14-3786 (2015).
37. Alisi, A. *et al.* Mirnome analysis reveals novel molecular determinants in the pathogenesis of diet-induced nonalcoholic fatty liver disease. *Laboratory investigation; a journal of technical methods and pathology* **91**, 283–293, doi:10.1038/labinvest.2010.166 (2011).
38. Feng, Y. Y. *et al.* Aberrant hepatic microRNA expression in nonalcoholic fatty liver disease. *Cellular physiology and biochemistry: international journal of experimental cellular physiology, biochemistry, and pharmacology* **34**, 1983–1997, doi:10.1159/000366394 (2014).
39. Paradies, G., Paradies, V., Ruggiero, F. M. & Petrosillo, G. Oxidative stress, cardiolipin and mitochondrial dysfunction in nonalcoholic fatty liver disease. *World journal of gastroenterology* **20**, 14205–14218, doi:10.3748/wjg.v20.i39.14205 (2014).
40. M, D. I. S. *et al.* A possible predictive marker of progression for hepatocellular carcinoma. *Oncology letters* **2**, 1247–1251, doi:10.3892/ol.2011.378 (2011).
41. Mardinoglu, A. *et al.* Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nature communications* **5**, 3083, doi:10.1038/ncomms4083 (2014).
42. Cox, T. R. & Erler, J. T. Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. *Disease models & mechanisms* **4**, 165–178, doi:10.1242/dmm.004077 (2011).
43. Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nature reviews. Molecular cell biology* **15**, 786–801, doi:10.1038/nrm3904 (2014).
44. Wells, R. G. Cellular sources of extracellular matrix in hepatic fibrosis. *Clinics in liver disease* **12**, 759–768, viii, doi:10.1016/j.cld.2008.07.008 (2008).
45. Son, M. K. *et al.* HS-173, a novel PI3K inhibitor, attenuates the activation of hepatic stellate cells in liver fibrosis. *Scientific reports* **3**, 3470, doi:10.1038/srep03470 (2013).
46. McPherson, S. *et al.* Evidence of NAFLD progression from steatosis to fibrosing-steatohepatitis using paired biopsies: implications for prognosis and clinical management. *Journal of hepatology* **62**, 1148–1155, doi:10.1016/j.jhep.2014.11.034 (2015).
47. Vempati, U. D. *et al.* Metadata Standard and Data Exchange Specifications to Describe, Model, and Integrate Complex and Diverse High-Throughput Screening Data from the Library of Integrated Network-based Cellular Signatures (LINCS). *Journal of biomolecular screening* **19**, 803–816, doi:10.1177/1087057114522514 (2014).
48. Moylan, C. A. *et al.* Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology* **59**, 471–482, doi:10.1002/hep.26661 (2014).
49. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).
50. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720, doi:10.1093/bioinformatics/btm563 (2008).
51. Demchak, B. *et al.* Cytoscape: the network visualization tool for GenomeSpace workflows. *F1000Research* **3**, 151, doi:10.12688/f1000research.4492.2 (2014).
52. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? *PLoS computational biology* **7**, e1001057, doi:10.1371/journal.pcbi.1001057 (2011).
53. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, doi:10.1093/nar/gkw377 (2016).

Acknowledgements

This work was supported in part through National Natural Science Foundation of China (81570524), Zhejiang Provincial Natural Science Foundation (LY15H070004), Hangzhou Health Science and technology plan (201763676), and Major project of Hangzhou health and family planning science and technology project (2015ZD02).

Author Contributions

J.Y. and J.P.S. conceived, designed, and supervised the overall study. Y.L., G.Y.T., and Y.S. carried out data processing and computational analysis. Y.L.L. and Y.D.C. performed samples preparation and expression analysis. All of the authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-05044-2](https://doi.org/10.1038/s41598-017-05044-2)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017