frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 27 March 2023
DOI 10.3389/fgene.2023.1070605

Identification of disease-related genes and construction of a gene co-expression database in non-alcoholic fatty liver disease

Hua Ye[1], Mengxia Sun[2], Mingli Su[1], Dahua Chen[1], Huiwei Liu[1], Yanyan Ma[1], Wenjing Luo[1], Hong Li[3] and Feng Xu[1]*

[1]Department of Gastroenterology, The Affiliated Lihuili Hospital, Ningbo University, Ningbo, Zhejiang, China, [2]Department of Clinical Medicine, Health Science Center, Ningbo University, Ningbo, Zhejiang, China, [3]Department of Hepatobiliary Surgery, The Affiliated Lihuili Hospital, Ningbo University, Ningbo, Zhejiang, China

OPEN ACCESS

EDITED BY
Mingjie Wang,
Shanghai Jiao Tong University, China

REVIEWED BY
Aaron Koenig,
The Ohio State University, United States
Yuting Dai,
Shanghai Institute of Hematology, China

*CORRESPONDENCE
Feng Xu,
✉ xufengxh19@163.com

SPECIALTY SECTION
This article was submitted to Statistical Genetics and Methodology, a section of the journal Frontiers in Genetics

RECEIVED 02 December 2022
ACCEPTED 15 March 2023
PUBLISHED 27 March 2023

CITATION
Ye H, Sun M, Su M, Chen D, Liu H, Ma Y, Luo W, Li H and Xu F (2023), Identification of disease-related genes and construction of a gene co-expression database in non-alcoholic fatty liver disease.
Front. Genet. 14:1070605.
doi: 10.3389/fgene.2023.1070605

Background: The mechanism of NAFLD progression remains incompletely understood. Current gene-centric analysis methods lack reproducibility in transcriptomic studies.

Methods: A compendium of NAFLD tissue transcriptome datasets was analyzed. Gene co-expression modules were identified in the RNA-seq dataset GSE135251. Module genes were analyzed in the R gProfiler package for functional annotation. Module stability was assessed by sampling. Module reproducibility was analyzed by the ModulePreservation function in the WGCNA package. Analysis of variance (ANOVA) and Student's t-test was used to identify differential modules. The receiver operating characteristic (ROC) curve was used to illustrate the classification performance of modules. Connectivity Map was used to mine potential drugs for NAFLD treatment.

Results: Sixteen gene co-expression modules were identified in NAFLD. These modules were associated with multiple functions such as nucleus, translation, transcription factors, vesicle, immune response, mitochondrion, collagen, and sterol biosynthesis. These modules were stable and reproducible in the other 10 datasets. Two modules were positively associated with steatosis and fibrosis and were differentially expressed between non-alcoholic steatohepatitis (NASH) and non-alcoholic fatty liver (NAFL). Three modules can efficiently separate control and NAFL. Four modules can separate NAFL and NASH. Two endoplasmic reticulum related modules were both upregulated in NAFL and NASH compared to normal control. Proportions of fibroblasts and M1 macrophages are positively correlated with fibrosis. Two hub genes *Aebp1* and *Fdft1* may play important roles in fibrosis and steatosis. m6A genes were strongly correlated with the expression of modules. Eight candidate drugs for NAFLD treatment were proposed. Finally, an easy-to-use NAFLD gene co-expression database was developed (available at https://nafld.shinyapps.io/shiny/).

Conclusion: Two gene modules show good performance in stratifying NAFLD patients. The modules and hub genes may provide targets for disease treatment.

KEYWORDS

module eigengene, M6A, connectivity, differential module, ROC

Introduction

Due to the drastically changed living style in modern life, non-alcoholic fatty liver disease (NAFLD) has become an epidemic and imposes a heavy burden on public health. NAFLD encompasses a series of progressive liver diseases developing from simple steatosis (NAFL) to hepatocyte cell death (ballooning) and inflammation (non-alcoholic steatohepatitis, NASH) (Lefebvre et al., 2017). NAFL is generally regarded as a reversible benign condition (Govaere et al., 2020). Mitochondrion plays a role in the phenotypic switching from NAFL to NASH(Pirola et al., 2013). Patients with NASH may progress to cirrhosis and hepatocellular carcinoma (HCC) (Arendt et al., 2015). Although not a prerequisite for diagnosis, fibrosis can also occur and is associated with adverse outcomes (Kozumi et al., 2021).

Transcriptomics is a powerful tool to investigate the expression of thousands of genes concurrently (Liu and Wang, 2020). Large-scale transcriptome data is valuable for developing diagnostic biomarkers, as well as for targeting therapy (Sookoian and Pirola, 2020). There is currently no approved therapy for non-alcoholic steatohepatitis (NASH). The systems biology method may help to dissect the disease mechanisms and to translate basic research into useful treatments (Sookoian and Pirola, 2019). The transcriptome of NAFLD patients has been profiled in several studies (Ahrens et al., 2013; Murphy et al., 2013; Arendt et al., 2015; Hoang et al., 2019; Azzu et al., 2021; Pantano et al., 2021). Some of the studies are based on the microarray, while some are based on state-of-the-art RNA-Seq technology. However, to what extent the results of these studies are reproducible is still not known. A systematic meta-analysis of microarray experiments on liver tissue of NAFLD patients has been performed, and four genes were identified as biomarkers for patients at risk of progression to severe NAFLD (Ryaboshapkina and Hammar, 2017). The study lacks RNA-Seq data. Only one recent report showed that four genes were consistently identified by all six NAFLD transcriptome studies (Pantano et al., 2021). The reason for the phenomenon may include different sample sizes, control samples chosen, platforms used, and data processing methods used (Ye and Liu, 2015).

As it is well known that complex human diseases such as NAFLD and cancer are rarely caused by a single gene but are more likely influenced by a network of interacting genes (Sookoian et al., 2020; Paci et al., 2021). Genetic redundancy accounts that a given biochemical function is redundantly encoded by two or more genes. Therefore, mutations (or defects) in one of these genes will have a smaller effect on the fitness of the organism than expected from the genes' function (Pearce et al., 2004). Differential gene lists from individual studies often lack reproducibility. Module-level differential analysis may help to overcome problem. A gene co-expression module is a more stable unit than a single gene in diseases (Zhou et al., 2019). An example is cancer where the aberrant cell cycle is recurrently present, but the change of a cell cycle-related gene may present in one dataset but not another (Zhou et al., 2019). The rationale of gene co-expression analysis is that gene expressions are correlated. Weighted gene co-expression network analysis (WGCNA) can reduce thousands of genes to tens of modules, which are relatively independent as genes in a module have similar expression patterns but are different from that of other modules. Thus, a module may represent a unique function of a bio-system. Therefore, it is reasonable to reanalyze these valuable datasets by gene co-expression network.

Here, we first applied WGCNA to the currently largest NAFLD tissue dataset GSE135251 which was profiled by RNA-Seq. Identified modules were used as a reference, where new datasets can be projected, making the comparisons between datasets possible. We found a sterol biosynthesis module and a collagen-containing extracellular matrix module. Both modules can efficiently and recurrently separate patients of NAFLD in different datasets. Proportions of fibroblasts and M1 macrophages are positively correlated with fibrosis. Eight candidate drugs were also identified based on the hub genes of differential modules. The analysis workflow is summarized in Figure 1.

Materials and methods

Datasets and preprocessing

A total of 11 NAFLD tissue datasets were downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) and the European Bioinformatics Institute (EBI) ArrayExpress. These datasets were listed in Table 1. Among these datasets, four are from RNA-Seq and seven are from the microarray. GSE135251 is currently the largest NAFLD dataset available from public repositories, which contains 216 NAFLD samples across the disease spectrum. To focus on the informative genes, the count matrix provided by the database was filtered with a mean count >200 and a standard deviation >0.1 before downstream analysis. Finally, 7,773 genes were retained for gene co-expression network analysis. Additional two datasets (GSE200186 and GSE119281) were used for drug targets validation. As datasets listed in Table 1 are from different studies and platforms and may have different clinical/histological information available, these data were analyzed individually not in a pooled form.

Weighted gene co-expression network analysis (WGCNA)

Gene co-expression module identification was performed according to the package manual (Langfelder and Horvath, 2008). Parameters were set as following: softPower = 16, corOptions = list (use = 'p', method = 'spearman'), networkType = "signed", minModuleSize = 30, deepSplit = 4, MEDissThres = 0.2. Briefly, the pairwise Spearman correlation coefficient was calculated for each gene in the gene expression matrix, and then an adjacency matrix was derived by raising the correlation matrix to a power 16, which generated a biologically meaningful scale-free network. The weighted network was transformed into a network of topological overlap (TO)—a metric that defines the relationship of two genes accounting for their correlation and shared neighborhood. Genes were hierarchically clustered based on their TO. Finally, co-expression gene modules were identified by the Dynamic Tree Cut algorithm. As genes in a module are highly correlated,

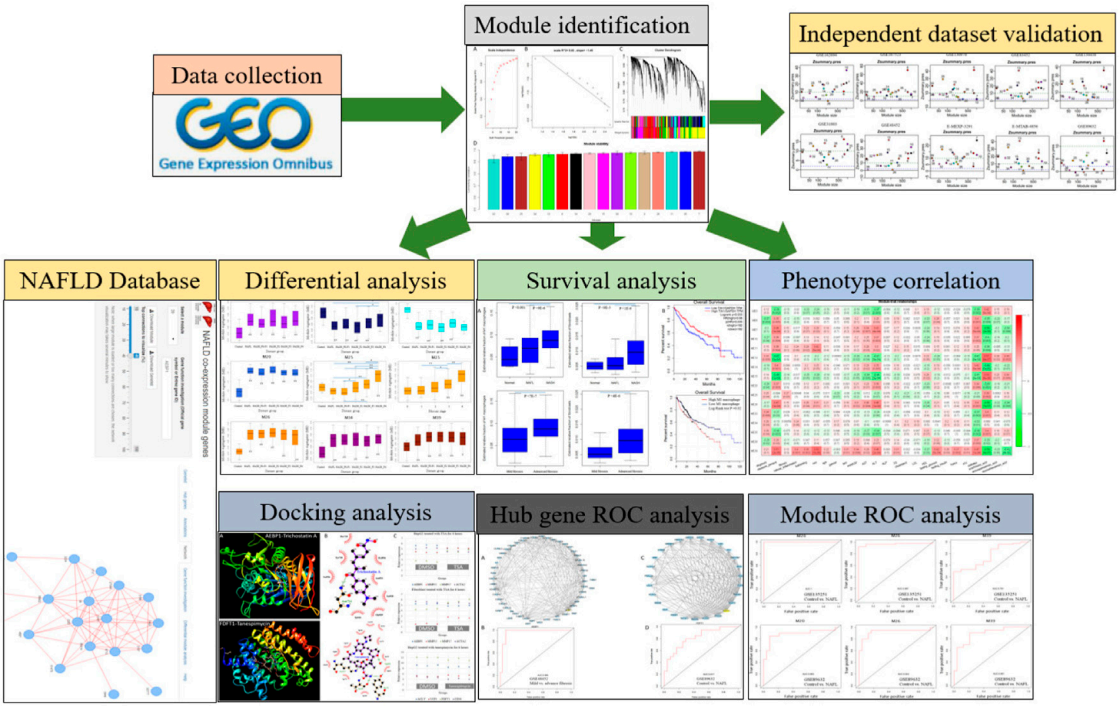


FIGURE 1
A schematic diagram for the analysis framework. NAFLD transcriptome datasets were collected and analyzed by gene co-expression network analysis. Independent datasets were used for network validation. Based on the identified network modules, the NAFLD database was constructed. Module-level differential analysis, ROC analysis, and phenotype correlation were performed. Module hub genes were mined for potential drug discovery.

TABLE 1 Datasets used in the study.

Accession	Sample	Platform	Clinical parameters
GSE135251	206 NAFLD, 10 controls	Illumina NextSeq 500	NAS, fibrosis
GSE162694	143 NASH patients of various fibrosis stages	Illumina HiSeq 3,000	Age, sex, fibrosis, NAS
GSE167523	98 NAFLD patients	Illumina HiSeq 3,000	Disease subtype, age, gender
GSE130970	72 NAFLD, 6 control	Illumina HiSeq 2,500	Sex, age, lobular inflammation grade, cytological ballooning grade, steatosis grade, NAS, fibrosis
GSE83452	126 NASH samples, 66 no NASH	Affymetrix Human Gene 2.0 ST Array	Age, gender
GSE134438	43 NAFLD patients	Affymetrix Human Gene 2.0 ST Array	Disease subtype
GSE49541	72 NAFLD samples	Affymetrix Human Genome U133 Plus 2.0 Array	Fibrosis
GSE48452	32 NAFLD, 41 control	Affymetrix Human Gene 1.1 ST Array	Fat, inflammation, sex, age, BMI, NAS, fibrosis, leptin, adiponectin
E-MEXP-3291	26 NAFLD samples, 19 control	Affymetrix GeneChip Human Gene 1.0 ST Array	Age, sex, diagnosis
E-MTAB-4856	88 NAFLD samples	Agilent Whole Human Genome Microarray 4 × 44K 014,850 G4112F	Age, sex, disease staging, BMI
GSE89632	29 NAFLD, 24 control	Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip	Diagnosis, steatosis, fibrosis, lobular inflammation, ballooning, NAS, age, gender, BMI, waist, AST, ALT, ALP, TG, TC, LDL, HDL, FPG, fasting insulin, HOMA-IR, HbA1c, diabetes

NAS: NAFLD, activity score, BMI: body mass index, AST: aspartate transaminase, ALT: alanine transaminase, ALP: alkaline phosphatase, TG: triglycerides, TC: total cholesterol, LDL: low-density lipoprotein cholesterol, HDL: high-density lipoprotein cholesterol, FPG: fasting plasma glucose, HOMA-IR: homa-insulin resistance, HbA1c: hemoglobin a1c.

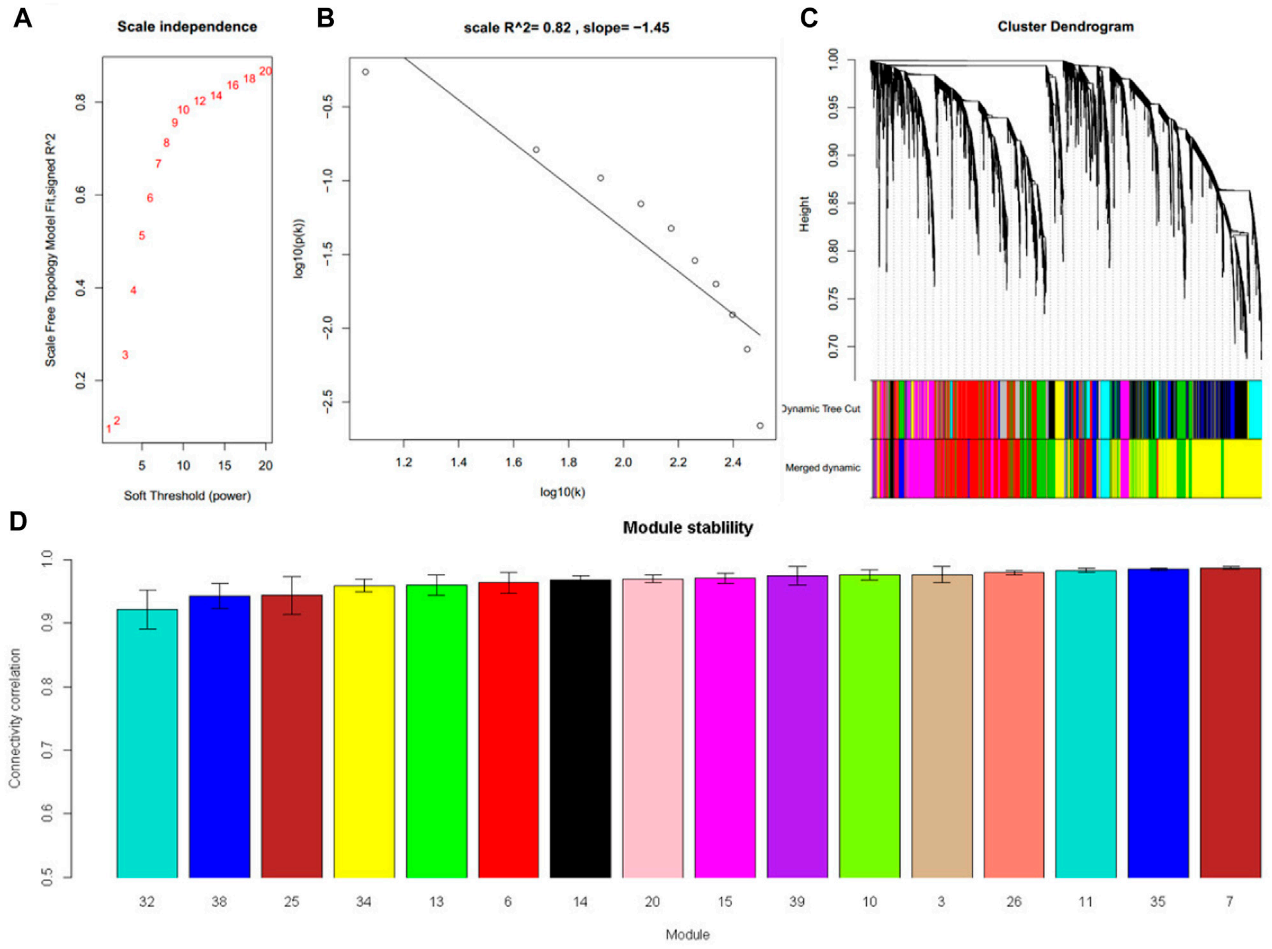


FIGURE 2
Sixteen modules were identified in the NAFLD dataset GSE135251. **(A)** The relationship between choosing power and corresponding scale-free topology model fit R^2 **(B)** When power was set at 16, the constructed network follows a power law. **(C)** Cluster dendrogram shows the partition of genes into co-expressed modules with different colors **(D)** The module stability was tested by sampling half of the samples. The stability was expressed as the correlation of intramodule connectivity between the original one and the sampled one.

module genes can be reduced to a module eigengene (ME) by singular value decomposition. ME represents the first principal component of module expression profiles (Zhang and Horvath, 2005). WGCNA also provides gene connectivity information, which is the sum of correlations of a gene with all other genes in the module or network. Hub gene in a co-expression module tends to have high connectivity and may play important roles in the network or module. For network module validation, the expression matrix was first intersected with the reference dataset, then its values were transformed to ranks before module projection. Module stability was tested by 1,000 half-samplings for each module (Liu et al., 2018a). The stability was presented by the correlation of intra-module connectivity between the original one and the sampled one in form of mean ± standard deviation. To test the reproducibility of these modules, other datasets were projected to the frozen reference for module preservation analysis (Langfelder et al., 2011). Parameters for module preservation were set as networkType = “signed”, nPermutations = 100. The module-level expression for other datasets was retrieved by the moduleEigengenes function.

Functional annotation of the modules

The gProfileR package was used for enrichment analysis of reference modules (Reimand et al., 2016). For drug screening, hub genes of differential modules were submitted to the Connectivity Map (https://portals.broadinstitute.org/cmap) (Lamb et al., 2006). Significant results were retrieved at the level of $p < 0.05$. Protein-ligand docking was performed in SwissDock (http://www.swissdock.ch/docking) (Grosdidier et al., 2011). LigPlot+ was used to generate a pose view of protein-ligand interaction (Laskowski and Swindells, 2011). The proportion of immune cell populations was estimated by TIMER2.0 (http://timer.comp-genomics.org/) (Li et al. , 2020). Immunohistochemistry images for hub genes *Aebp1* and *Fdft1* were retrieved from the Human Protein Atlas database (https://www.proteinatlas.org/). Survival analysis for genes and

TABLE 2 Functional enrichment analysis for the sixteen modules identified in the NAFLD dataset GSE135251.

Module (No. genes)	Function (*p*-value)	Hub	TIMER cell type with highest correlation
3 (813)	Nucleus (2E-30) Factor: E2F (9E-15)	*Taf1c*	Memory CD4 T cells
6 (813)	Ribosome (3E-92) Translation (2E-81)	*Ndufa2*	Macrophage
7 (2,410)	Factor: HDAC2 (1E-61) hsa-miR-21–5p (2E-54)	*Tmem106b*	Th1 CD4 T cell (−)
10 (1,621)	Factor: Churchill (8E-76) Factor: Sp1 (1E-75)	*Anapc2*	Th1 CD4 T cell
11 (205)	Vesicle (2E-13) Mitochondrion (5E-13)	*Tm9sf2*	Memory CD4 T cells (−)
13 (171)	Immune response (1E-21) Myeloid leukocyte activation (3E-16)	*Mpeg1*	Macrophage
14 (170)	Mitochondrion (5E-13) Factor: ER81 (2E-11)	*Rbis*	Lymphoid progenitor
15 (153)	Nucleus (8E-17) RNA metabolism (1E-15)	*Tia1*	M1 Macrophage (−)
20 (337)	Vesicle (2E-18) Endoplasmic reticulum (5E-11) Sp1 (3E-10) hsa-miR-484 (5E-10)	*Clptm1*	Lymphoid progenitor (−)
25 (69)	Collagen-containing extracellular matrix (1E-26)	*Aebp1*	Fibroblast
26 (234)	Endomembrane system (6E-9) Endoplasmic reticulum (2E-4)	*Kctd20*	Macrophage (−)
32 (54)	Leukocyte activation involved in immune response (7E-8)	*Stk10*	Microenvironment score*
34 (92)	Nucleoplasm (3E-6)	*Trrap*	Lymphoid progenitor (−)
35 (386)	hsa-miR-21–5p (1E-12) Factor: ETF (2E-9)	*Dennd4c*	Granulocyte monocyte progenitor
38 (33)	Translation (1E-11) Structural constituent of ribosome (3E-10) Mitochondrial inner membrane (7E-10)	*Stoml2*	CD4 T cell (−)
39 (30)	Sterol biosynthesis (8E-39) Factor: YB-1 (1E-5)	*Fdft1*	

(−), the estimated fraction of cell type is negatively correlated with module eigengene (ME); All the correlations in the “TIMER, cell type with highest correlation” column are >0.6 and $p < 0.01$.
[a]Microenvironment score, the sum of all immune and stromal cell types.

immune cells was retrieved from the GEPIA2021 database (http://gepia2021.cancer-pku.cn/).

Statistical analysis

Differential module analyses were performed using the Student’s t-test or ANOVA in R package aov. Tukey’s HSD was used to calculate *p* values of multiple pairwise comparisons. An adjusted *p*-value smaller than 0.05 was considered statistically significant.

Results

Sixteen modules were identified in the NAFLD

RNA-Seq dataset GSE135251 was used to construct a frozen reference network, to which other datasets would be projected. A scale-free network was constructed (Figure 2A, B), and then a total of 16 gene co-expression modules were identified (Figure 2C). The top hub gene with high connectivity for each module was provided in Table 2. Functional annotation shows that these modules were associated with nucleus, translation, transcription factors, vesicle, immune response, mitochondrion, collagen, and sterol biosynthesis (Table 2). For a full list of genes, their assigned modules, connectivity, and gene description, readers can refer to Supplementary Table S1.

In module reproducibility analysis, all the modules had an average connectivity correlation larger than 0.9 (Figure 2D). Ten NAFLD datasets from different platforms (Illumina, Affymetrix, and Agilent in Table 1) were projected on the frozen reference modules to test reproducibility (Figure 3). All of the modules have an average Zsummary. pres statistic larger than 4.4 and the average Zsummary. pres statistic of all modules was 12.4, indicating very strong preservation of modules except GSE89632 which has weak to moderate preservation Zsummary. pres statistic.

Modules are correlated with clinical parameters

By correlating modules with clinical parameters, we can identify which module contributes to disease. Module-trait relationship analysis revealed several connections. M25 is positively correlated with fibrosis in all six datasets that provide fibrosis data (Supplementary Table S2). As dataset GSE89632 has abundant clinical parameters, thus we correlated module expression and these parameters in the dataset. M39 and M35 are positively and negatively correlated with hepatic fat content in two datasets. M39 is positively correlated with NAFLD activity score (NAS), fasting glucose, and diabetes status in GSE89632. M13 and M39 are positively and negatively correlated with arachidonic acid (Figure 4). Module-trait relationship results for the other datasets are provided in Supplementary Figure S1. We also performed TIMER analysis to estimate the immune cell populations in the

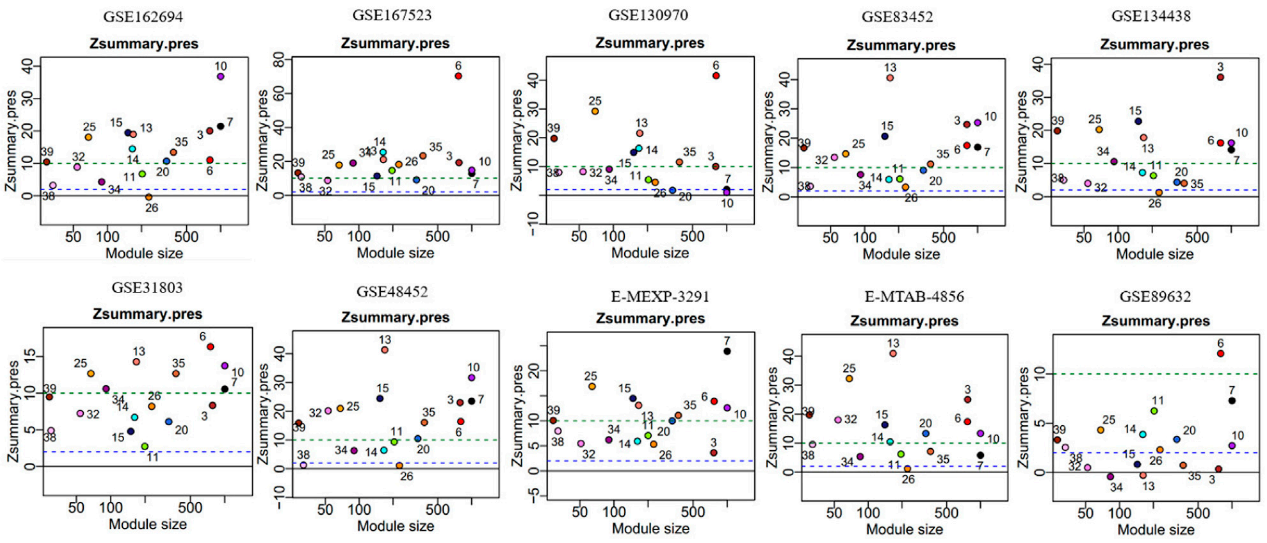
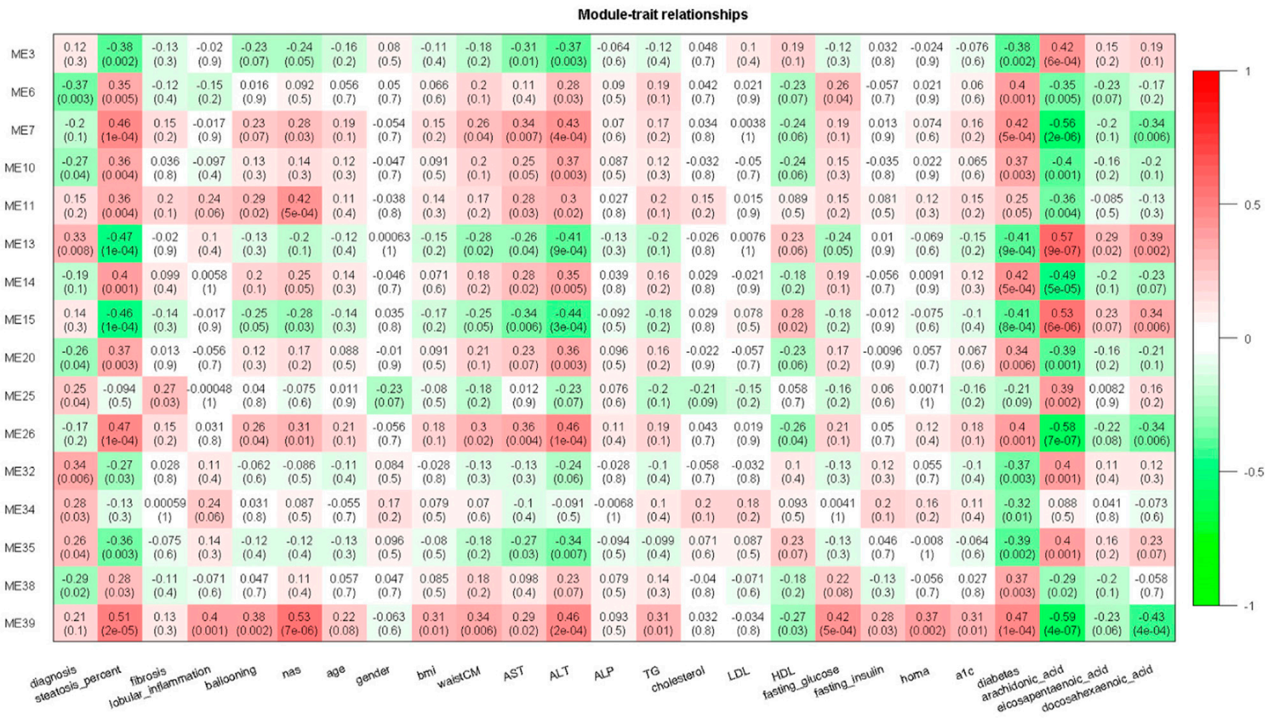


FIGURE 3
Gene network modules from GSE135251 are well preserved in the other ten datasets. The *y*-axis represents preservation statistics and the *x*-axis is the number of genes in each module. The dashed blue and green lines indicate the thresholds Z = 2 and Z = 10, respectively. Zsummary <2 implies no evidence for module preservation, 2 < Zsummary <10 implies weak to moderate evidence, and Zsummary >10 implies strong evidence for module preservation.



	diagnosis	steatosis_percent	fibrosis	lobular_inflammation	ballooning	nas	age	gender	bmi	waistCM	AST	ALT	ALP	TG	cholesterol	LDL	HDL	fasting_glucose	fasting_insulin	homa	a1c	diabetes	arachidonic_acid	eicosapentaenoic_acid	docosahexaenoic_acid
ME3	0.12 (0.3)	-0.38 (0.002)	-0.13 (0.3)	-0.02 (0.9)	-0.23 (0.07)	-0.24 (0.05)	-0.16 (0.2)	0.08 (0.5)	-0.11 (0.4)	-0.18 (0.2)	-0.31 (0.01)	-0.37 (0.003)	-0.064 (0.6)	-0.12 (0.4)	0.048 (0.7)	0.1 (0.4)	0.19 (0.1)	-0.12 (0.3)	0.032 (0.8)	-0.024 (0.9)	-0.076 (0.6)	-0.38 (0.002)	0.42 (6e-04)	0.15 (0.2)	0.19 (0.1)
ME6	-0.37 (0.003)	0.35 (0.005)	-0.12 (0.4)	-0.15 (0.2)	0.016 (0.9)	0.092 (0.5)	0.056 (0.7)	0.05 (0.7)	0.066 (0.6)	0.2 (0.1)	0.11 (0.4)	0.28 (0.03)	0.09 (0.5)	0.19 (0.1)	0.042 (0.7)	0.021 (0.9)	-0.23 (0.07)	0.26 (0.04)	-0.057 (0.7)	0.021 (0.9)	0.06 (0.6)	0.4 (0.001)	-0.35 (0.005)	-0.23 (0.07)	-0.17 (0.2)
ME7	-0.2 (0.1)	0.46 (1e-04)	0.15 (0.2)	-0.017 (0.9)	0.23 (0.07)	0.28 (0.03)	0.19 (0.1)	-0.054 (0.7)	0.15 (0.2)	0.26 (0.04)	0.34 (0.007)	0.43 (4e-04)	0.07 (0.6)	0.17 (0.2)	0.034 (0.8)	0.0038 (1)	-0.24 (0.06)	0.19 (0.1)	0.013 (0.9)	0.074 (0.6)	0.16 (0.2)	0.42 (5e-04)	-0.56 (2e-06)	-0.2 (0.1)	-0.34 (0.006)
ME10	-0.27 (0.04)	0.36 (0.004)	0.036 (0.8)	-0.097 (0.4)	0.13 (0.3)	0.14 (0.3)	0.12 (0.3)	-0.047 (0.7)	0.091 (0.5)	0.2 (0.1)	0.25 (0.05)	0.37 (0.003)	0.087 (0.5)	0.12 (0.3)	-0.032 (0.8)	-0.05 (0.7)	-0.24 (0.06)	0.15 (0.3)	-0.035 (0.8)	0.022 (0.9)	0.065 (0.6)	0.37 (0.003)	-0.4 (0.001)	-0.16 (0.2)	-0.2 (0.1)
ME11	0.15 (0.2)	0.36 (0.004)	0.2 (0.1)	0.24 (0.06)	0.29 (0.02)	0.42 (5e-04)	0.11 (0.4)	-0.038 (0.8)	0.14 (0.3)	0.17 (0.2)	0.28 (0.03)	0.3 (0.02)	0.027 (0.8)	0.2 (0.1)	0.15 (0.2)	0.015 (0.9)	0.089 (0.5)	0.15 (0.2)	0.081 (0.5)	0.12 (0.3)	0.15 (0.2)	0.25 (0.05)	-0.36 (0.004)	-0.085 (0.5)	-0.13 (0.3)
ME13	0.33 (0.008)	-0.47 (1e-04)	-0.02 (0.9)	0.1 (0.4)	-0.13 (0.3)	-0.2 (0.1)	-0.12 (0.4)	0.00063 (1)	-0.15 (0.2)	-0.28 (0.02)	-0.26 (0.04)	-0.41 (9e-04)	-0.13 (0.3)	-0.2 (0.1)	-0.026 (0.8)	0.0076 (1)	0.23 (0.06)	-0.24 (0.05)	0.01 (0.9)	-0.069 (0.6)	-0.15 (0.2)	-0.41 (9e-04)	0.57 (9e-07)	0.29 (0.02)	0.39 (0.002)
ME14	-0.19 (0.1)	0.4 (0.001)	0.099 (0.4)	0.0058 (1)	0.2 (0.1)	0.25 (0.05)	0.14 (0.3)	-0.046 (0.7)	0.071 (0.6)	0.18 (0.2)	0.28 (0.02)	0.35 (0.005)	0.039 (0.8)	0.16 (0.2)	0.029 (0.8)	-0.021 (0.9)	-0.18 (0.2)	0.19 (0.1)	-0.056 (0.7)	0.0091 (0.9)	0.12 (0.3)	0.42 (5e-04)	-0.49 (5e-05)	-0.2 (0.1)	-0.23 (0.07)
ME15	0.14 (0.3)	-0.46 (1e-04)	-0.14 (0.3)	-0.017 (0.9)	-0.25 (0.05)	-0.28 (0.03)	-0.14 (0.3)	0.035 (0.8)	-0.17 (0.2)	-0.25 (0.05)	-0.34 (0.006)	-0.44 (3e-04)	-0.092 (0.5)	-0.18 (0.2)	0.029 (0.8)	0.078 (0.5)	0.28 (0.02)	-0.18 (0.2)	-0.012 (0.9)	-0.075 (0.6)	-0.1 (0.4)	-0.41 (8e-04)	0.53 (6e-06)	0.23 (0.07)	0.34 (0.006)
ME20	-0.26 (0.04)	0.37 (0.003)	0.013 (0.9)	-0.056 (0.7)	0.12 (0.3)	0.17 (0.2)	0.088 (0.5)	-0.01 (0.9)	0.091 (0.5)	0.21 (0.1)	0.23 (0.07)	0.36 (0.003)	0.096 (0.5)	0.16 (0.2)	-0.022 (0.9)	-0.057 (0.7)	-0.23 (0.06)	0.17 (0.2)	-0.0096 (0.9)	0.057 (0.7)	0.067 (0.6)	0.34 (0.006)	-0.39 (0.001)	-0.16 (0.2)	-0.21 (0.1)
ME25	0.25 (0.04)	-0.094 (0.5)	0.27 (0.03)	-0.00048 (1)	0.04 (0.8)	-0.075 (0.6)	0.011 (0.9)	-0.23 (0.07)	-0.08 (0.5)	-0.18 (0.2)	0.012 (0.9)	-0.23 (0.07)	0.076 (0.6)	-0.2 (0.1)	-0.21 (0.09)	-0.15 (0.2)	0.058 (0.7)	-0.16 (0.2)	0.06 (0.6)	0.0071 (1)	-0.16 (0.2)	-0.21 (0.09)	0.39 (0.002)	0.0082 (0.9)	0.16 (0.2)
ME26	-0.17 (0.2)	0.47 (1e-04)	0.15 (0.2)	0.031 (0.8)	0.26 (0.04)	0.31 (0.01)	0.21 (0.1)	-0.056 (0.7)	0.18 (0.1)	0.3 (0.02)	0.36 (0.004)	0.46 (1e-04)	0.11 (0.4)	0.19 (0.1)	0.043 (0.7)	0.019 (0.9)	-0.26 (0.04)	0.21 (0.1)	0.05 (0.7)	0.12 (0.4)	0.18 (0.1)	0.4 (0.001)	-0.58 (7e-07)	-0.22 (0.08)	-0.34 (0.006)
ME32	0.34 (0.006)	-0.27 (0.03)	0.028 (0.8)	0.11 (0.4)	-0.062 (0.6)	-0.086 (0.5)	-0.11 (0.4)	0.084 (0.5)	-0.028 (0.8)	-0.13 (0.3)	-0.13 (0.3)	-0.24 (0.06)	-0.028 (0.8)	-0.1 (0.4)	-0.058 (0.7)	-0.032 (0.8)	0.1 (0.4)	-0.13 (0.3)	0.12 (0.3)	0.055 (0.7)	-0.1 (0.4)	-0.37 (0.003)	0.4 (0.001)	0.11 (0.4)	0.12 (0.3)
ME34	0.28 (0.03)	-0.13 (0.3)	0.00059 (1)	0.24 (0.06)	0.031 (0.8)	0.087 (0.5)	-0.055 (0.7)	0.17 (0.2)	0.079 (0.5)	0.07 (0.6)	-0.1 (0.4)	-0.091 (0.5)	-0.0068 (1)	0.1 (0.4)	0.2 (0.1)	0.18 (0.2)	0.093 (0.5)	0.0041 (1)	0.2 (0.1)	0.16 (0.2)	0.11 (0.4)	-0.32 (0.01)	0.088 (0.5)	0.041 (0.8)	-0.073 (0.6)
ME35	0.26 (0.04)	-0.36 (0.003)	-0.075 (0.6)	0.14 (0.3)	-0.12 (0.4)	-0.12 (0.4)	-0.13 (0.3)	0.096 (0.5)	-0.08 (0.5)	-0.18 (0.2)	-0.27 (0.03)	-0.34 (0.007)	-0.094 (0.5)	-0.099 (0.4)	0.071 (0.6)	0.087 (0.5)	0.23 (0.07)	-0.13 (0.3)	0.046 (0.7)	-0.008 (1)	-0.064 (0.6)	-0.39 (0.002)	0.4 (0.001)	0.16 (0.2)	0.23 (0.07)
ME38	-0.29 (0.02)	0.28 (0.03)	-0.11 (0.4)	-0.071 (0.6)	0.047 (0.7)	0.11 (0.4)	0.057 (0.7)	0.047 (0.7)	0.085 (0.5)	0.18 (0.2)	0.098 (0.4)	0.23 (0.07)	0.079 (0.5)	0.14 (0.3)	-0.04 (0.8)	-0.071 (0.6)	-0.18 (0.2)	0.22 (0.08)	-0.13 (0.3)	-0.056 (0.7)	0.027 (0.8)	0.37 (0.003)	-0.29 (0.02)	-0.2 (0.1)	-0.058 (0.7)
ME39	0.21 (0.1)	0.51 (2e-05)	0.13 (0.3)	0.4 (0.001)	0.38 (0.002)	0.53 (7e-06)	0.22 (0.08)	-0.063 (0.6)	0.31 (0.01)	0.34 (0.006)	0.29 (0.02)	0.46 (2e-04)	0.093 (0.5)	0.31 (0.01)	0.032 (0.8)	-0.034 (0.8)	-0.27 (0.03)	0.42 (5e-04)	0.28 (0.03)	0.37 (0.002)	0.31 (0.01)	0.47 (1e-04)	-0.59 (4e-07)	-0.23 (0.06)	-0.43 (4e-04)

FIGURE 4
Module-trait relationship heatmap for GSE89632. Cells represent the correlation between modules expression and clinical parameters. Numbers in the bracket indicate the statistical significance.

liver tissues of patients with NAFLD dataset GSE135251. Fibroblasts and M1 macrophages are positively correlated with fibrosis. The correlation coefficients are 0.42 and 0.35, and the *p* values are 2E-10 and 1E-7 respectively. Further analysis shows that the two components are upregulated during NAFLD progression. Patients with NASH and advanced fibrosis had greater fractions of fibroblasts

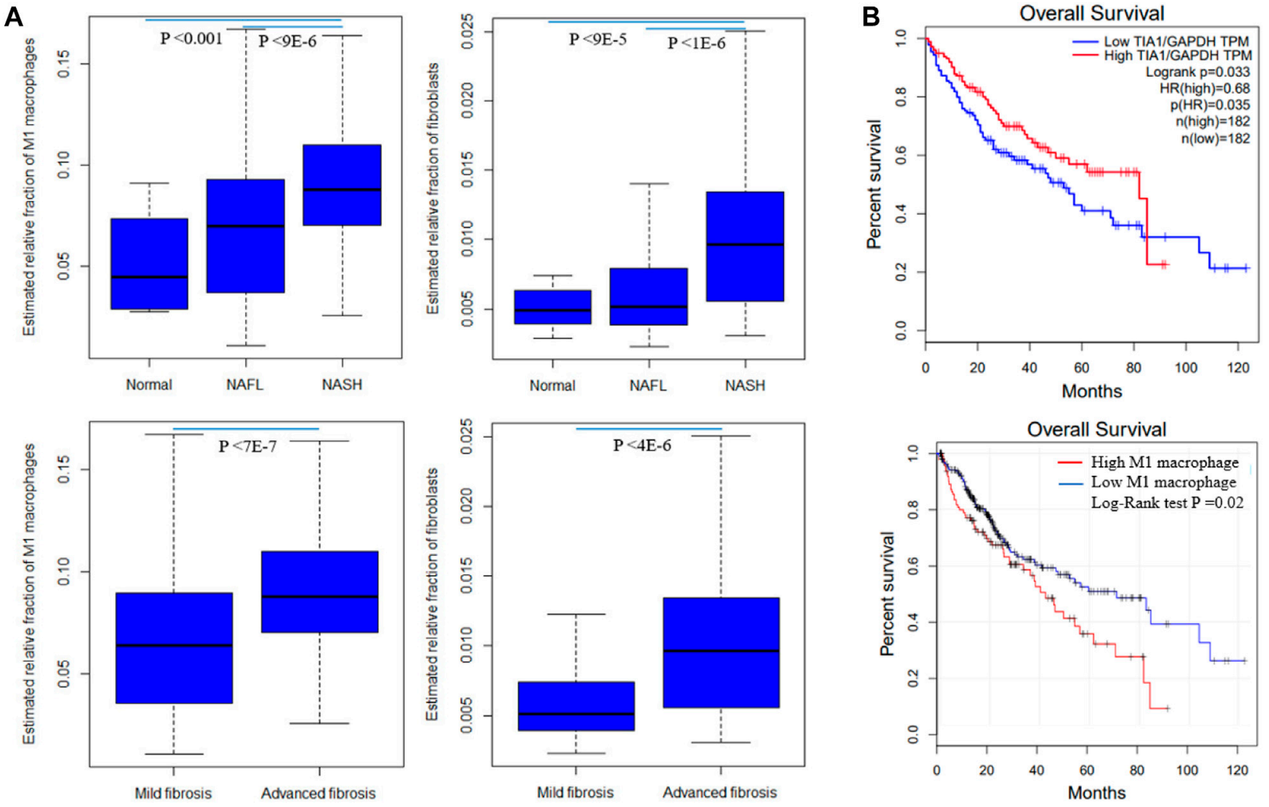


FIGURE 5
Fibroblast and M1 macrophage in NAFLD and HCC. **(A)** Differentially expressed fractions of M1 macrophages and fibroblasts in patients with different NAFLD stages and fibrosis stages in GSE135251 **(B)** M1 macrophage and hub gene *Tia1* are associated with HCC patients survival in the TCGA dataset. *Tia1* is the hub gene of M15 which is negatively correlated with M1 macrophage proportion. The log-rank derived *p*-value indicates the significance of the comparison between groups. Statistical significance was calculated by Student's t-test for mild and advanced fibrosis comparison, or ANOVA for normal, NAFL, and NASH comparison.

and M1 macrophages than patients with NAFL and mild fibrosis (Figure 5A). The fraction of M2 macrophages is downregulated during NAFLD progression but not significant. NAFLD-related HCC has been reported in a significant number of patients (Margini and Dufour, 2016). Thus, we also analyzed the two components in the TCGA HCC dataset and found that low M1 macrophage fraction group patients had longer overall survival (Figure 5B). M15 is negatively correlated with M1 macrophage proportion (Table 2). M15 hub gene TIA1 expression suggests a positive contribution to HCC patient survival (Figure 5B).

Differentially expressed modules associated with steatosis, fibrosis, and steatohepatitis

M25 was consistently upregulated in advanced fibrosis compared to mild fibrosis in five datasets (*t*-test, $p < 0.05$). In GSE135251 only M25 was differently expressed across fibrosis stages (Figure 6). M14 and M20 were the most significantly upregulated in NAFL samples compared to the control. M25 and M39 were the only two differential modules upregulated in NASH compared to NAFL ($p = 0.0003$ and $p = 0.02$). M25 and M14 were the most significantly upregulated and downregulated in NASH samples compared to no NASH ($p = 0.001$ and $p = 0.04$) (Figure 7). The box figures for all the modules can be viewed in the "Differential module analysis" tab of the NAFLD database at http://nafld.shinyapps.io/shiny/.

Modules can consistently separate NAFLD patients

To check if these modules can efficiently separate patients, we calculated specificity and sensitivity to plot the receiver operating characteristic (ROC) curve. Modules such as M20, M26, and M39 can recurrently separate control and NAFL patients in different datasets with AUCs ranging from 0.76 to 1.00 (Supplementary Figure S2). M20, M25, M26, M32 can separate samples from NAFL and NASH with AUCs larger than 0.88 (Supplementary Figure S3). M25 could consistently separate samples from mild fibrosis and advanced fibrosis in six datasets with AUCs 0.89, 0.96, 0.97, 0.84, 0.82, and 0.76. (Supplementary Figure S4).

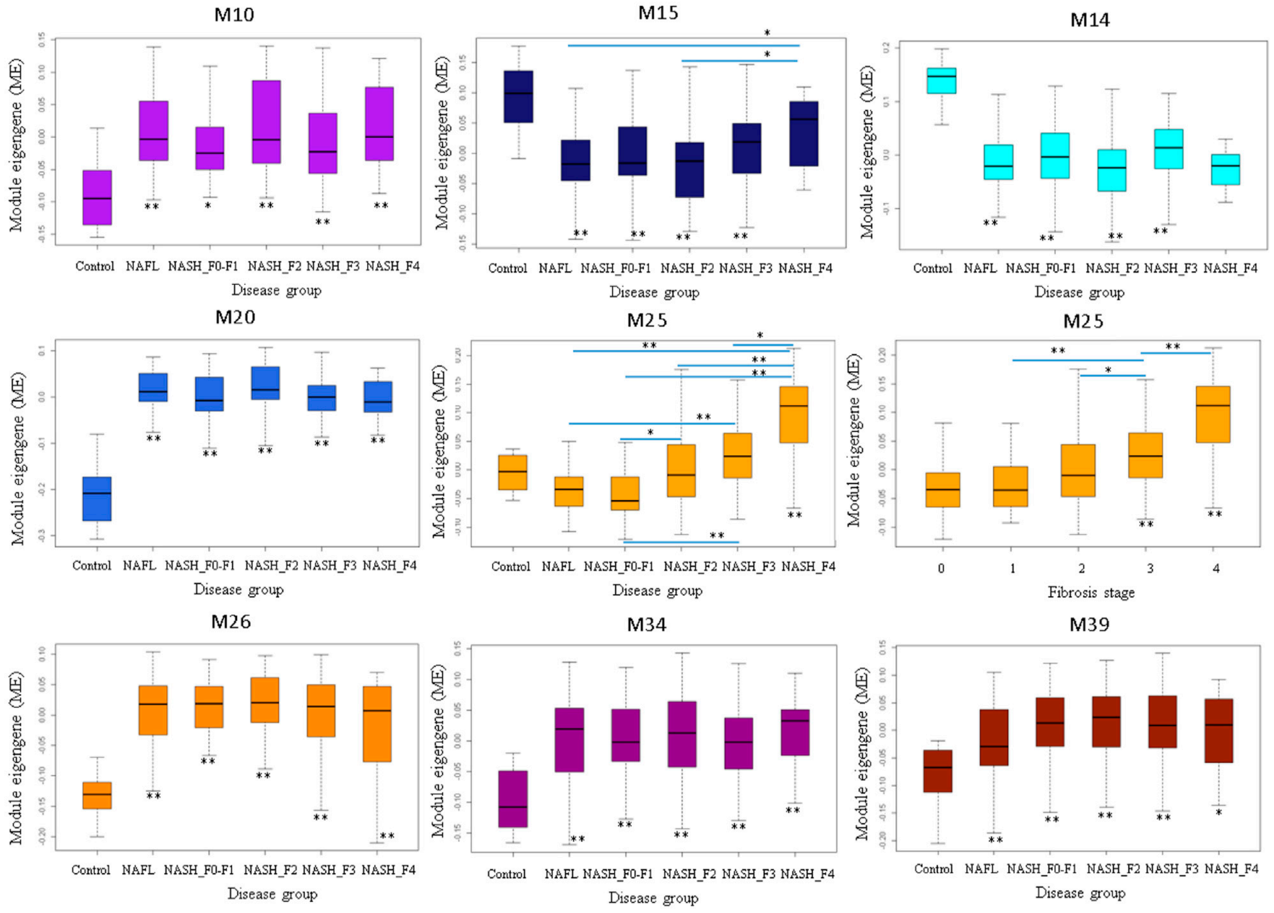


FIGURE 6
Differentially expressed modules associated with steatohepatitis stages in GSE135251. The asterisk under the box indicates the significance of the comparison between disease and normal control by ANOVA test with adjusted p values. *: $p < 0.05$, **: $p < 0.01$.

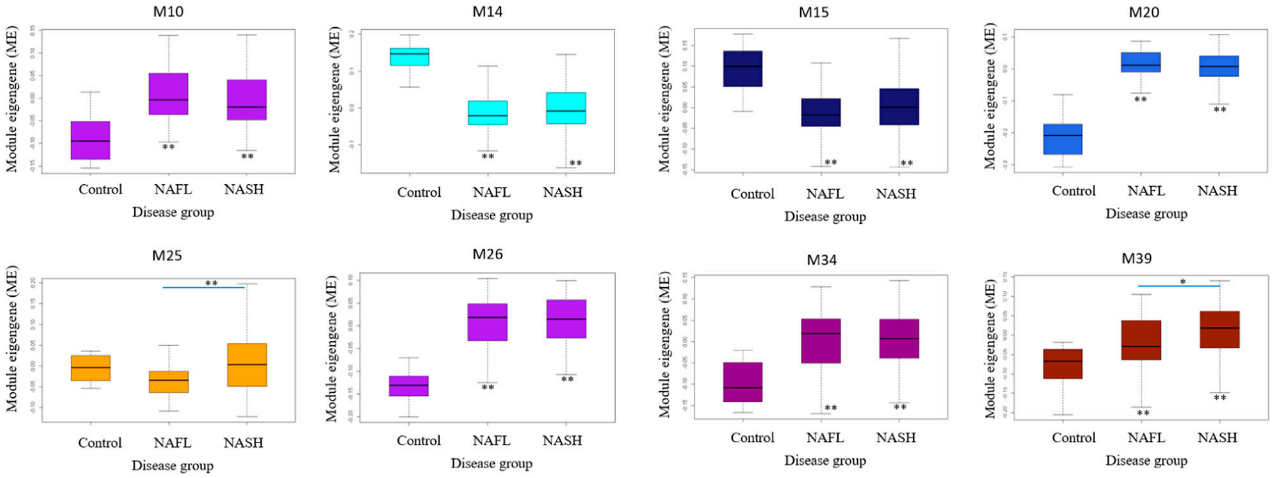


FIGURE 7
Differentially expressed modules associated with NAFLD disease stages in GSE135251. The asterisk under the box indicates the significance of the comparison between disease and normal control by ANOVA test with adjusted p values. *: $p < 0.05$, **: $p < 0.01$.

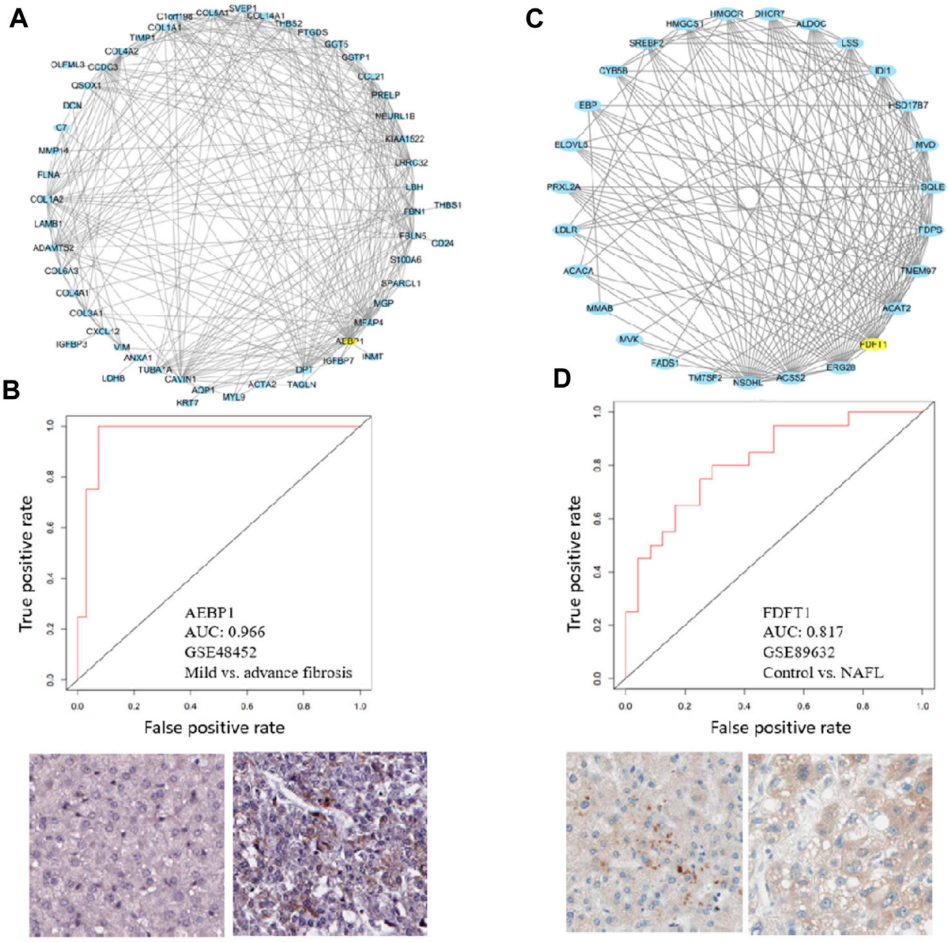


FIGURE 8
Hub genes in M25 and M39 can separate patients. **(A, C)** Module visualization for the top 100 connections within M25 and M27 by using Cytoscape. Yellow nodes indicate the hub gene of modules **(B, D)** ROC curves show that hub genes of M25 and M39, *Aebp1*, and *Fdft1* can separate advanced fibrosis or NAFL patients. Immunohistochemical plots also show the positive staining of AEBP1 and FDFT1 in normal and HCC tissues.

Hub genes are important in NAFLD progression

To demonstrate the utility of the co-expression modules identified, two important modules M25 and M39 were visualized. *Aebp1* is the hub gene of M25 (Figure 8A). It has been reported that adipocyte enhancer-binding protein 1 (AEBP1) is a ubiquitously expressed, multifunctional protein and is a transcriptional repressor involved in adipogenesis, inflammation, cholesterol homeostasis, and atherogenesis (Majdalawieh et al., 2020). AEBP1 expression increases with the severity of fibrosis in NASH possibly by encoding the aortic carboxypeptidase-like protein (ACLP) that associates with collagens in the extracellular matrix (ECM) (Blackburn et al., 2018; Gerhard et al., 2019). In our analysis, M25 indeed is associated with fibrosis, and *Aebp1* is highly correlated with ECM genes *Col1a1, Col1a2, Col3a1, Col4a1, Col4a2, Col5a1, Col6a3, And Col14a1* (Figure 8A). *Aebp1* has a good performance in separating mild and advanced fibrosis patients (Figure 8B). *Fdft1* is the hub gene of M39 (Figure 8C). *Fdft1* is one of the causative loci for steatosis, NAS, degree of fibrosis, lobular inflammation, and serum levels of alanine aminotransferase (ALT) (Stattermayer et al., 2014; Sharma et al., 2015). Interestingly, we found that M39 is highly correlated with steatosis, NAS, lobular inflammation, ALT, and diabetes (Figure 4). *Fdft1* has a moderate performance in separating healthy control and NAFL patients (Figure 8D). The two proteins were positively stained in HCC liver tissues. As hepatic fibrosis presents in a majority of HCC patients, we also checked the two gene expression in TCGA pan-cancer datasets and found that the two genes were not liver-specific genes and were not prognostic for HCC overall survival (Supplementary Figure S5). All the module information can be explored at the NAFLD co-expression database at http://nafld.shinyapps.io/shiny/, where users can browse gene lists, hub genes, functional annotation, network, correlation analysis, and differential module expression information.

Correlation between modules and enzymes regulating m^{6}A mRNA methylation

N6-methyladenosine (m^{6}A) modification contributes to metabolic reprogramming in NAFLD (Qin et al., 2021). To discover potential m^{6}A regulators of modules, we selected

TABLE 3 Candidate NAFLD drugs identified by Connectivity Map.

No.	Drug	*p*-value
1	Trichostatin A	2E-7
2	Geldanamycin	4E-4
3	Tanespimycin	0.015
4	Diphenylpyraline	0.024
5	6-bromoindirubin-3′-oxime	0.024
6	Pha-00851261E	0.028
7	Lisuride	0.028
8	Clebopride	0.029

24 related genes within the dataset and correlated them with MEs by Spearman correlation. Some of the m^{6}A genes were highly correlated with module expression. For example, YTHDF3, YTHDC2, METTL14, CBLL1, and M7 all had significant correlations larger than 0.75. YTHDF3, YTHDC2, METTL14, CBLL1, and M10 had significant correlations smaller than −0.70. FTO-M26, Eif3A-M35, and VIRMA-M35 all had significant correlations larger than 0.70 (Supplementary Table S3). These results were consistent with the reverse correlations with Th1 CD4 T cells between M7 and M10. M26 was correlated with macrophage proportion, which is important in NASH. *Fto* is correlated with M26, indicating the role of *Fto* in NASH progression.

In silico drugs screening based on highly connected genes in differentially expressed modules

To identify candidate drugs for NAFLD, we selected ten genes from differentially expressed modules and submitted them to the Connectivity Map tool. As listed in Table 3, eight candidates were found with $p < 0.05$. Protein-ligand docking analysis showed the interactions between hub genes *Aebp1* (M25), *Fdft1* (M39) and Trichostatin A (TSA), Tanespimycin (Figure 9A). Pose view analysis showed that the interactions may occur at sites Asn724 in AEBP1 and Lys117, Thr50, Tyr73 in FDFT1 (Figure 9B). Independent datasets were used to validate the results. Interestingly, TSA can significantly downregulated *Aebp1* in HepaG2 and fibroblast cell lines. Additionally, the expression of fibrosis-related genes *Mmp15, Mmp17,* and *Acta2* also changed after 6 h TSA treatment (Figure 9C). As for tanespimycin, *Fdft1* was significantly downregulated after

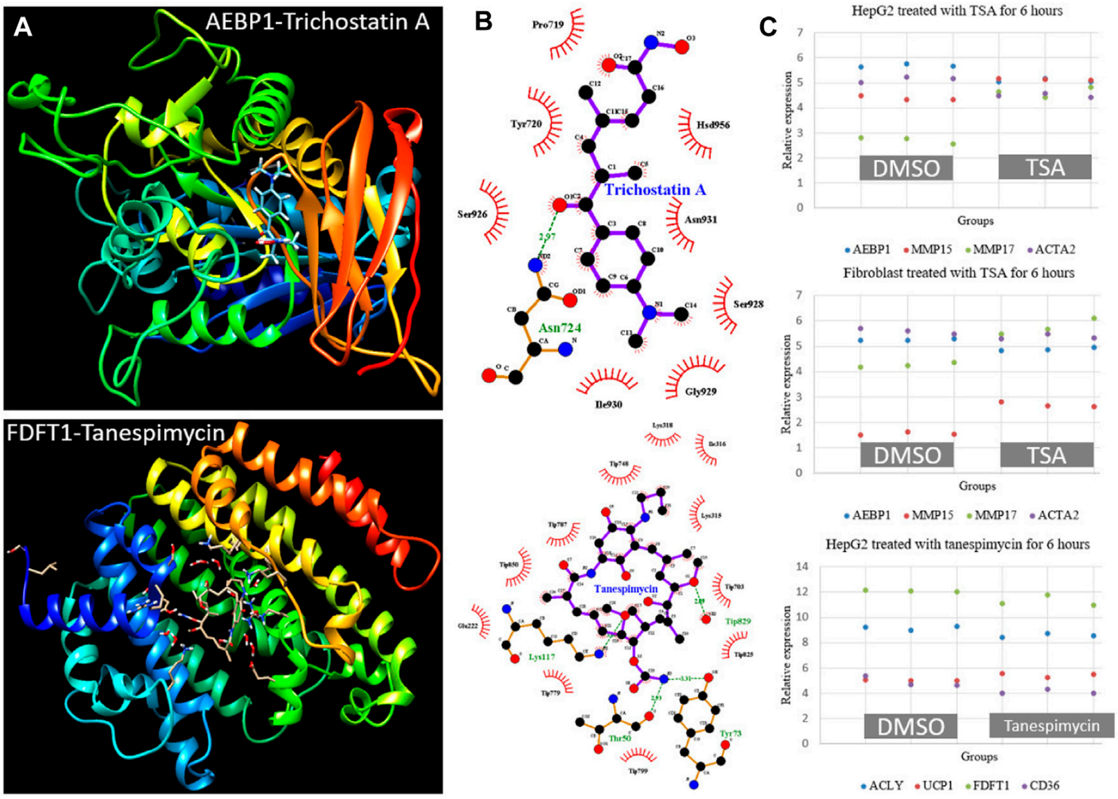


FIGURE 9
Protein-ligand docking analysis and expression validation for hub genes AEBP1 in M25 and FDFT1 in M39. **(A)** 3-D structure models for AEBP1-Trichostatin A and FDFT1-Tanespimycin **(B)** Pose view for the interaction sites. The hydrogen bonds were visualized in the green dashed line **(C)** AEBP1 and fibrosis-related genes MMP15, MMP17, and ACTA2 expression changed after TSA treatment in HepG2 and fibroblast cell lines from dataset GSE200186. FDFT1 and lipid metabolism-related genes ACLY, UCP1, and CD36 expression changed after tanespimycin treatment in HepG2 from dataset GSE119281. All the gene expression was detected with three replicates and had a $p < 0.05$ by Student's t-test.

treatment in HepaG2. Additionally, the expression of genes *Acly* (lipogenic gene), *Ucp1* (thermogenic gene), and *Cd36* (lipid uptake) also changed after 6 h tanespimycin treatment (Figure 9C). All the candidate drugs are worth validation in future studies.

Discussion

We collected a comprehensive compendium of NAFLD transcriptome datasets. Traditional transcriptome data analysis is based on differential analysis methods. However, the number of sample groups in a study may not just be two, if considering different fibrosis stages or other parameters. The subject size varies in studies. Thus, it is somewhat hard to reach a favorable overlap of differential genes between studies. Only four genes, namely, *Col1a2, Efemp2, Fbln5*, and *Thbs2* were found consistently across six published transcriptomic studies on NASH(Pantano et al., 2021). There can be many reasons for the low overlap, such as the inherent complexity of the liver which is composed of heterogeneous cell types, different biopsy sites or methods, different sample preparation workflows, and different transcriptome profiling platforms. We turned to the module-centric analysis, as the module represents a high level of the regulatory scenario (Wang et al., 2008). Several genes may participate in a common biological function. The gene change in one dataset may not replicate in other datasets, but the biological function may replicate in other datasets. We used one dataset for reference module identification, and verify modules in the other 10 datasets. Results suggest that gene co-expression networks from microarray and RNA-Seq are generally reproducible. Based on these modules, we can correlate modules with different clinical data from different studies.

We identified several biological processes that have been known in NAFLD progressions, such as mitochondrion, endoplasmic reticulum, collagen, sterol biosynthesis, and leukocyte activation (Simoes et al., 2018). Some of the identified modules do not have obvious function annotations, such as M35. In our analysis, M35 was downregulated in NAFL compared to normal (GSE89632) and in higher-grade fibrosis compared to control (GSE162694 and GSE130970). Recent reports suggest that activated hsa-miR-21–5p can promote steatosis and fibrosis (Calo et al., 2016; Tadokoro et al., 2021). The Hub gene of M35 *Dennd4c* has not been reported in NAFLD. Thus, it can serve as a potential therapeutic target in future studies. The Hub gene of M38 *Stoml2* is also currently not reported to be associated with NAFLD. In our analysis, M38 was downregulated in NAFL compared with the control. Therefore, it can serve as a potential therapeutic target for steatosis in future studies.

In our module-based analysis, the AUCs for predicting NAFL and advanced fibrosis were 0.93 (M20) and 0.89 (M25), respectively, which were comparable to a recent report (Kozumi et al., 2021). The M20 hub *Clptm1* has a perfect performance (AUC = 0.99) in separating NAFL and normal control. CLPTM1 is located in the endoplasmic reticulum (ER) and is involved in ER stress (Liu et al., 2018b). M20 is the most significant differential module in the spectrum of NAFLD, indicating the important role of ER in disease progression.

One of the important findings in this study is that the gene co-expression modules can efficiently separate NAFLD patients. Although signature genes have been proposed for patient stratification, few genes showed good reproducibility across studies (Vandel et al., 2021). The present study incorporates data from a large number of NAFLD patients and has been validated with multiple databases, showing a satisfactory performance in NAFLD prognosis. In addition, the hub genes of the two modules can be used as biomarkers for NAFLD patient stratification.

We also found that modules were correlated with NAFLD clinical features. As the number of clinical features is variable across datasets, we focused on the common correlations. It may be hard to explain the differences between datasets, as the reason for differences is hard to trace. Correlations detected in more than two datasets may be more reliable module-trait associations. Besides modules M39 and M35, we found M32 was associated with fibrosis and NAS in four datasets (Supplementary Figure S1).

Finally, we proposed drug repurposing by analyzing highly connected genes in the Connectivity Map tool. Some of the identified drugs have been indicated in roles of fibrosis, inflammation, or fat metabolism although not in the liver. For example, TSA is an HDAC inhibitor that could alleviate atrial fibrosis and subsequent atrial fibrillation (Liu et al., 2008). It also can reduce systemic inflammation and improve the survival of the mice model of sepsis (Cui et al., 2019). 17-AAG also known as tanespimycin, a derivative of the antibiotic geldanamycin that has a higher affinity to HSP90, could inhibit fibroblast activation and reduce ECM production (Sontake et al., 2017). Tanespimycin is an Hsp90 inhibitor that can prolong survival, attenuate inflammation, and reduce lung injury in mouse models of sepsis (Chatterjee et al., 2007). Thus, these drugs are promising candidates for the treatment of NASH. Only a recent study showed that antidepressants such as diphenylpyraline could activate FAM3A to suppress hepatic gluconeogenesis and lipogenesis, finally improving hyperglycemia and steatosis in obese diabetic mice (Chen et al., 2020). It was found that inhibition or downregulation of the canonical Wnt/β-catenin pathway contributes to the disease progression of NAFLD (Shree Harini and Ezhilarasan, 2022). While Glycogen synthase kinase 3 (GSK3) inhibitors could activate canonical Wnt/β-catenin signaling and then promote hepatocyte differentiation (Huang et al., 2017). In an animal model, it was proved that GSK3 inhibitor 6-bromoindirubin-3′-oxime (6BIO) could modulate bioenergetic pathways and decrease lipid and glucose tissue load (Tsakiri et al., 2017). Thus, the two drugs are promising candidates for future validation.

However, there are still some limitations to our study. WGCNA generates an undirected network, lacking information on the regulation direction between genes. The collected transcriptome is still not large enough, and some lack detailed clinical information, making the integration of datasets from studies difficult. The identified modules can be used as biomarkers for prognosis, but not sufficiently to reveal of disease mechanism. More work is needed to validate upstream genes that control the co-expression of modules.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: The datasets analyzed in the study are available in public repositories at NCBI GEO (https://www.ncbi.nlm.nih.gov/geo/) and EBI ArrayExpress (https://www.ebi.ac.uk/arrayexpress/).

Author contributions

HY: conceived and designed the experiments. MS: collected data and performed the WGCNA. MS: collected data and curation. DC: performed the WGCNA. HL: visualization. YM: visualization. WL: visualization. HL: analyzed the data. FX: project administration and revised the manuscript.

Funding

This study was funded by the Zhejiang Provincial Natural Science Foundation of China (Nos. LGF19H030006 and LQ20H030001), Ningbo Science and Technology Project (No. 2019C50100), and Ningbo Clinical Medicine Research Center Project (No. 2019A21003). The Natural Science Foundation of Ningbo (202003N4234).

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1070605/full#supplementary-material

References

Ahrens, M., Ammerpohl, O., von Schonfels, W., Kolarova, J., Bens, S., Itzel, T., et al. (2013). DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell metab.* 18, 296–302. doi:10.1016/j.cmet.2013.07.004

Arendt, B. M., Comelli, E. M., Ma, D. W., Lou, W., Teterina, A., Kim, T., et al. (2015). Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. *Hepatology* 61, 1565–1578. doi:10.1002/hep.27695

Azzu, V., Vacca, M., Kamzolas, I., Hall, Z., Leslie, J., Carobbio, S., et al. (2021). Suppression of insulin-induced gene 1 (INSIG1) function promotes hepatic lipid remodelling and restrains NASH progression. *Mol. Metab.* 48, 101210. doi:10.1016/j.molmet.2021.101210

Blackburn, P. R., Xu, Z., Tumelty, K. E., Zhao, R. W., Monis, W. J., Harris, K. G., et al. (2018). Bi-Allelic alterations in AEBP1 lead to defective collagen assembly and connective tissue structure resulting in a variant of ehlers-danlos syndrome. *Am. J. Hum. Genet.* 102, 696–705. doi:10.1016/j.ajhg.2018.02.018

Calo, N., Ramadori, P., Sobolewski, C., Romero, Y., Maeder, C., Fournier, M., et al. (2016). Stress-activated miR-21/miR-21* in hepatocytes promotes lipid and glucose metabolic disorders associated with high-fat diet consumption. *Gut* 65, 1871–1881. doi:10.1136/gutjnl-2015-310822

Chatterjee, A., Dimitropoulou, C., Drakopanayiotakis, F., Antonova, G., Snead, C., Cannon, J., et al. (2007). Heat shock protein 90 inhibitors prolong survival, attenuate inflammation, and reduce lung injury in murine sepsis. *Am. J. Respir. Crit. care Med.* 176, 667–675. doi:10.1164/rccm.200702-291OC

Chen, Z., Liu, X., Luo, Y., Wang, J., Meng, Y., Sun, L., et al. (2020). Repurposing doxepin to ameliorate steatosis and hyperglycemia by activating FAM3A signaling pathway. *Diabetes* 69, 1126–1139. doi:10.2337/db19-1038

Cui, S. N., Chen, Z. Y., Yang, X. B., Chen, L., Yang, Y. Y., Pan, S. W., et al. (2019). Trichostatin A modulates the macrophage phenotype by enhancing autophagy to reduce inflammation during polymicrobial sepsis. *Int. Immunopharmacol.* 77, 105973. doi:10.1016/j.intimp.2019.105973

Gerhard, G. S., Hanson, A., Wilhelmsen, D., Piras, I. S., Still, C. D., Chu, X., et al. (2019). AEBP1 expression increases with severity of fibrosis in NASH and is regulated by glucose, palmitate, and miR-372-3p. *PloS one* 14, e0219764. doi:10.1371/journal.pone.0219764

Govaere, O., Cockell, S., Tiniakos, D., Queen, R., Younes, R., Vacca, M., et al. (2020). Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci. Transl. Med.* 12, eaba4448. doi:10.1126/scitranslmed.aba4448

Grosdidier, A., Zoete, V., and Michielin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic acids Res.* 39, W270–W277. doi:10.1093/nar/gkr366

Hoang, S. A., Oseini, A., Feaver, R. E., Cole, B. K., Asgharpour, A., Vincent, R., et al. (2019). Gene expression predicts histological severity and reveals distinct molecular profiles of nonalcoholic fatty liver disease. *Sci. Rep.* 9, 12541. doi:10.1038/s41598-019-48746-5

Huang, J., Guo, X., Li, W., and Zhang, H. (2017). Activation of Wnt/β-catenin signalling via GSK3 inhibitors direct differentiation of human adipose stem cells into functional hepatocytes. *Sci. Rep.* 7, 40716. doi:10.1038/srep40716

Kozumi, K., Kodama, T., Murai, H., Sakane, S., Govaere, O., Cockell, S., et al. (2021). Transcriptomics identify thrombospondin-2 as a biomarker for NASH and advanced liver fibrosis. *Hepatology* 74, 2452–2466. doi:10.1002/hep.31995

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., et al. (2006). The connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935. doi:10.1126/science.1132939

Langfelder, P., and Horvath, S. (2008). Wgcna: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559

Langfelder, P., Luo, R., Oldham, M. C., and Horvath, S. (2011). Is my network module preserved and reproducible? *PLoS Comput. Biol.* 7, e1001057. doi:10.1371/journal.pcbi.1001057

Laskowski, R. A., and Swindells, M. B. (2011). LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* 51, 2778–2786. doi:10.1021/ci200227u

Lefebvre, P., Lalloyer, F., Bauge, E., Pawlak, M., Gheeraert, C., Dehondt, H., et al. (2017). Interspecies NASH disease activity whole-genome profiling identifies a fibrogenic role of PPARα-regulated dermatopontin. *JCI insight* 2, e92264. doi:10.1172/jci.insight.92264

Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., et al. (2020). TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic acids Res.* 48, W509–W514. doi:10.1093/nar/gkaa407

Liu, F., Levin, M. D., Petrenko, N. B., Lu, M. M., Wang, T., Yuan, L. J., et al. (2008). Histone-deacetylase inhibition reverses atrial arrhythmia inducibility and fibrosis in cardiac hypertrophy independent of angiotensin. *J. Mol. Cell. Cardiol.* 45, 715–723. doi:10.1016/j.yjmcc.2008.08.015

Liu, W., Tu, W., Li, L., Liu, Y., Wang, S., Li, L., et al. (2018). Revisiting Connectivity Map from a gene co-expression network analysis. *Exp. Ther. Med.* 16, 493–500. doi:10.3892/etm.2018.6275

Liu, W., and Wang, Y. (2020). Large-scale analysis of zebrafish (*Danio rerio*) transcriptomes identifies functional modules associated with phenotypes. *Mar. genomics* 53, 100770. doi:10.1016/j.margen.2020.100770

Liu, Y. S., Guo, X. Y., Hirata, T., Rong, Y., Motooka, D., Kitajima, T., et al. (2018). N-Glycan-dependent protein folding and endoplasmic reticulum retention regulate GPI-anchor processing. *J. Cell Biol.* 217, 585–599. doi:10.1083/jcb.201706135

Majdalawieh, A. F., Massri, M., and Ro, H. S. (2020). AEBP1 is a novel oncogene: Mechanisms of action and signaling pathways. *J. Oncol.* 2020, 8097872. doi:10.1155/2020/8097872

Margini, C., and Dufour, J. F. (2016). The story of HCC in NAFLD: From epidemiology, across pathogenesis, to prevention and treatment. *Liver Int. official J. Int. Assoc. Study Liver* 36, 317–324. doi:10.1111/liv.13031

Murphy, S. K., Yang, H., Moylan, C. A., Pang, H., Dellinger, A., Abdelmalek, M. F., et al. (2013). Relationship between methylome and transcriptome in patients with nonalcoholic fatty liver disease. *Gastroenterology* 145, 1076–1087. doi:10.1053/j.gastro.2013.07.047

Paci, P., Fiscon, G., Conte, F., Wang, R.-S., Farina, L., and Loscalzo, J. (2021). Gene co-expression in the interactome: Moving from correlation toward causation via an integrated approach to disease module discovery. *npj Syst. Biol. Appl.* 7, 3. doi:10.1038/s41540-020-00168-0

Pantano, L., Agyapong, G., Shen, Y., Zhuo, Z., Fernandez-Albert, F., Rust, W., et al. (2021). Molecular characterization and cell type composition deconvolution of fibrosis in NAFLD. *Sci. Rep.* 11, 18045. doi:10.1038/s41598-021-96966-5

Pearce, A. C., Senis, Y. A., Billadeau, D. D., Turner, M., Watson, S. P., and Vigorito, E. (2004). Vav1 and vav3 have critical but redundant roles in mediating platelet activation by collagen. *J. Biol. Chem.* 279, 53955–53962. doi:10.1074/jbc.M410355200

Pirola, C. J., Gianotti, T. F., Burgueno, A. L., Rey-Funes, M., Loidl, C. F., Mallardi, P., et al. (2013). Epigenetic modification of liver mitochondrial DNA is associated with histological severity of nonalcoholic fatty liver disease. *Gut* 62, 1356–1363. doi:10.1136/gutjnl-2012-302962

Qin, Y., Li, B., Arumugam, S., Lu, Q., Mankash, S. M., Li, J., et al. (2021). m(6 A mRNA methylation-directed myeloid cell activation controls progression of NAFLD and obesity. *Cell Rep.* 37, 109968. doi:10.1016/j.celrep.2021.109968

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., et al. (2016). g: Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic acids Res.* 44, W83–W89. doi:10.1093/nar/gkw199

Ryaboshapkina, M., and Hammar, M. (2017). Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Sci. Rep.* 7, 12361. doi:10.1038/s41598-017-10930-w

Sharma, M., Mitnala, S., Vishnubhotla, R. K., Mukherjee, R., Reddy, D. N., and Rao, P. N. (2015). The riddle of nonalcoholic fatty liver disease: Progression from nonalcoholic fatty liver to nonalcoholic steatohepatitis. *J. Clin. Exp. hepatology* 5, 147–158. doi:10.1016/j.jceh.2015.02.002

Shree Harini, K., and Ezhilarasan, D. (2022). Wnt/beta-catenin signaling and its modulators in nonalcoholic fatty liver diseases. *Hepatobiliary Pancreat. Dis. Int.* 20, S1499-S3872(22)00239-9. doi:10.1016/j.hbpd.2022.10.003

Simoes, I. C. M., Fontes, A., Pinton, P., Zischka, H., and Wieckowski, M. R. (2018). Mitochondria in non-alcoholic fatty liver disease. *Int. J. Biochem. Cell Biol.* 95, 93–99. doi:10.1016/j.biocel.2017.12.019

Sontake, V., Wang, Y., Kasam, R. K., Sinner, D., Reddy, G. B., Naren, A. P., et al. (2017). Hsp90 regulation of fibroblast activation in pulmonary fibrosis. *JCI insight* 2, e91454. doi:10.1172/jci.insight.91454

Sookoian, S., and Pirola, C. J. (2020). Precision medicine in nonalcoholic fatty liver disease: New therapeutic insights from genetics and systems biology. *Clin. Mol. hepatology* 26, 461–475. doi:10.3350/cmh.2020.0136

Sookoian, S., and Pirola, C. J. (2019). Review article: Shared disease mechanisms between non-alcoholic fatty liver disease and metabolic syndrome - translating knowledge from systems biology to the bedside. *Alimentary Pharmacol. Ther.* 49, 516–527. doi:10.1111/apt.15163

Sookoian, S., Pirola, C. J., Valenti, L., and Davidson, N. O. (2020). Genetic pathways in nonalcoholic fatty liver disease: Insights from systems biology. *Hepatology* 72, 330–346. doi:10.1002/hep.31229

Stattermayer, A. F., Rutter, K., Beinhardt, S., Wrba, F., Scherzer, T. M., Strasser, M., et al. (2014). Role of FDFT1 polymorphism for fibrosis progression in patients with chronic hepatitis C. *Liver Int. official J. Int. Assoc. Study Liver* 34, 388–395. doi:10.1111/liv.12269

Tadokoro, T., Morishita, A., and Masaki, T. (2021). Diagnosis and therapeutic management of liver fibrosis by MicroRNA. *Int. J. Mol. Sci.* 22, 8139. doi:10.3390/ijms22158139

Tsakiri, E. N., Gaboriaud-Kolar, N., Iliaki, K. K., Tchoumtchoua, J., Papanagnou, E. D., Chatzigeorgiou, S., et al. (2017). The indirubin derivative 6-bromoindirubin-3'-oxime activates proteostatic modules, reprograms cellular bioenergetic pathways, and exerts antiaging effects. *Antioxidants redox Signal.* 27, 1027–1047. doi:10.1089/ars.2016.6910

Vandel, J., Dubois-Chevalier, J., Gheeraert, C., Derudas, B., Raverdy, V., Thuillier, D., et al. (2021). Hepatic molecular signatures highlight the sexual dimorphism of nonalcoholic steatohepatitis (NASH). *Hepatology* 73, 920–936. doi:10.1002/hep.31312

Wang, X., Dalkic, E., Wu, M., and Chan, C. (2008). Gene module level analysis: Identification to networks and dynamics. *Curr. Opin. Biotechnol.* 19, 482–491. doi:10.1016/j.copbio.2008.07.011

Ye, H., and Liu, W. (2015). Transcriptional networks implicated in human nonalcoholic fatty liver disease. *Mol. Genet. genomics MGG* 290, 1793–1804. doi:10.1007/s00438-015-1037-3

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4, Article17. doi:10.2202/1544-6115.1128

Zhou, X. H., Chu, X. Y., Xue, G., Xiong, J. H., and Zhang, H. Y. (2019). Identifying cancer prognostic modules by module network analysis. *BMC Bioinforma.* 20, 85. doi:10.1186/s12859-019-2674-z