

RESEARCH ARTICLE

# Identification of hub gene for the pathogenic mechanism and diagnosis of MASLD by enhanced bioinformatics analysis and machine learning

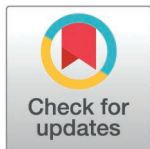
Hong Lu<sup>1✉‡</sup>, Ziyong Mao<sup>2✉‡</sup>, Mengyao Zheng<sup>3,4✉‡</sup>, Min Zhang<sup>2</sup>, Heqing Huang<sup>1</sup>, Yiling Chen<sup>1</sup>, Long Lv<sup>5\*</sup>, Zutao Chen<sup>1,6,7\*</sup>

**1** Infectious Disease Department, the First Affiliated Hospital of Soochow University, Suzhou, Jiangsu Province, China, **2** BamRock Research Department, Suzhou BamRock Biotechnology Ltd., Suzhou, Jiangsu Province, China, **3** College of Life Sciences, Qufu Normal University, Qufu, Shandong Province, China, **4** Department of Biological Sciences, University at Albany, Albany, New York, United States of America, **5** Central Laboratory, The Affiliated Gaochun Hospital of Jiangsu University, Nanjing, Jiangsu Province, China, **6** MOE Key Laboratory of Geriatric Diseases and Immunology, Suzhou Key Laboratory of Pathogen Bioscience and Anti-infective Medicine, the First Affiliated Hospital of Soochow University, Suzhou, Jiangsu Province, China, **7** Infectious Disease Department, the Fourth Affiliated Hospital of Soochow University, Suzhou, Jiangsu Province, China

✉ These authors contributed equally to this work.

‡ co-first authors.

\* [chenzt@suda.edu.cn](mailto:chenzt@suda.edu.cn) (ZC); [13851596501@163.com](mailto:13851596501@163.com) (LL)



## OPEN ACCESS

**Citation:** Lu H, Mao Z, Zheng M, Zhang M, Huang H, Chen Y, et al. (2025) Identification of hub gene for the pathogenic mechanism and diagnosis of MASLD by enhanced bioinformatics analysis and machine learning. PLoS One 20(5): e0324972. <https://doi.org/10.1371/journal.pone.0324972>

**Editor:** Akif Altinbas, University of Connecticut, UNITED STATES OF AMERICA

**Received:** February 19, 2025

**Accepted:** May 5, 2025

**Published:** May 28, 2025

**Copyright:** © 2025 Lu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** Gene expression data and clinical information used to support the findings of this study are downloaded from the laboratory data porta of professor Huang, including GepLiver-bulk-05 (GSE126848, 31 NAFLD), GepLiver-bulk-06 (GSE130970,

## Abstract

Metabolic dysfunction-associated steatotic liver disease (MASLD) is a heterogeneous disease caused by multiple etiologies. It is characterized by excessive fat accumulation in the liver. Without intervention, MASLD can progress from steatosis to metabolic dysfunction-associated steatohepatitis (MASH), fibrosis and even to cirrhosis and hepatocellular carcinoma. However, the pathogenesis of MASH and the mechanism underlying the development of fibrosis remain poorly understood, posing challenges for accurate diagnosis of MASH and fibrosis. In this study, we analyzed tissue RNA-seq data and clinical information of healthy individuals and MASLD patients from multiple datasets, the key genes and pathways involved in the occurrence and progression of MASLD, MASH, and fibrosis were screened respectively. Our findings reveal that the development of MASLD, MASH and fibrosis is associated with lipid metabolism processes. Based on the RNA expression profiles of identified hub genes, we established three alternative diagnostic models for MASLD, MASH, and fibrosis. These models demonstrated excellent performance in the diagnosis of MASLD, MASH, and fibrosis, with AUC values exceeding 0.9, implicating its potential clinical values in disease diagnosis.

## Introduction

MASLD is a complex process and closely linked to cardiometabolic disease, whilst excluding alcohol, genetic factors and other causes as contributors to liver disease

72 NAFLD), GepLiver-bulk-07 (GSE135251, 192 NAFLD), GepLiver-bulk-08 (GSE162694, 31 Normal, 81 NAFLD), GepLiver-bulk-09 (GSE167523, 98 NAFLD), GepLiver-bulk-13 (E-MTAB-6863, 10 NAFLD) and GepLiver-bulk-01 (GTEx database, 226 Normal) (<http://www.gepliver.org/#/download>). The source code of random forest algorithm has been uploaded and is available from GitHub ([https://github.com/Bamrock/Machine\\_learning/](https://github.com/Bamrock/Machine_learning/)).

**Funding:** Funding: This work was supported by the Suzhou Science and Technology Bureau (SKY2022047) and high quality project of gaochun people's hospital (GYK-2021-004).

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** MAFLD, Metabolic dysfunction-associated fatty liver disease; MASLD, Metabolic dysfunction-associated steatosis liver disease; MASH, Metabolic dysfunction-associated Steatohepatitis; NAS, NAFLD Activity Score; GEO, Gene Expression Omnibus; PPI, Protein-protein interaction; GO, Gene Ontology; MF, Molecular Functions; CC, Cellular Components; BP, Biological Processes; KEGG, Kyoto Encyclopedia of Genes and Genomes; ROC, Receiver operating characteristic

[1–3]. As a result of the diabetes and metabolic syndrome epidemics, MASLD has risen to the top of the list of chronic liver diseases worldwide. The spectrum of MASLD ranges from initial metabolic dysfunction-associated fatty liver to MASH, and advanced conditions such as MASH-associated cirrhosis and hepatocellular carcinoma [4]. However, the disease progression is poorly understood.

This research builds on previous research data on metabolic liver disease, where it was previously known as non-alcoholic fatty liver disease (NAFLD) [5]. It is advocated that the severity evaluation of MASLD is based on inflammation activity rather than the presence or absence of inflammation. The NAFLD Activity Score (NAS) system remains in use, evaluating histologic features into three distinct categories: lobular inflammation (0–3), steatosis (0–3), and ballooning (0–2) [6]. MASH is diagnosed with a NAS value of 5 or greater, a NAS value lower than 3 cannot be diagnosed with MASH. Currently, studies have found that an increase in NAS score is associated with the progression of liver fibrosis in the presence of steatohepatitis, while a decrease in NAS score is associated with resolution of fibrosis. Chronic inflammation activates the immune system and releases a large amount of inflammatory mediators (such as cytokines and chemokines). These substances stimulate the activation of hepatic stellate cells (HSCs), which in turn secrete excessive extracellular matrix (ECM), forming fibrous scars [7–9]. In mouse models of fatty liver disease, the dysregulation of some key genes has been shown to modulate hepatic inflammation and lipid accumulation, thereby contributing to the development of MASLD and fibrosis [31,32]. However, there is inherent difference between murine and human pathophysiology. The hub genes and pathways that involved in the development and progression of human MASLD, MASH and fibrosis remain unclear.

The histology of liver biopsy remains the gold standard for diagnosis. However, biopsy driven diagnosis has its limitations, namely bias and variable interpretation. Histological analysis of biopsy specimens requires experience and skill, but there is still a subjective bias that tends to vary within and between observers. For example, the agreement of liver pathologists on the diagnosis of fibrosis staging within observers is 60%~90%, and the consensus among observers is 70%~90% [10].

In this study, we employed an integrative analysis of multiple RNA-seq datasets to identify key genes and pathways implicated in the onset and progression of MASLD, MASH, and fibrosis. Through bioinformatic modeling and functional annotation, we investigated the biological roles and mechanistic contributions of these genes. Next, we explore whether there is a core set of genes and pathways that play a key role in the above process. Finally, alternative diagnostic models for MASLD, MASH and fibrosis were established through RNA expression profiles of liver tissue samples using a small set of hub genes, respectively. The implementation of these models may facilitate the standardization, objectification, and automation of diagnostic process. The results of this study are expected to provide valuable references for further research on MASLD disease mechanisms, diagnosis, and discovery of drug therapeutic targets.

## Methods

### Data collection and clinical information analysis

Gene expression data and clinical information of 346 MASLD and 226 normal liver tissues were acquired for model training from the laboratory data porta of professor Huang (<http://www.gepliver.org/#/download>) [11], including GepLiver-bulk-06 (GSE130970\_MASLD), GepLiver-bulk-07 (GSE135251\_MASLD) and GepLiver-bulk-08 (GSE162694\_MASLD), and GepLiver-bulk-01 (GTEx database\_Normal). Several independent MASLD datasets (GepLiver-bulk-05\_GSE126848, GepLiver-bulk-09\_GSE167523, GepLiver-bulk-13\_E-MTAB-6863) are used to verify our model. The detailed source and population demographics (sex and age) are described (see [S1 File](#), Supplemental Digital Content, showed source and clinical information). The Pearson correlation coefficient was calculated to check the correlation between population demographics factors (sex and age) and MASH or fibrosis by using SPSS software ([S6 File](#)). The specific processing flow is described in detail in “GepLiver: a dynamic, integrative liver expression atlas spanning developmental stages and liver disease phases, figshare”. Our study used publicly available data, therefore, ethical IRB approval is not required.

### Differential expression analyses

In order to explore the differential expression genes (DEGs) that related to MASLD, MASH and fibrosis respectively, we divided the data into two distinct groups according to the following three distinct methods:

- (1) To explore the DEGs associated with MASLD, the MASLD patients composed the disease group and the healthy individuals composed the control group.
- (2) In order to investigate the DEGs associated with MASH, MASLD patients with  $NAS \geq 5$  were placed in disease group and MASLD patients with  $NAS < 3$  composed the control group.
- (3) To examine the DEGs associated with fibrosis, MASLD patients with a fibrosis stage  $> 0$  were placed in disease group and MASLD without fibrosis (stage = 0) composed the control group. To reduce variability in fibrosis assessment, we investigated the inflammation difference between the fibrosis and control groups, and we confirmed that there is no inflammation-related bias.

Because our aim is to investigate the genes involved in fibrosis within the context of liver inflammation. Therefore, we excluded patients who lacked liver inflammation. Given that the majority of patients labeled as MAFLD in the GSE135251 database did not exhibit liver inflammation, we excluded these MAFLD patients from our analysis. Similarly, patients with an inflammation grade of 0 in the GSE130970 database were also excluded. It is noteworthy that the GSE162694 database was not considered for the fibrosis group analysis due to the absence of specific inflammatory information for individual patients.

The RNA-Seq data of liver tissues were analyzed using the Limma package of R language (<http://www.bioconductor.org/packages/release/bioc/html/limma.html>) to identify DEGs. The data from Gepliver database are preprocessed and there is no missing data. There were no outliers. The low-expression genes (expression level  $< 1$ ) were filtered out. The voom method was used to analyze DEGs: (1) the raw counts were converted to log<sub>2</sub> CPM (counts per million reads), all counts were added by 0.5 to avoid taking logarithmic zeros; (2) the logCPM value matrix was normalized, the Non-Linear Least-Square Minimization and Compute Contrasts from Linear Model Fit was used to fit and compare the data; (3) the compared model was transformed into an empirical Bayesian model by eBayes, and the variance part of the t-test was adjusted. Finally, logFC, average expression, p value, t, q, value and B value were extracted. DEGs were selected based on the following criteria:  $q\text{-value} < 0.05$  and  $|\log_2 \text{Fold Change}| (|\log_2 \text{FC}|) > 1$ . Heatmaps were generated using pheatmap package and volcano plots were also conducted in R software.

## Correlation analysis between gene expression and clinical phenotype

To examine the correlation between continuous variables (gene expression) and discrete variables (e.g., NAS or fibrosis stage), we used the ordinal logistic regression. The “polr” function in the R package MASS was used to conduct this analysis. The pseudo R square was calculated using the likelihood ratio method:

$$R^2 = \frac{1 - (\log\text{Lik}(\text{model}))}{(\log\text{Lik}(\text{nullmodel}))}$$

In which the “logLik” is the log likelihood of a model, and the pseudo R square reflects the magnitude of correlation between variables.

The FDR method was used to get the q-value. We used the threshold q-value < 0.01 and R square > 0 to screen the significant correlation genes.

## Weighted gene co-expression network analysis

To identify gene co-expression modules associated with MASLD, we performed a weighted gene co-expression network analysis (WGCNA) on RNA seq data of MASLD using the R package “WGCNA” (<https://CRAN.R-project.org/package=WGCNA>). In this study, we chose the soft threshold  $\beta = 12$  (scale free  $R^2 = 0.98$ ). The genes within modules were selected as the candidate hub genes.

## Go and KEGG enrichment analysis

To investigate the biological pathways that might be involved in the occurrence and development of MASLD, MASH and fibrosis, DEGs were subjected to Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichments using the R package “clusterProfiler”. The GO analysis involved three categories, namely molecular functions (MF), cellular components (CC), and biological processes (BP). The threshold was set as q-value < 0.05.

## Protein-protein interaction networks

The protein-protein interaction (PPI) network for DEGs was constructed using STRING, BioGrid, OmniPath and InWeb\_IM [12–15]. Only STRING (physical score > 0.132) and physical interactions in BioGrid were used. The MCODE plugin was used to find important modules enriched in the PPI network. Genes in crucial (top score) modules were selected as candidate genes.

## Machine learning-based model training, testing and independent validation

To analyze the feasibility of candidate genes in identifying MASLD, MASH and fibrosis, receiver operating characteristic (ROC) curve prediction was performed according to clinical data (S6 File).

We used bootstrap method for estimating the confidence intervals [16,17]. Briefly, samples were repeatedly extracted 100 times (randomly sampling) from the original sample data. 80% cases were extracted and used for model construction each time. The 95% confidence intervals of AUC, sensitivity and specificity were estimated using 100 times repeated sampling. Hanley Mcneil method was used to test the significant differences between AUC values [18]. Because the negative and positive samples are unbalance in the Fibrosis dataset, over-sampling (also called up-sampling) methods are used to increase the number of negative samples in a minority class. The hyperparameters used for the random forest algorithm were selected using grid search each time, the hyperparameters give the best performance in test dataset are chosen to construct models. The optimized hyperparameters in our study including nestimators and max\_features. Other hyperparameters are default in the software. The hyperparameters for the best MASLD diagnosis model are 200 and 20 respectively using grid search.

Several independent MASLD datasets (GepLiver-bulk-05\_GSE126848, GepLiver-bulk-09\_GSE167523, GepLiver-bulk-13\_E-MTAB-6863) are used to verify our MASLD model. There is no information of NAS score or fibrosis stages in these independent datasets, and we cannot find other independent MASH and Fibrosis dataset that containing information of NAS score or fibrosis stages, thus MASH and Fibrosis models are not verified using independent dataset in our study.

## Results

### Correlation of sex and age with MASLD, MASH, fibrosis

The Pearson correlation coefficients of MASLD, MASH, fibrosis with sex were 0.356, 0.236, 0.285 respectively, and the Pearson correlation coefficients of MASLD, MASH, fibrosis with age were 0.247, 0.312, 0.268 respectively (Table 1). The results indicated that age and sex had low correlation with MASLD, MASH and fibrosis, which means these factors will not significantly affect our analysis.

### Genes, modules, and pathways associated with MASLD

Total 1,068 DEGs (279 upregulated and 789 downregulated) were screened between MASLD and normal liver tissues. The heatmap of DEGs were presented in Fig 1A, which revealed that MASLD samples can be distinguished from the normal samples. Volcano plots and venn plot shows the distribution of DEGs between MASLD and normal controls (Fig 1B and 1C). These identified DEGs are expected to have possible roles in the development and progression of MASLD.

In WGCNA, a total of 5 co-expression modules were identified in MASLD. The turquoise and yellow modules comprised 1237 positive correlation genes, and the blue modules comprised 421 negatively correlated genes (Fig 1D). Among these genes, a total of 448 genes overlapped with DEGs.

Subsequently, these 448 genes were analyzed by GO and KEGG enrichments to clarify the biological functions. The top 10 enriched GO and KEGG pathways of DEGs in MASLD are listed in Table 2. In GO analysis, most of the top 10 enriched pathways are associated with catabolism and metabolism of organic compounds (Table 2, Fig 1F). In KEGG analysis, most OF TOP 10 enriched pathways are associated with substance (e.g., xenobiotics, drug, carbon) metabolism (Table 2, Fig 1E). The complete result of KEGG, GO (BP, CC, MF) analysis of DEGs for MASLD were summarized in S2 File of Supplemental Digital Content.

The PPI network of these 448 genes containing 161 edges was constructed by STRING, which provides critical assessment and integration of protein-protein interactions, including direct (physical) and indirect (functional) correlations (Fig 1G). In the PPI network, the MCODE plugin identified 12 significant enriched modules. 25 genes (*UGT2B7*, *UGT1A9*, *UGT1A4SAA1*, *PON1*, *PGM1*, *PC*, *MDH2*, *LSR*, *LIPC*, *IDH2*, *HSD17B14*, *HMGCS2*, *GOT2*, *GOT1*, *GLUD1*, *GDI2*, *FDPS*, *EGFR*, *DBI*, *APOM*, *APOH*, *AKR1C2*, *AKR1C1*, *ACO2*) in the first score of clusters were selected as the candidate hub genes associated with MASLD. These candidate hub genes were enriched in the pathway of glycolipid metabolism. The AUC values of the 25 candidate genes were showed in Fig 4A, most of which were greater than 0.8.

### Genes, modules, and pathways associated with MASH

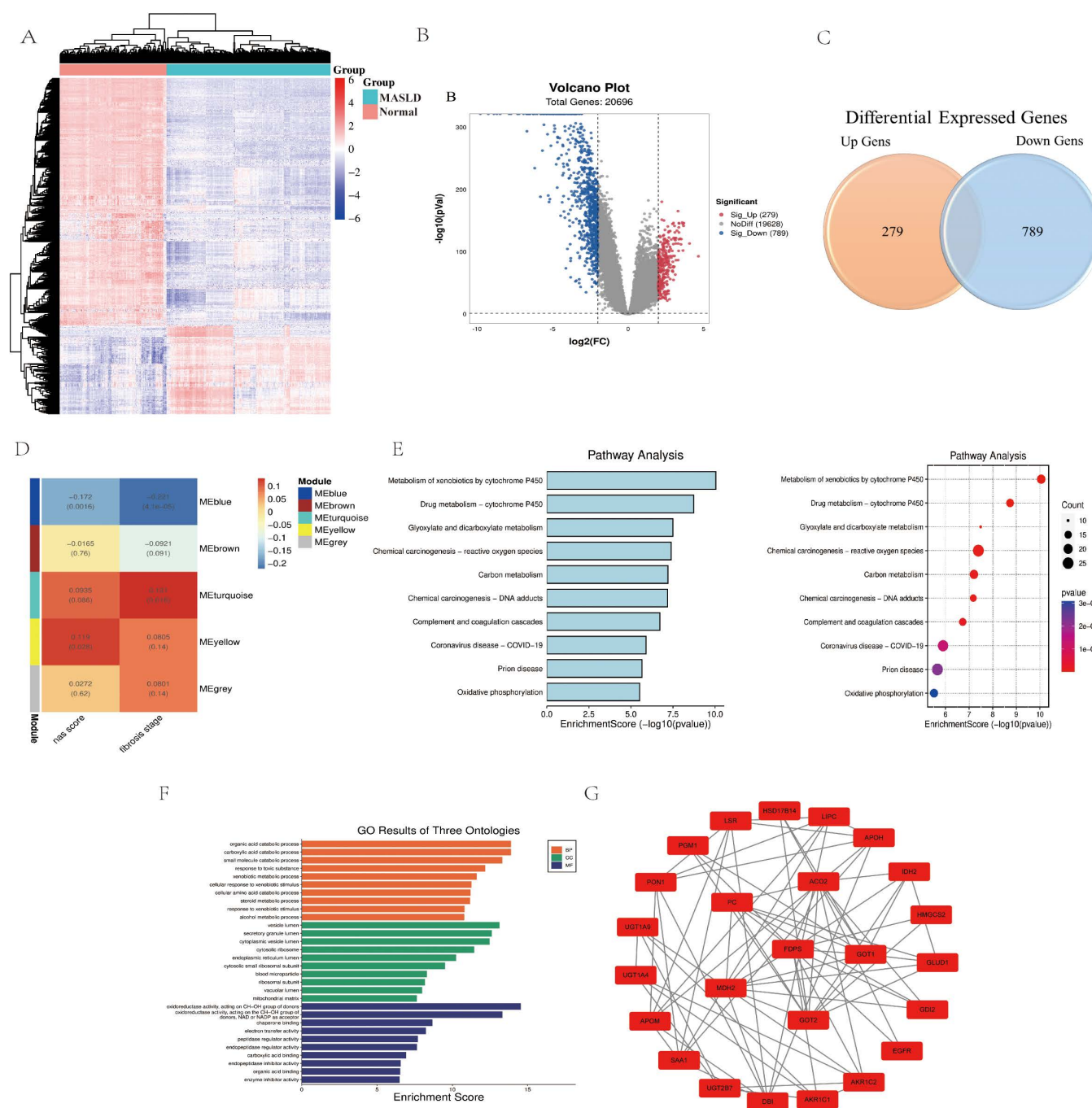
A total of 108 DEGs (54 upregulated and 54 downregulated) were identified between MASLD samples with  $NAS \geq 5$  and  $NAS < 3$ . The heatmap (Fig 2A) revealed that MASH ( $NAS \geq 5$ ) samples can be obviously distinguished from the MAFLD

**Table 1. The Pearson correlation coefficients of MASLD, MASH, fibrosis with sex and age.**

Phenotype	Sex	Age
MASLD	0.365	0.247
MASH	0.236	0.312
Fibrosis	0.285	0.268

<https://doi.org/10.1371/journal.pone.0324972.t001>





**Fig 1. Genes, modules, and pathways associated with MASLD.** (A) Heatmap of significantly DEGs between normal and MASLD liver tissues. The color from blue to red represents the progression from low expression to high expression. (B) Volcano plot of differentially expressed genes between Normal and MASLD. The red dots in the plot represents upregulated genes and blue dots represents downregulated genes with statistical significance. Gray dots represent there is no differential expression in those genes. (C) Venn diagram of differentially expressed genes between Normal and MASLD. (D) WGCNA module plot of genes in MASLD. Each row represents a co-expression module. Each column represents a trait attribute. Blue color represents negative correlation and red color represents positive correlation. (E) KEGG pathways of DEGs. (F) Go analysis. The figure represents BP, CC and MF of DEGs. (G) The most significant enriched module in the PPI network.

<https://doi.org/10.1371/journal.pone.0324972.g001>

**Table 2. Functional Enrichment Analysis of GO and KEGG Pathways of DEGs in the MASLD, MASH and Fibrosis.**

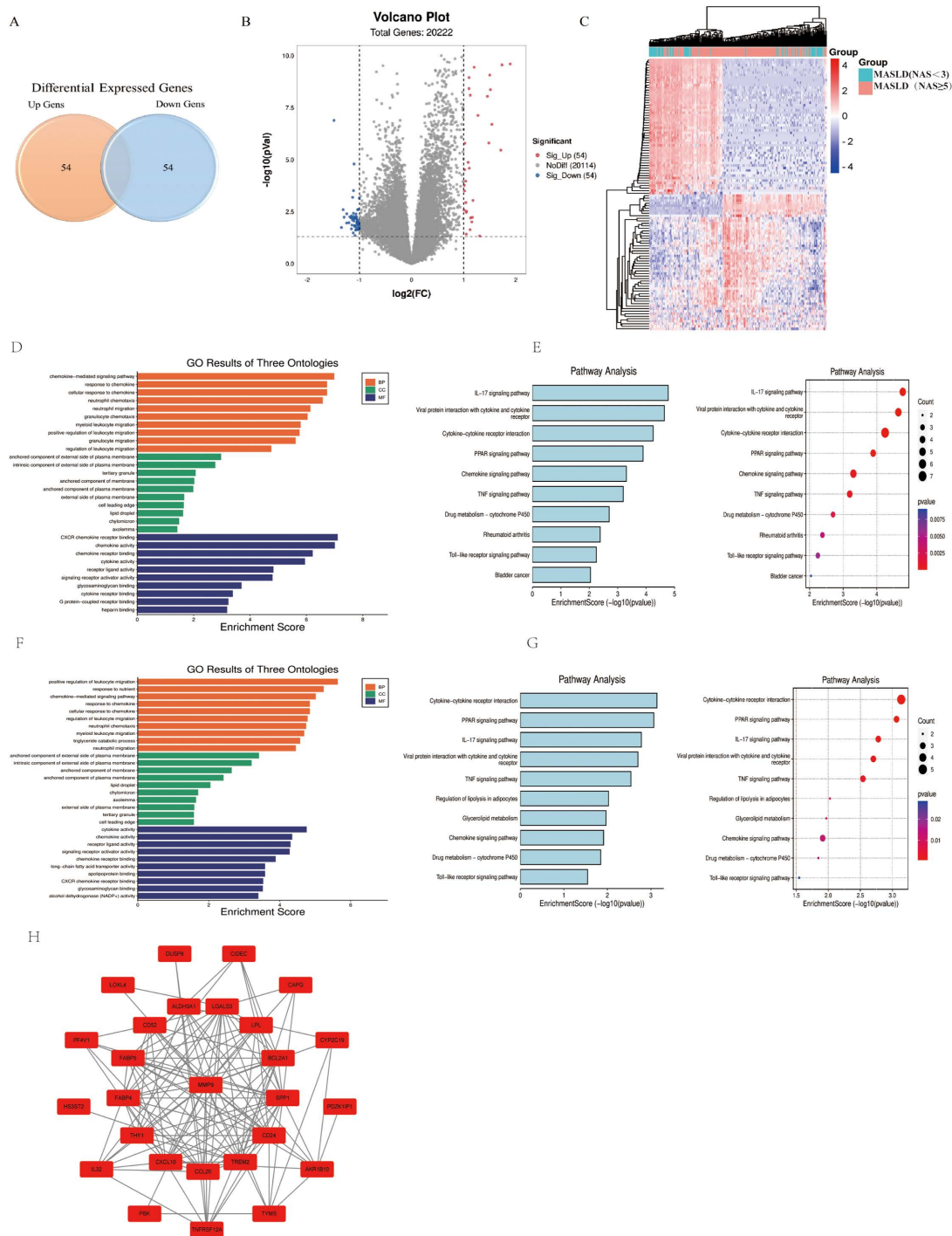
Phenotype	GO			KEGG
	Biological Process	Cell component	Molecular Function	
MASLD	Organic acid catabolic process, carboxylic acid catabolic process, small molecule catabolic process, response to toxic substance, xenobiotic metabolic process, cellular response to xenobiotic stimulus, cellular amino acid catabolic process, steroid metabolic process, response to xenobiotic stimulus, alcohol metabolic process, etc.	vesicle lumen, secretory granule lumen, cytoplasmic vesicle lumen, cytosolic ribosome, endoplasmic reticulum lumen, cytosolic small ribosomal subunit, blood microparticle, ribosomal subunit, vacuolar lumen, mitochondrial matrix, etc	oxidoreductase activity, acting on CH-OH group of donors, oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor, chaperone binding, electron transfer activity, peptidase regulator activity, endopeptidase regulator activity, carboxylic acid binding, endopeptidase inhibitor activity, organic acid binding, enzyme inhibitor activity, etc.	Metabolism of xenobiotics by cytochrome P450, Drug metabolism – cytochrome P450, Glyoxylate and dicarboxylate metabolism, Chemical carcinogenesis – reactive oxygen species, Carbon metabolism, Chemical carcinogenesis DNA adducts, Complement and coagulation cascades, Coronavirus disease - COVID-19, Prion disease, Oxidative phosphorylation, etc.
MASH	positive regulation of leukocyte migration, response to nutrient, chemokine-mediated signaling pathway, response to chemokine, cellular response to chemokine, regulation of leukocyte migration, neutrophil chemotaxis, myeloid leukocyte migration, triglyceride catabolic process, neutrophil migration, etc.	anchored component of external side of plasma membrane, intrinsic component of external side of plasma membrane, anchored component of membrane, anchored component of plasma membrane, lipid droplet, chylomicron, axolemma, external side of plasma membrane, tertiary granule, cell leading edge, etc.	cytokine activity, chemokine activity, receptor ligand activity, signaling receptor activator activity, chemokine receptor binding, long-chain fatty acid transporter activity, apolipoprotein binding, CXCR chemokine receptor binding, glycosaminoglycan binding, alcohol dehydrogenase (NADP+) activity, etc.	Cytokine-cytokine receptor interaction, PPAR signaling pathway, IL-17 signaling pathway, Viral protein interaction with cytokine and cytokine receptor, TNF signaling pathway, Regulation of lipolysis in adipocytes, Glycerolipid metabolism, Chemokine signaling pathway, Drug metabolism – cytochrome P450, Toll-like receptor signaling pathway, etc.
Fibrosis	detection of chemical stimulus involved in sensory perception of taste, detection of chemical stimulus involved in sensory perception of bitter taste, sensory perception of bitter taste, protein K48-linked deubiquitination, covalent chromatin modification, histone modification, regulation of small GTPase mediated signal transduction, peptidyl-lysine modification, positive regulation of GTPase activity, regulation of chromosome organization, etc.	cell leading edge, centriole, site of polarized growth, site of DNA damage, growth cone, PML body, site of double-strand break, chromosomal region, cytoplasmic ribonucleoprotein granule, pericentriolar material, etc.	bitter taste receptor activity, taste receptor activity, ATPase activity, guanyl-nucleotide exchange factor activity, nucleoside-triphosphatase regulator activity, GTPase activator activity, GTPase regulator activity, DNA-dependent ATPase activity, protein serine/threonine kinase activity, small GTPase binding, etc.	Taste transduction, ABC transporters, Signaling pathways regulating pluripotency of stem cells, Phosphatidylinositol signaling system, Lysine degradation, Ubiquitin mediated proteolysis, Thyroid hormone signaling pathway, Hepatocellular carcinoma, Adherens junction, Colorectal cancer, etc.

<https://doi.org/10.1371/journal.pone.0324972.t002>

(NAS < 3) samples based on DEGs. Volcano plots and venn plots shows the distribution of DEGs (Fig 2B and 2C). These identified DEGs are expected to have possible roles in the development and progression of MASH.

GO analysis of DEGs in MASH revealed that most of top 10 enriched pathways are linked to chemokine-mediated signaling and immune cells (e.g., leukocyte, neutrophil, granulocyte) migration (Fig 2D). KEGG analysis of DEGs in MASH revealed that most of top 10 enriched pathways are linked to immune-inflammatory regulation (e.g., cytokine, chemokine, IL-17, Toll-like receptor) signaling pathways (Fig 2E). Complete results of the KEGG and GO analyses (BP, CC, MF) for MASH-associated DEGs are provided in S3 File (Supplemental Digital Content).

A total of 2,528 genes were identified significantly correlated with NAS using ordinal logistic regression and 38 genes were overlapped with DEGs. GO analysis of the 38 candidate genes revealed that most of top10 enriched pathways are linked to chemokine-mediated signaling and immune cells (e.g., leukocyte, neutrophil,) migration (Table 2, Fig 2F). KEGG analysis of 38 candidate genes in MASH revealed that most of top 10 enriched pathways are linked to



**Fig 2. Genes, modules, and pathways associated with MASH.** (A) Venn diagram of DEGs with MASH. (B) Volcano plot of DEGs with MASH. The red dots in the plot represents upregulated genes and blue dots represents downregulated genes with statistical significance. Gray dots represent there is no differential expression in those genes. (C) Heatmap of significantly DEGs with MASH. The color from blue to red represents the progression from low expression to high expression. (D) Go analysis. The figure represents BP, CC and MF of the DEGs. (E) The most significant KEGG pathways of the DEGs. (F) Go analysis. The figure represents BP, CC and MF of the DEGs overlapped with those screened by correlation analysis. (G) The most significant KEGG pathways of DEGs overlapped with those screened by correlation analysis. (H) The most significant enriched module in the PPI network.

<https://doi.org/10.1371/journal.pone.0324972.g002>



immune-inflammatory regulation (e.g., cytokine, chemokine, TNF, Toll-like receptor) signaling pathways (Table 2, Fig 2G). The complete result of KEGG, GO (BP, CC, MF) analysis of 38 candidate genes for MASH were summarized in S4 File (Supplemental Digital Content).

The PPI network of these 38 genes containing 125 edges was constructed (Fig 2H). The 12 genes with a degree score greater than 10 were identified as candidate hub genes for MASH, including five key genes related to inflammation (*MMP9*, *CXCL10*, *CCL20*, *BCL2A1*, *THY1*), and the remaining eight key genes were related to lipid metabolism (*TREM2*, *LPL*, *CD52*, *SPP1*, *LGALS3*, *CD24*, *FABP4*, *FABP5*). The AUC values of the 12 candidate genes for MASH detection were showed in the range of 0.6–0.8 (Fig 4B).

### Genes, modules, and pathways associated with the fibrosis in MASLD

A total of 1,814 DEGs (54 upregulated and 1,760 downregulated) were identified between MASLD samples with fibrosis stage=0 and fibrosis stage>0. The heatmap (Fig 3A) revealed that samples with fibrosis can be obviously distinguished from the samples without fibrosis based on DEGs. Volcano plots and Venn plot shows the distribution of DEGs (Fig 3B and 3C). The identified DEGs are expected to have possible roles in the development and progression of fibrosis.

After the ordinal logistic regression analysis, a total of 8,257 genes that significantly correlated with the fibrosis stage were screened, in which 1,583 genes were overlapped with DEGs. GO analysis of the 1,583 candidate genes revealed that most of top10 pathways are associated with protein deubiquitination, chromatin and histone modification, chemical stimulus etc (Table 2, Fig 3E). KEGG analysis revealed that most of top10 pathways are linked to energy metabolism (ATP/GTP regulation) and cellular responses (signaling, structural remodeling) (Table 2, Fig 3D). The complete result of KEGG, GO (BP, CC, MF) analysis of DEGs for fibrosis of MASLD were summarized in S5 File of Supplemental Digital Content. The enriched pathways were discovered to be predominantly associated with ubiquitin and lipid metabolism.

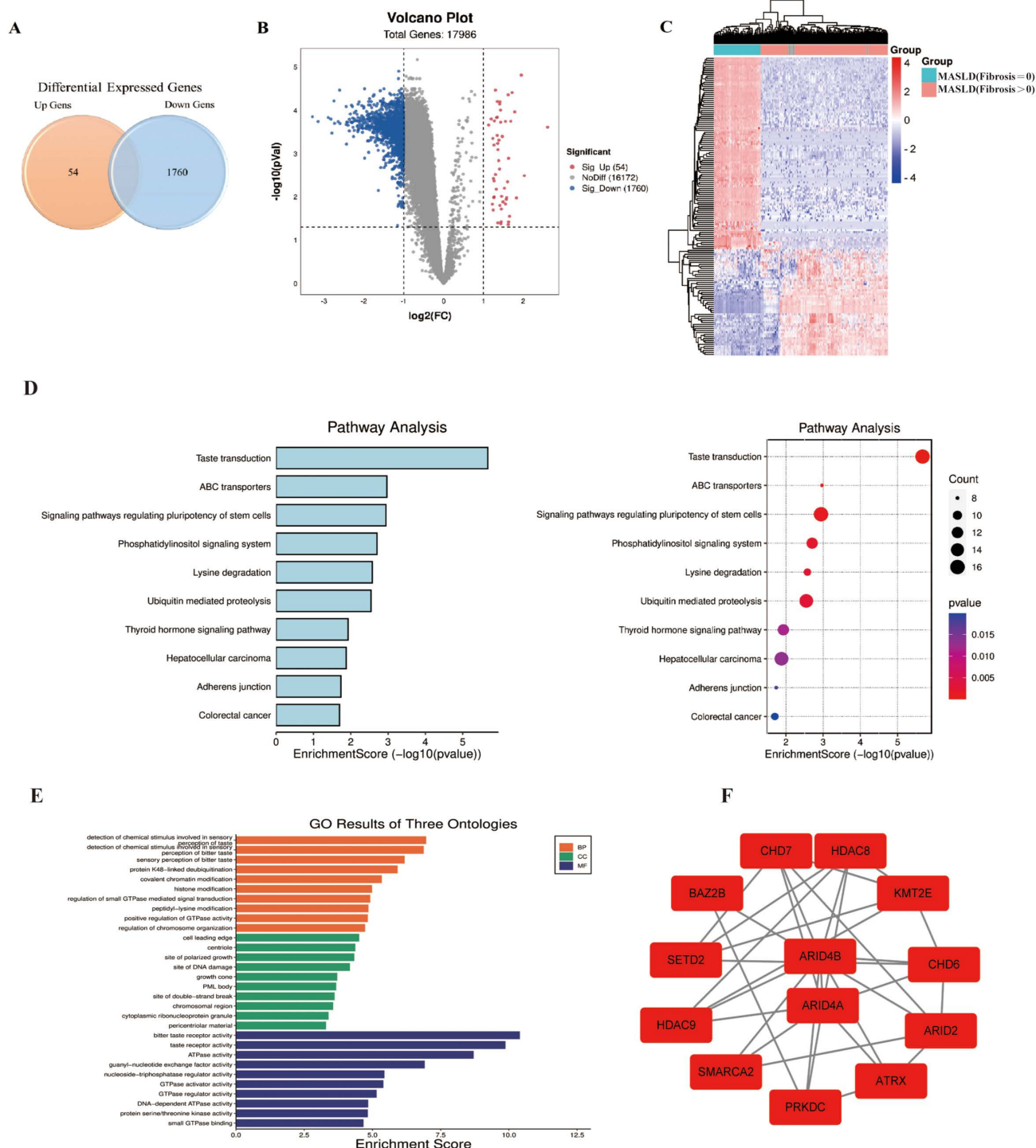
Protein-protein interaction (PPI) enrichment analysis was performed on these 1,583 genes. Six important modules were enriched in the PPI network. Thirteen genes (*BAZ2B*, *ARID2*, *ARID4B*, *CHD6*, *HDAC8*, *SMARCA2*, *HDAC9*, *SETD2*, *PRKDC*, *ATRX*, *ARID4A*, *CHD7*, *KMT2E*) in the cluster with highest score were selected as the candidate hub genes associated with fibrosis in MASLD (Fig 3F). The AUC values of the 13 candidate genes were showed in Fig 4C.

### Diagnostic models for MASLD, MASH and fibrosis

We constructed three alternative diagnostic models for MASLD, MASH, and fibrosis using 25, 13, and 13 candidate genes, respectively. Using bootstrap methods, the average AUC, sensitivity, specificity for distinguishing MASLD from normal individuals was 0.999 (95%CI: 0.990–1.00), 0.987 (95%CI: 0.951–1.000) and 0.991 (95%CI: 0.960–1.000) respectively. In MASH detection, the average AUC, sensitivity and specificity were 0.917 (95%CI: 0.882–0.952), 0.807 (95%CI: 0.651–0.963) and 0.873 (95%CI: 0.744–1) respectively. In the fibrosis detection, the average AUC, sensitivity and specificity were 0.986 (95%CI: 0.953–1.000), 0.908 (95%CI: 0.897–1.000) and 0.986 (95%CI: 0.925–1.000) respectively (Fig 4D). The AUC difference between MASLD and MASH is significant ( $p=0.023$ ). No significant difference of AUC is found between MASLD and fibrosis ( $p=0.32$ ) or between MASH and fibrosis ( $p=0.071$ ). To assess the generalizability and potential clinical utility of MASLD, the best MASLD model was constructed with grid search method and independent datasets were used to validate the model. The AUC for distinguishing MASLD from normal individuals in train, test and independent validation dataset were 1.000, 0.999 and 0.957 respectively (Fig 4E).

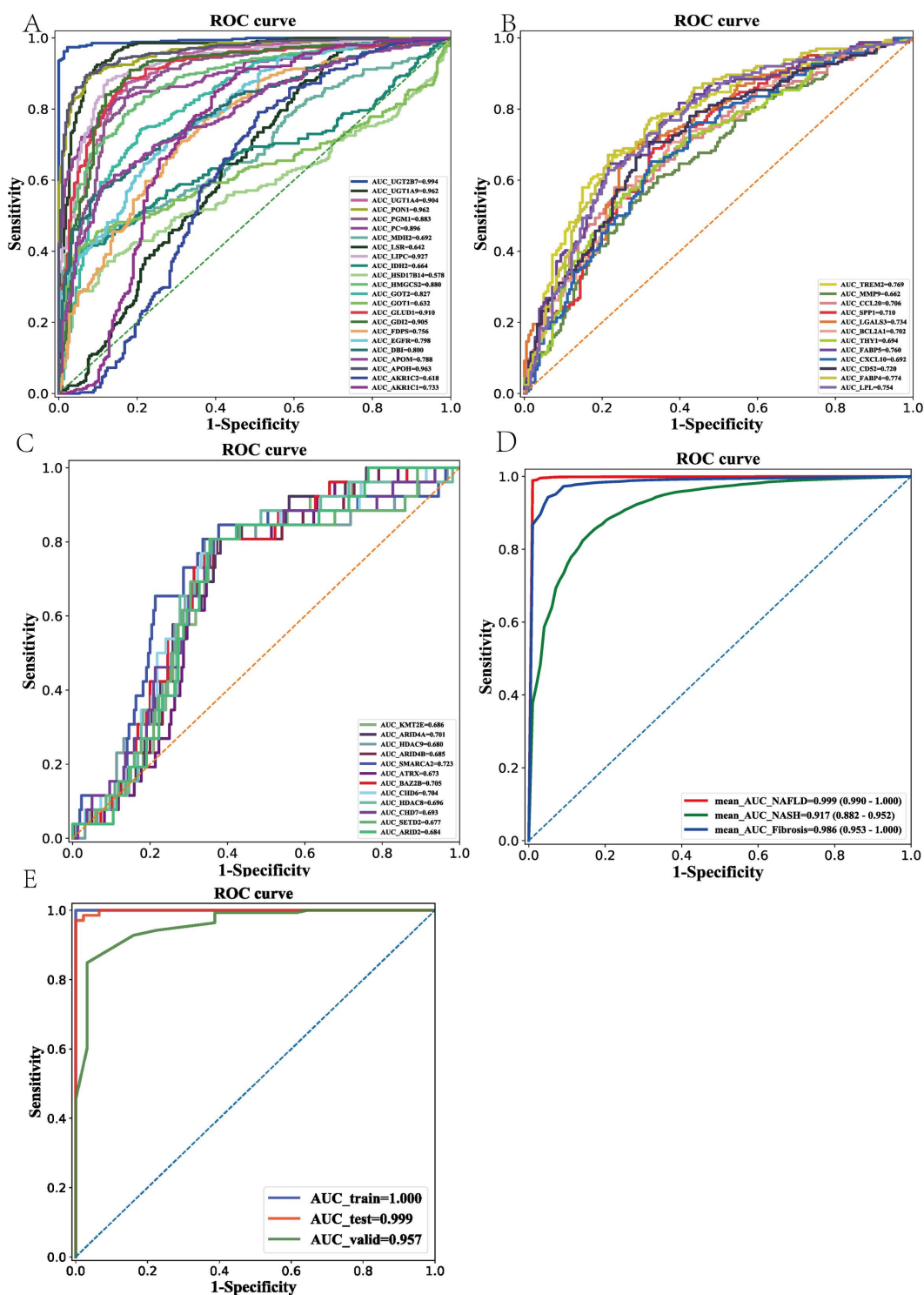
### Discussion

MASLD is a complex disease characterized by the accumulation of fat droplets in liver cells, which manifest as steatosis in the early stages. Ongoing oxidative stress and inflammation can trigger further liver damage that progresses to MASH [7]. Approximate 40 percent of patients with MASH progressively deteriorate and progress to liver fibrosis and cirrhosis [19].



**Fig 3. Genes, modules, and pathways associated with the fibrosis in MASLD.** (A) Venn diagram of DEGs with the fibrosis. (B) Volcano plot of DEGs with the fibrosis. The red dots in the plot represents upregulated genes and blue dots represents downregulated genes with statistical significance. Gray dots represent no differentially expressed genes. (C) Heatmap of significantly DEGs with the fibrosis. The color from blue to red represents the progression from low expression to high expression. (D) The most significant KEGG pathways of DEGs. (E) Go analysis. The figure represents s BP, CC and MF of the DEGs overlapped with those screened by correlation analysis. (F) The most significant enriched module in the PPI network.

<https://doi.org/10.1371/journal.pone.0324972.g003>



**Fig 4. Diagnostic models for MASLD, MASH and fibrosis.** (A) ROC curve of each single candidate gene for MASLD diagnosis. (B) ROC curve of each single candidate gene for MASH diagnosis. (C) ROC curve of each single candidate gene for fibrosis diagnosis. (D) ROC curves of models with combined candidate genes for MASLD, MASH and fibrosis diagnoses. (E) ROC curves of models with combined candidate genes for MASLD, MASH and fibrosis diagnoses in independent datasets.

<https://doi.org/10.1371/journal.pone.0324972.g004>

MASH is a key in the progression of MASLD. Besides the abnormal liver lipid deposition, MASH is also accompanied by pathological changes such as hepatocyte swelling and liver inflammation, and the pathogenesis of MASH is still unclear. There are significant challenges in the diagnosis and treatment of MASLD complicated with fibrosis, and the molecular mechanisms underlying the progression of liver fibrosis in patients with MASLD are unclear [20]. The progression of liver fibrosis in patients with MASH remains untreated, and more research is needed to ascertain molecular processes of liver fibrosis in MASH and to identify important therapeutic targets [21,22]. Therefore, it is necessary to further explore the occurrence, pathogenic factors and related mechanisms of MASH, liver fibrosis. Moreover, alternative diagnostic models for MASLD, MASH, and fibrosis are anticipated to be developed through the investigation of genes and pathways associated with their progression, the refinement of pathogenic factors, and the identification of key genes.

Several omics studies have demonstrated that the occurrence of MAFLD is mainly associated with steatosis, while the occurrence of MASH is predominantly related to inflammation and fibrosis. Gene expression profiles in liver tissue change with the progression of MASLD. For example, the expression of lipid metabolism-related genes is significantly increased after hepatic lipodeposition and steatosis, and is also significantly increased after MASH progression or liver fibrosis [23–26]. Expression of genes related to inflammation and fibrosis is also significantly elevated during the progressive phase of MASLD. Therefore, there is a significant correlation between the expression of certain genes in liver tissue and the severity of MASLD disease activity and the stage of clinical progression [23]. Thus, measuring the expression levels of these genes may help determine the clinical subtype or disease severity of MASLD or help to predict the risk of disease progression [27,28].

MASLD related hub genes were found to enrich in pathways related to glycolipid metabolism. MASLD is often correlated with obesity, diabetes mellitus type 2 (T2DM), and other disorders associated with dysfunction in glucose and lipid metabolism. Studies have demonstrated that T2DM predisposes to the development of MASLD through insulin resistance and hyperglycemia, and this may consequently contribute to the risk of cirrhosis and malignant tumors of the liver [29,30]. In fatty liver mice model, downregulation of *Apolipoprotein H* gene was found to aggravate fatty liver and induce gut microbiota dysbiosis by dysregulating bile acids [31]. Liu et al. found a significant inhibitory of *FDPS* gene can attenuate the mouse NASH-related phenotype. Overexpression of *FDPS* in mice causes increased lipid accumulation, inflammation, and fibrosis, while hepatic *FDPS* deficiency protects mice from NASH progression [32].

MASH is considered as an inflammatory subtype of MASLD with steatosis and evidence of hepatocellular damage and interactions among multiple immune cells [33]. Our data found that 38 candidate genes (e.g., *FABP4* and *MMP9*) in MASH were enriched in inflammation and lipid metabolism pathways. A recent study discovered that the expression levels of *FABP4* and *MMP9* can effectively identify patients who are at risk of progressing from MAFLD to MASH, as well as those at risk of advancing from MASH to cirrhosis and HCC, respectively [34]. Zhao [35] et al. further demonstrated that the protein expression level of MMP-9 was significantly increased when exposed to free fatty acids through in vitro experiments on HepG2 cells. In previous studies, it was found that the serum FABP4 level in MASLD was positively correlated with the severity of hepatic steatosis. The expression of *FABP4* was mainly distributed in liver sinusoidal endothelial cells, which was significantly increased in mice with high fat diet [36]. Subsequently, Zhou et al. used flow cytometry to identify macrophages and found that *FABP4* in liver sinusoidal endothelial cells may play a pathogenic role in the course of MASLD by activating NF- $\kappa$ B/p65 signaling to promote CXCL10-mediated macrophage M1 polarization and CXCR3 macrophage infiltration [37].

DEGs in fibrosis were found to enrich in the Lysine degradation, Ubiquitin mediated proteolysis pathways, and so on. Lysine acetylation can regulate Smad2 transcriptional activity, which inhibits hepatic stellate cells activation and liver fibrosis [38]. Ubiquitin-mediated proteolysis plays a significant role in the pathogenesis of liver fibrosis, a condition characterized by the excessive accumulation of extracellular matrix proteins. This process involves the tagging of proteins with ubiquitin, marking them for degradation by the proteasome, which is crucial for maintaining cellular homeostasis and regulating various biological processes [39]. In agreement, The E3 ubiquitin ligase RNF41 has been shown to orchestrate



macrophage-driven fibrosis resolution and hepatic regeneration [40]. Another study using multi-omics integration emphasized the role of ubiquitination in CCl<sub>4</sub>-induced liver fibrosis [41]. Moreover, the therapeutic role of some ubiquitin-like proteins has been addressed in the context of liver fibrosis, indicating potential targets for therapeutic intervention [42].

Interestingly, we discovered that the development of MASLD, MASH and fibrosis were all associated with the lipid metabolism processes. The discovered pathways (or GO terms) related to the lipid metabolism process were the common ones among MASLD, MASH and fibrosis. Therefore, lipid metabolism may not only involve in the development of MASLD but also may play a synergistic role in MASH and related liver fibrosis. This suggests that the pathways and genes related to lipid metabolism may be one of the key points that we need to pay attention to in the diagnosis, treatment and drug development of MASLD.

Currently, liver biopsy is the gold standard method that can reliably distinguish MASH from MASLD and accurately measure the degree of fibrosis. However, the consistency and accuracy of the diagnosis may be compromised due to the experiential limitations and subjective bias of pathologists [6,43]. For example, the judgment of fibrosis grade and MASH often varies among pathologists [44]. In our study, we established alternative diagnostic models for MASLD, MASH, and fibrosis by integrating RNA expression profiles of a small set of genes, respectively. The gene expression quantification assays, such as gene expression microarray and high-throughput target sequencing, have been standardized, guaranteeing that the outcomes are objective and devoid of human subjective biases. Furthermore, once trained, the model generates precise and consistent results based on specific data inputs. Consequently, the integration of these models with gene expression quantification experiments can contribute to automate, batch, and objectify the diagnostic processes.

It is possible that the samples used in our study are only a partial representation of the MASLD population, which may lead to the loss identification of some other important hub genes. Some batch effects may also exist in our collected datasets and causes misidentification of hub genes. This also means the constructed model that based on the identified hub genes in our study may be not robustness and not have a universal applicability in all MASLD population, especially the MASH and Fibrosis models which lacks the independent validation dataset. Another limitation of our study is we didn't address the influence of other possible variability (e.g., race, ethnicity) to the MASLD, MASH and fibrosis diseases due to the lack of those information. In our future study, we will collect some MASLD, MASH and Fibrosis samples to verify the robustness and clinical application of our models. We will also pay a close attention to the public datasets, which have detail population demographics and clinic information about NAS score and fibrosis stage, to further validate our models.

In conclusion, our study integrates multiple datasets to identify the hub genes that involved in the development and progression of MASLD, MASH, and fibrosis. Three diagnostic models based on identified hub genes show good performance in the diagnosis of MASLD, MASH, and fibrosis, implicating its potential clinical values in disease diagnosis. However, further validation and refinement of these models is necessary before they can be applied in clinical practice.

## Supporting information

**S1 File. Supplemental Digital Content.** Description of detailed data sources and clinical information.  
(XLSX)

**S2 File. Supplemental Digital Content.** Results of GO and KEGG enrichment analysis with DEGs in MASLD.  
(XLSX)

**S3 File. Supplemental Digital Content.** Results of GO and KEGG enrichment analysis with DEGs in NASH.  
(XLSX)

**S4 File. Supplemental Digital Content.** Results of GO and KEGG enrichment analysis with DEGs that overlapped with screened genes using ordinal logistic regression in NASH.  
(XLSX)

**S5 File. Supplemental Digital Content.** Results of GO and KEGG enrichment analysis of DEGs with the fibrosis in MASLD.

(XLSX)

**S6 File. Description of correlation analysis and ROC analysis.**

(DOCX)

## Author contributions

**Formal analysis:** Hong Lu, MengYao Zheng, Min Zhang, Heqing Huang.

**Investigation:** Hong Lu, Yiling Chen.

**Methodology:** Ziyong Mao.

**Writing – original draft:** Hong Lu, MengYao Zheng.

**Writing – review & editing:** Ziyong Mao, Long Lv, Zutao Chen.

## References

1. Younossi ZM. Non-alcoholic fatty liver disease - a global public health perspective. *J Hepatol.* 2019;70(3):531–44. <https://doi.org/10.1016/j.jhep.2018.10.033> PMID: [30414863](https://pubmed.ncbi.nlm.nih.gov/30414863/)
2. Kechagias S, Ekstedt M, Simonsson C, Nasr P. Non-invasive diagnosis and staging of non-alcoholic fatty liver disease. *Hormones (Athens).* 2022;21(3):349–68. <https://doi.org/10.1007/s42000-022-00377-8> PMID: [35661987](https://pubmed.ncbi.nlm.nih.gov/35661987/)
3. Theodorakis N, Nikolaou M. From cardiovascular-kidney-metabolic syndrome to cardiovascular-renal-hepatic-metabolic syndrome: proposing an expanded framework. *Biomolecules.* 2025;15(2):213. <https://doi.org/10.3390/biom15020213> PMID: [40001516](https://pubmed.ncbi.nlm.nih.gov/40001516/)
4. Omran M, Omer M, Mohamed AA, Abdelghafour RA, Muharram NM, Hassan MB, et al. Development and validation of nonalcoholic fatty liver disease test: a simple sensitive and specific marker for early diagnosis of nonalcoholic fatty liver disease. *Eur J Gastroenterol Hepatol.* 2023;35(8):874–80. <https://doi.org/10.1097/MEG.0000000000002575> PMID: [37395240](https://pubmed.ncbi.nlm.nih.gov/37395240/)
5. Wang J-L, Jiang S-W, Hu A-R, Zhou A-W, Hu T, Li H-S, et al. Non-invasive diagnosis of non-alcoholic fatty liver disease: current status and future perspective. *Heliyon.* 2024;10(5):e27325. <https://doi.org/10.1016/j.heliyon.2024.e27325> PMID: [38449611](https://pubmed.ncbi.nlm.nih.gov/38449611/)
6. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology.* 2005;41(6):1313–21. <https://doi.org/10.1002/hep.20701> PMID: [15915461](https://pubmed.ncbi.nlm.nih.gov/15915461/)
7. Perakakis N, Polyzos SA, Yazdani A, Sala-Vila A, Kountouras J, Anastasilakis AD, et al. Non-invasive diagnosis of non-alcoholic steatohepatitis and fibrosis with the use of omics and supervised learning: a proof of concept study. *Metabolism.* 2019;101:154005. <https://doi.org/10.1016/j.metabol.2019.154005> PMID: [31711876](https://pubmed.ncbi.nlm.nih.gov/31711876/)
8. Ratzliff V, Sanyal A, Harrison SA, Wong VW-S, Francque S, Goodman Z, et al. Cenicriviroc treatment for adults with nonalcoholic steatohepatitis and fibrosis: final analysis of the phase 2b CENSAUR study. *Hepatology.* 2020;72(3):892–905. <https://doi.org/10.1002/hep.31108> PMID: [31943293](https://pubmed.ncbi.nlm.nih.gov/31943293/)
9. Kotsiliti E, Leone V, Schuehle S, Govaere O, Li H, Wolf MJ, et al. Intestinal B cells license metabolic T-cell activation in NASH microbiota/antigen-independently and contribute to fibrosis by IgA-FcR signalling. *J Hepatol.* 2023;79(2):296–313. <https://doi.org/10.1016/j.jhep.2023.04.037> PMID: [37224925](https://pubmed.ncbi.nlm.nih.gov/37224925/)
10. Reinson T, Buchanan RM, Byrne CD. Noninvasive serum biomarkers for liver fibrosis in NAFLD: current and future. *Clin Mol Hepatol.* 2023;29(Suppl):S157–70. <https://doi.org/10.3350/cmh.2022.0348> PMID: [36417894](https://pubmed.ncbi.nlm.nih.gov/36417894/)
11. Li Z, Zhang H, Li Q, Feng W, Jia X, Zhou R, et al. GepLiver: an integrative liver expression atlas spanning developmental stages and liver disease phases. *Sci Data.* 2023;10(1):376. <https://doi.org/10.1038/s41597-023-02257-1> PMID: [37301898](https://pubmed.ncbi.nlm.nih.gov/37301898/)
12. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607–13. <https://doi.org/10.1093/nar/gky1131> PMID: [30476243](https://pubmed.ncbi.nlm.nih.gov/30476243/)
13. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 2006;34(Database issue):D535–9. <https://doi.org/10.1093/nar/gkj109> PMID: [16381927](https://pubmed.ncbi.nlm.nih.gov/16381927/)
14. Türei D, Korcsmáros T, Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat Methods.* 2016;13(12):966–7. <https://doi.org/10.1038/nmeth.4077> PMID: [27898060](https://pubmed.ncbi.nlm.nih.gov/27898060/)
15. Li T, Wernersson R, Hansen RB, Horn H, Mercer J, Slodkowicz G, et al. A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat Methods.* 2017;14(1):61–4. <https://doi.org/10.1038/nmeth.4083> PMID: [27892958](https://pubmed.ncbi.nlm.nih.gov/27892958/)

16. Noma H, Shinozaki T, Iba K, Teramukai S, Furukawa TA. Confidence intervals of prediction accuracy measures for multivariable prediction models based on the bootstrap-based optimism correction methods. *Stat Med*. 2021;40(26):5691–701. <https://doi.org/10.1002/sim.9148>
17. Lai MHC. Bootstrap confidence intervals for multilevel standardized effect size. *Multivariate Behav Res*. 2021;56(4):558–78. <https://doi.org/10.1080/00273171.2020.1746902> PMID: 32279536
18. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983;148(3):839–43. <https://doi.org/10.1148/radiology.148.3.6878708> PMID: 6878708
19. Athyros VG, Boutari C, Stavropoulos K, Anagnostis P, Imprialos KP, Doumas M, et al. Statins: an under-appreciated asset for the prevention and the treatment of NAFLD or NASH and the related cardiovascular risk. *Curr Vasc Pharmacol*. 2018;16(3):246–53. <https://doi.org/10.2174/1570161115666170621082910> PMID: 28676019
20. Schwabe RF, Tabas I, Pajvani UB. Mechanisms of fibrosis development in nonalcoholic steatohepatitis. *Gastroenterology*. 2020;158(7):1913–28. <https://doi.org/10.1053/j.gastro.2019.11.311> PMID: 32044315
21. Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol*. 2018;15(1):11–20. <https://doi.org/10.1038/nrgastro.2017.109> PMID: 28930295
22. Friedman SL, Neuschwander-Tetri BA, Rinella M, Sanyal AJ. Mechanisms of NAFLD development and therapeutic strategies. *Nat Med*. 2018;24(7):908–22. <https://doi.org/10.1038/s41591-018-0104-9> PMID: 29967350
23. Cazanave S, Podtelezhnikov A, Jensen K, Seneshaw M, Kumar DP, Min H-K, et al. The transcriptomic signature of disease development and progression of nonalcoholic fatty liver disease. *Sci Rep*. 2017;7(1):17193. <https://doi.org/10.1038/s41598-017-17370-6> PMID: 29222421
24. Kozumi K, Kodama T, Murai H, Sakane S, Govaere O, Cockell S, et al. Transcriptomics identify thrombospondin-2 as a biomarker for NASH and advanced liver fibrosis. *Hepatology*. 2021;74(5):2452–66. <https://doi.org/10.1002/hep.31995> PMID: 34105780
25. Govaere O, Cockell S, Tiniakos D, Queen R, Younes R, Vacca M, et al. Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci Transl Med*. 2020;12(572):eaba4448. <https://doi.org/10.1126/scitranslmed.aba4448> PMID: 33268509
26. Haas JT, Vonghia L, Mogilenko DA, Verrijken A, Molendi-Coste O, Fleury S, et al. Transcriptional network analysis implicates altered hepatic immune function in NASH development and resolution. *Nat Metab*. 2019;1(6):604–14. <https://doi.org/10.1038/s42255-019-0076-1> PMID: 31701087
27. Fujiwara N, Kubota N, Crouchet E, Koneru B, Marquez CA, Jajoriya AK, et al. Molecular signatures of long-term hepatocellular carcinoma risk in nonalcoholic fatty liver disease. *Sci Transl Med*. 2022;14(650):eabo4474. <https://doi.org/10.1126/scitranslmed.abo4474> PMID: 35731891
28. Vandel J, Dubois-Chevalier J, Gheeraert C, Derudas B, Raverdy V, Thuillier D, et al. Hepatic molecular signatures highlight the sexual dimorphism of nonalcoholic steatohepatitis (NASH). *Hepatology*. 2021;73(3):920–36. <https://doi.org/10.1002/hep.31312> PMID: 32394476
29. Chao H-W, Chao S-W, Lin H, Ku H-C, Cheng C-F. Homeostasis of glucose and lipid in non-alcoholic fatty liver disease. *Int J Mol Sci*. 2019;20(2):298. <https://doi.org/10.3390/ijms20020298> PMID: 30642126
30. Zhang Z, Ji G, Li M. Glucokinase regulatory protein: a balancing act between glucose and lipid metabolism in NAFLD. *Front Endocrinol (Lausanne)*. 2023;14:1247611. <https://doi.org/10.3389/fendo.2023.1247611> PMID: 37711901
31. Liu Y, Zhao Y, Liu Q, Li B, Daniel PV, Chen B, et al. Effects of apolipoprotein H downregulation on lipid metabolism, fatty liver disease, and gut microbiota dysbiosis. *J Lipid Res*. 2024;65(1):100483. <https://doi.org/10.1016/j.jlr.2023.100483> PMID: 38101620
32. Liu J, Zhang X, Zhang Y, Qian M, Yang M, Yang S, et al. Farnesyl diphosphate synthase exacerbates nonalcoholic steatohepatitis via the activation of AHR-CD36 axis. *FASEB J*. 2023;37(7):e23035. <https://doi.org/10.1096/fj.202300433RR> PMID: 37310396
33. MacParland SA, Liu JC, Ma X-Z, Innes BT, Bartczak AM, Gage BK, et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat Commun*. 2018;9(1):4383. <https://doi.org/10.1038/s41467-018-06318-7> PMID: 30348985
34. Coilly A, Desterke C, Guettier C, Samuel D, Chiappini F. FABP4 and MMP9 levels identified as predictive factors for poor prognosis in patients with nonalcoholic fatty liver using data mining approaches and gene expression analysis. *Sci Rep*. 2019;9(1):19785. <https://doi.org/10.1038/s41598-019-56235-y> PMID: 31874999
35. Zhao Y, Yakufu M, Ma C, Wang B, Yang J, Hu J. Transcriptomics reveal a molecular signature in the progression of nonalcoholic steatohepatitis and identifies PAI-1 and MMP-9 as biomarkers in in vivo and in vitro studies. *Mol Med Rep*. 2024;29(1):15. <https://doi.org/10.3892/mmr.2023.13138>
36. Yang H, Deng Q, Ni T, Liu Y, Lu L, Dai H, et al. Targeted inhibition of LPL/FABP4/CPT1 fatty acid metabolic axis can effectively prevent the progression of nonalcoholic steatohepatitis to liver cancer. *Int J Biol Sci*. 2021;17(15):4207–22. <https://doi.org/10.7150/ijbs.64714> PMID: 34803493
37. Zhou C, Shen Z, Shen B, Dai W, Sun Z, Guo Y, et al. FABP4 in LSECs promotes CXCL10-mediated macrophage recruitment and M1 polarization during NAFLD progression. *Biochim Biophys Acta Mol Basis Dis*. 2023;1869(7):166810. <https://doi.org/10.1016/j.bbadis.2023.166810> PMID: 37487374
38. Zhang J, Li Y, Liu Q, Huang Y, Li R, Wu T, et al. Sirt6 alleviated liver fibrosis by deacetylating conserved lysine 54 on Smad2 in hepatic stellate cells. *Hepatology*. 2021;73(3):1140–57. <https://doi.org/10.1002/hep.31418> PMID: 32535965
39. Shen W, Zhang Z, Ma J, Lu D, Lyu L. The ubiquitin proteasome system and skin fibrosis. *Mol Diagn Ther*. 2021;25(1):29–40. <https://doi.org/10.1007/s40291-020-00509-z> PMID: 33433895
40. Crunkhorn S. RNF41 reverses liver fibrosis. *Nat Rev Drug Discov*. 2023;22(9):697. <https://doi.org/10.1038/d41573-023-00125-6> PMID: 37528208

41. Mercado-Gómez M, Lopitz-Otsoa F, Azkargorta M, Serrano-Maciá M, Lachiondo-Ortega S, Goikoetxea-Usandizaga N, et al. Multi-omics integration highlights the role of ubiquitination in CCl<sub>4</sub>-induced liver fibrosis. *Int J Mol Sci.* 2020;21(23):9043. <https://doi.org/10.3390/ijms21239043> PMID: [33261190](https://pubmed.ncbi.nlm.nih.gov/33261190/)
42. Lachiondo-Ortega S, Mercado-Gómez M, Serrano-Maciá M, Lopitz-Otsoa F, Salas-Villalobos TB, Varela-Rey M, et al. Ubiquitin-like post-translational modifications (Ubl-PTMs): small peptides with huge impact in liver fibrosis. *Cells.* 2019;8(12):1575. <https://doi.org/10.3390/cells8121575>
43. Pai RK, Kleiner DE, Hart J, Adeyi OA, Clouston AD, Behling CA, et al. Standardising the interpretation of liver biopsies in non-alcoholic fatty liver disease clinical trials. *Aliment Pharmacol Ther.* 2019;50(10):1100–11. <https://doi.org/10.1111/apt.15503> PMID: [31583739](https://pubmed.ncbi.nlm.nih.gov/31583739/)
44. Sanyal AJ, Chalasani N, Kowdley KV, McCullough A, Diehl AM, Bass NM, et al. Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. *N Engl J Med.* 2010;362(18):1675–85. <https://doi.org/10.1056/NEJMoa0907929> PMID: [20427778](https://pubmed.ncbi.nlm.nih.gov/20427778/)