**Integrative** **multi-omics analysis** **identifies** **distinct molecular subtypes of NAFLD in a Chinese population**

Jingjing Ding1, 2, 3, 4#, Huaizheng Liu5#, Xiaoxun Zhang1, 2, 3, 4#, Nan Zhao1, 2, 3, 4#, Ying Peng1, 2, 3, 4#, Junping Shi6#, Jinjun Chen7#, Xiaoling Chi8#, Ling Li1, 2, 3, 4, Mengni Zhang1, 2, 3, 4, Wen-Yue Liu9, Liangjun Zhang1, 2, 3, 4, Jiafeng Ouyang1, 2, 3, 4, Qian Yuan1, 2, 3, 4, 10, Min Liao1, 2, 3, 4, Ya Tan1, 2, 3, 4, Mingqiao Li1, 2, 3, 4, Ziqian Xu1, 2, 3, 4, Wan Tang1, 2, 3, 4, Chuanming Xie11, Yi Li12, Qiong Pan1, 2, 3, 4, Ying Xu13, Shi-Ying Cai14, Christopher D. Byrne15, Giovanni Targher16, Xinshou Ouyang17\*\*, Liqun Zhang12\*\*, Zhongyong Jiang18\*\*, Ming-Hua Zheng19, 20\*\*, Fengjun Sun10\*\*, Jin Chai1, 2, 3, 4\*

1Department of Gastroenterology,the First Affiliated Hospital (Southwest Hospital), Third Military Medical University (Army Medical University), Chongqing 400038, China;

2Institute of Digestive Diseases of PLA, Third Military Medical University (Army Medical University), Chongqing 400038, China;

3Cholestatic Liver Diseases Center, the First Affiliated Hospital (Southwest Hospital), Third Military Medical University (Army Medical University), Chongqing 400038, China;

4 Metabolic Dysfunction-Associated Fatty Liver Disease (MASLD) Medical Research Center,the First Affiliated Hospital (Southwest Hospital), Third Military Medical University (Army Medical University), Chongqing 400038, China;

5Department of Emergency, the Third Xiangya Hospital, Central South University, Changsha 410013, China;

6Department of Infectious Diseases and Hepatology, the Affiliated Hospital of Hangzhou Normal University, Hangzhou 310015, Zhejiang, China;

7Hepatology Unit, Department of Infectious Diseases, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China;

8Department of Hepatology, Guangdong Provincial Hospital of Chinese Medicine，the Second Affiliated Hospital of Guangzhou University of Chinese Medicine，Guangzhou 510405，China;

9Department of Endocrinology, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, China;

10Department of Pharmacy, the First Affiliated Hospital (Southwest Hospital), Third Military Medical University (Army Medical University), Chongqing 400038, China;

11Institute of Hepatobiliary Surgery, the First Affiliated Hospital (Southwest Hospital), Third Military Medical University (Army Medical University), Chongqing 400038, China;

12Department of Clinical Laboratory, the Second Affiliated Hospital, Army Medical University, Chongqing 400037, China;

13School of Clinical Medicine & the First Affiliated Hospital of Chengdu Medical College, Chengdu 610500, China;

14Department of Internal Medicine and Liver Center, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520, USA;

15Southampton National Institute for Health and Care Research Biomedical Research Centre, University Hospital Southampton and University of Southampton, Southampton General Hospital, Southampton SO16 6YD, UK;

16Metabolic Diseases Research Unit, IRCCS Sacro Cuore - Don Calabria Hospital, Negrar di Valpolicella 37024, Italy;

17Department of Internal Medicine, Section of Digestive Diseases, Yale University School of Medicine, New Haven, CT 06520, USA;

18Department of Medical Laboratory, Cheng du Seventh People's Hospital (Affiliated Cancer Hospital of Chengdu Medical College), Sichuan 610213, China;

19MAFLD Research Center, Department of Hepatology, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, China;

20Key Laboratory of Diagnosis and Treatment for the Development of Chronic Liver Disease in Zhejiang Province, Wenzhou 325000, China.

#These authors contributed equally to this study.

\*Jin Chai, E-mail: [jin.chai@cldcsw.org](mailto:jin.chai@cldcsw.org)

\*\*Fengjun Sun, E-mail: fj-sun@tmmu.edu.cn;

\*\*Ming-Hua Zheng, E-mail: [zhengmh@wmu.edu.cn](mailto:zhengmh@wmu.edu.cn);

\*\*Zhongyong Jiang, E-mail: jiangzhongyong@cmc.edu;

\*\*Liqun Zhang, E-mail: liqunzhang@tmmu.edu.cn;

\*\*Xinshou Ouyang, E-mail: xinshou.ouyang@yale.edu.

**Overline: LIVER DISEASE**

**One Sentence Summary:** Integrative multi-omics analysis identifies three molecular subtypes of NAFLD associated with distinct clinical risks.

**Abstract**

Non-alcoholic fatty liver disease (NAFLD) has become a common healthcare burden worldwide. The high heterogeneity of NAFLD remains elusive and impairs outcomes of clinical diagnosis and pharmacotherapy. Several NAFLD classifications have been proposed, based on clinical, genetic, alcoholic or serum metabolic analyses. Yet, accurately predicating the progression to cirrhosis or hepatocellular carcinoma (HCC) in patients with NAFLD remains a challenge. Here, based on a Chinese cohort of patients, we classified NAFLD into three distinct molecular subtypes (NAFLD-mSI, -mSII, and -mSIII) using integrative multi-omics including whole genome sequencing (WGS), proteomics, phosphoproteomics, lipidomics, and metabolomics across a broad range of liver, blood, and urine specimens. We found that NAFLD-mSI had higher expression of CYP1A2 and CYP3A4, which alleviate hepatic steatosis through mediating free fatty acid/bile acid-mTOR-FXR/PPARα signaling. NAFLD-mSII displayed an elevated risk of liver cirrhosis along with increased hepatic infiltration of M1 and M2 macrophages due to lipid-triggered hepatic CCL2 and CRP production. NAFLD-mSIII exhibited a potential risk for HCC development by increased transcription of CEBPB- and ERCC3-regulated oncogenes due to activation of the EGF-EGFR/CHKA/PI3K-PDK1-AKT cascade. Next, we validated the existence of these three NAFLD molecular subtypes in an external cohort comprising 92 patients with NAFLD across three different Chinese hospitals. These findings may aid in understanding the molecular features underlying NAFLD heterogeneity, thereby facilitating clinical diagnosis and treatment strategies with the aim of preventing the development of liver cirrhosis and HCC.

**INTRODUCTION**

The incidence of non-alcoholic fatty liver disease (NAFLD) is rapidly increasing and is becoming a common healthcare burden worldwide (*1*). NAFLD and its progressive form non-alcoholic steatohepatitis (NASH) are characterized by hepatic lipid accumulation and are associated with increased risk of cirrhosis or hepatocellular carcinoma (HCC) (*2*). There was no specific drug for NASH treatment until recently, when the Food and Drug Administration approved resmetirom as the first medication for this purpose. Notably, treatment with resmetiron only benefits 29.9% of patients with NASH (*3, 4*), and it remains unclear which patients would benefit from treatment. In addition, many studies have reported that agonists for nuclear receptors farnesoid X receptor (FXR) (such as obeticholic acid, OCA) and peroxisome proliferator-activated receptor-α (PPARα) (such as fenofibrate) can ameliorate hepatic steatosis and fibrosis in mouse models of NASH (*5, 6*). Unfortunately, the histological outcomes of treatment with OCA or fenofibrate remain unsatisfactory in most patients with NASH (*7, 8*). This lack of consistency is likely attributed to the high heterogeneity of human NAFLD, because the susceptibility to developing NAFLD is highly variable and influenced by a variety of risk factors including aging, genotype, sex, ethnicity, dietary intake, bile acid metabolism, and insulin resistance (*9*).

To address its high heterogeneity, several classifications of NAFLD have been proposed. Lean NAFLD, which constitutes ~20% of NAFLD cases, has distinct metabolomic profiles that differ from their obese counterparts (*10*, *11*). Nevertheless, both obese NAFLD and lean NAFLD can progress into NASH, cirrhosis, and eventually into HCC (*12*). Furthermore, so-called ‘*PNPLA3* NAFLD’ has been recently proposed, with a I148M mutation in the patatin-like phospholipase domain-containing protein 3 (*PNPLA3*) gene that can increase hepatic fat accumulation, but without insulin resistance (*13*). However, patients with ‘*PNPLA3* NAFLD’ can also carry other gene polymorphisms contributing to the development and progression of NAFLD (*14*). A recent study analyzed the serum metabolomes of patients to identify two major metabolic subtypes of NAFLD and to differentiate simple steatosis from NASH (*15*). Along with cardiovascular disease and genetic risk factors, metabolomic signatures have been used for the classification of three serum metabolic subgroups of NAFLD, which were independent of histological disease severity (*16*). However, NAFLD may occur without insulin resistance, which is a key component of the metabolic syndrome, and NAFLD may also progress into HCC without coexisting cirrhosis or metabolic syndrome (*17*). Thus, advanced classifications of NAFLD supporting clinical practice are required.

To investigate the advanced classifications of NAFLD, we performed multi-omics analysis, including whole genome sequencing (WGS), proteomic, phosphoproteomic, lipidomic, and metabolomic analyses on liver, serum, and urine specimens from several cohorts of patients. Here, we define three NAFLD molecular subtypes and determine their clinical features, molecular mechanisms, omics characteristics, and associated markers. Our findings may shed light on precision diagnosis and potential therapeutic targets for treating and preventing NAFLD and its adverse clinical outcomes.

**RESULTS**

***Overview of integrative multi-omics analysis and machine learning to identify and validate NAFLD molecular subtypes in multiple cohorts***

To explore the use of multi-omics data for classifying the NAFLD spectrum, we conducted an unbiased approach by performing WGS, proteomics, phosphoproteomics, lipidomics, and metabolomics in liver, blood, and urine specimens from patients with NAFLD or non-NAFLD and healthy controls in the internal cohort (Fig. 1, left, fig. S1, and tables S1 to S3). Principal component analysis (PCA) revealed a spatial separation between control and NAFLD (fig. S2).Non-negative matrix factorization (NMF) using the resulting 1346\*103 protein expression matrix then classified NAFLD into three molecular subtypes (mSI, mSII, and mSIII). Detailed information on the number of patients in each of the three subtypes is shown in table S4.

We next analyzed whether NAFLD-related genetic variants affected the clustering of these three molecular subtypes. We used PCA to reduce the high-dimensional protein expression matrix and analyzed the correlations between the top 10 PCs and NAFLD-associated genetic variants. There was no significant association (*P* >= 0.05) between PCs and these genetic variants (fig. S3A), suggesting that they did not influence these three NAFLD molecular classifications. Furthermore, there was no significant difference (*P* >= 0.05) in the distribution of these genetic variants among the NAFLD molecular subtypes (fig. S3, B and C). Next, hepatic proteins associated with the NAFLD molecular subtypes were characterized and validated by machine learning using proteomic or multiplex IF data. Last, the distinct clinical and molecular characteristics of the three NAFLD subtypes observed in the internal cohort were validated in the external cohort (Fig. 1, right).

***Liver proteomic profiling separates NAFLD into three distinct molecular subtypes in the internal cohort***

We performed proteomics on 17 surgical liver samples from a different non-NAFLD control individuals and 103 biopsied liver samples from consecutively recruited patients with NAFLD. Two experienced liver pathologists examined these samples using the NASH Clinical Research Network Scoring System (Fig. 2A) as described previously (*18*). 7205 proteins were quantified by proteomics for both non-NAFLD and NAFLD patients (Fig. 2B). 921 proteins were significantly up-regulated and 321 down-regulated (*P* < 0.05) in patients with NAFLD compared with non-NAFLD (Fig. 2C). Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis revealed seven pathways that were enriched in NAFLD samples (fig. S4), consistent with previous studies (*19-22*).

Next, 1346 proteins (those with the top 25% coefficient of variation) in the NAFLD group were analyzed by NMF (fig. S5). Cophenetic correlation coefficient and consensus clustering analyses defined three distinct NAFLD molecular subtypes with different molecular and histological characteristics (Fig. 2, D to F). Notably, a comparison analysis of the NAFLD molecular and liver histological subtypes revealed that the proportion of patients with NASH in the NAFLD-mSI subtype was lower than those in the NAFLD-mSII and mSIII subtypes (table S5). Conversely, there was no significant difference (*P* >= 0.05) in the proportion of patients with NAFL or NASH and fibrosis between the NAFLD-mSII and NAFLD-mSIII subtypes (table S5). These data support that the three NAFLD molecular subtypes differ from traditional liver histological subtypes (from NAFL to NASH, with and without fibrosis).

**NAFLD-mSI maintains high hepatic CYP1A2 and CYP3A4 against hepatic steatosis by mediating free fatty acid (FFA)/BAs-mTOR-FXR/PPARα signaling**

Liver proteomics and bioinformatics analysis revealed 103 differentially expressed proteins (DEPs) in the NAFLD-mSI group, compared to NAFLD-mSII and NAFLD-mSIII groups (Fig. 3A). These DEPs were enriched in 10 KEGG pathways involving xenobiotics, endogenous or drug metabolism (Fig. 3 A and B). Four proteins, CYP1A2, CYP3A4, GSTA2 and GSTT2B, were involved in the enriched KEGG pathways and were higher in the NAFLD-mSI than the NAFLD-mSII and NAFLD-mSIII groups (Fig. 3, C and D). Hepatic CYP1A2 and CYP3A4, but not GSTA2 and GSTT2B, were negatively correlated with hepatic steatosis and histologic NAS scores in this population (Fig. 3E). Hepatic CYP1A2 and CYP3A4 were inversely correlated with serum lipids (Fig. 3F), the lipids were significantly lower (*P* < 0.05) in the NAFLD-mSI group than the other two groups (fig. S6), suggesting that these proteins may regulate lipid accumulation. Furthermore, siRNA-mediated knockdown of *CYP1A2* or *CYP3A4* significantly increased (*P* < 0.05) the accumulation of lipid droplets in LO2 cells after PO (palmitic acid/oleic acid, PA/OA) treatment (fig. S7). Hence, down-regulated hepatic CYP1A2 and CYP3A4 expression likely aggravates hepatic lipid accumulation during NAFLD.

Additional analyses revealed that hepatic CYP1A2 and CYP3A4 proteins correlated with the abundances of circulating conjugated BAs (taurocholic acid, TCA; glycocholic acid, GCA) and FFAs (FFA 11:1; 18:1) in all patients with NAFLD (Fig. 3F). Serum lipidomic and metabolomic data revealed that the NAFLD-mSI group had lower TCA and GCA and higher FFAs (FFA 11:1 and 18:1) compared to the NAFLD-mSII and SIII groups (Fig. 3G). Western blotting analysis showed that the NAFLD-mSI group also had higher hepatic CYP1A2 and CYP3A4 but lower FXR and PPARα protein abundance (Fig. 3H). Furthermore, CD36 protein, but not NTCP, was lower in the NAFLD-mSI group than the other two groups (fig. S8). These data suggest that hepatic conjugated BAs and FFAs may be reduced in the NAFLD-mSI group. Of note, FFAs and BAs can activate PPARα and FXR, inhibiting hepatic CYP3A4 (*23*). We found that FXR or PPARα activation by GW4064 or fenofibrate treatment markedly reduced CYP1A2 and CYP3A4 in LO2 cells (Fig. 3, I and J). These findings support the notion that The conjugated BAs or FFAs were taken up by the liver via NTCP or CD36, enhancing FXR or PPARα activation and subsequently repressed CYP1A2 and CYP3A4, thereby aggravating hepatic fat accumulation during NAFLD progression.

We further performed phosphoproteomics of liver tissues from non-NAFLD individuals (n=16) and patients with NAFLD (n=90) (Fig. 3K and L). KEGG analysis showed that hepatic mTOR signaling was highly enriched in the NAFLD-mSI group (Fig. 3M). Hepatic mTOR phosphorylation was also notably higher in NAFLD-mSI (Fig. 3N). Moreover, there was a potential protein-protein interaction (PPI) network predicted among the hepatic phosphoproteins comprising mTOR signaling, FXR, PPARα, CYP1A2, and CYP3A4 proteins (Fig. 3O). FFAs can activate mTOR signaling (*24*). We observed that conjugated BAs TCA and GCA also activated mTOR signaling in PLC/PRF/5-*ASBT* cells (Fig. 3P). Treatment with OSI-027, a specific mTOR inhibitor, markedly increased FXR and PPARα proteins in BA-treated PLC/PRF/5-*ASBT* cells and PO-treated LO2 cells (Fig. 3Q), indicating that mTOR signaling activation represses FXR and PPARα. Therefore, low concentrations of liver BAs and FFAs regulated by hepatic transporters NTCP and CD36 maintained high hepatic CYP1A2 and CYP3A4 abundance against hepatic steatosis by controlling mTOR signaling-suppressed FXR and PPARα activities in NAFLD-mSI (Fig. S9).

**NAFLD-mSII displays a high risk for liver cirrhosis through lipid-triggered CRP and CCL2 production that promotes increased hepatic M1/M2 macrophage infiltration**

KEGG analysis of proteomic data revealed that ECM-receptor interaction and focal adhesion pathways were enriched in NAFLD-mSII (Fig. 4A). LAMA2, LAMA4, LAMB1, LAMC1, FLNA, FLNC, MYL2, and MYLPF were enriched in these two pathways and significantly higher (*P* < 0.05) in the NAFLD-mSII group than the other groups (Fig. 4, B and C). LAMs and FLNs contribute to increased liver fibrogenesis (*25-27*), and we found their abundance was associated with liver fibrosis and NAS scores in patients with NAFLD, and also significantly correlated (*P* < 0.05) with hepatic fibrotic biomarkers (Fig. 4, D and E). Proteomic analysis revealed that hepatic α-SMA, COL1A1, COL1A2, and COL3A1 protein abundances were also significantly higher (*P* < 0.05) in NAFLD-mSII than the other groups (Fig. 4F). Consistent with this, a similar pattern of protein abundance was detected by Western-blotting analysis (Fig. 4G and fig. S10). Moreover, hepatic TGFβ and SMAD2/3 phosphorylation were higher in NAFLD-mSII (Fig. 4H). These data indicated that enhanced activation of hepatic TGFβ-SMAD2/3 signaling promoted liver fibrosis in the NAFLD-mSII group. In addition, there was a significant enrichment in hepatic complement and coagulation cascades (*P* < 0.05) and systemic lupus erythematosus (*P* < 0.05) pathway-related inflammation and hepatic proinflammatory TNFα, IL-1β, and IL-6 proteins were also higher in NAFLD-mSII (Fig. 4I), highlighting the importance of inflammation in disease pathogenesis in this group.

KEGG analysis of liver phosphoproteomics revealed enrichment of the leukocyte transendothelial migration pathway in NAFLD-mSII (Fig. 4J). Bioinformatic analysis of liver proteomics and leukocyte-related genes identified distinct hepatic proteins in NAFLD-mSII compared to those in neutrophils, lymphocytes, monocytes, and macrophages from the Molecular Signatures Database (MSigDB) (fig. S11, A to D). In addition, hepatic CD68, CD11c, and CD163, but not MPO and CD8α, were higher in the NAFLD-mSII group than in the other two groups (Fig. 4K). Liver-infiltrated M1 and M2 macrophages (Fig. 4L) and hepatic CCL2, a known chemokine recruiting monocytes and macrophages (Fig. 4M), were higher in the NAFLD-mSII group. CRP can stimulate the differentiation of monocytes into M1 macrophages (*28*). We found that hepatic CRP protein was elevated in the liver and overlapped with macrophage-related genes in NAFLD-mSII, which was associated with enriched inflammatory and immune response pathways (fig. S11, D to F). These findings suggest that CRP promoted monocytes recruited by hepatic CCL2 to differentiate into M1 macrophages, contributing to pathogenesis in patients classified in NAFLD-mSII.

Compared with the other two subtypes, PKM, TPM1, MRC2, and MYL3 were enriched in both serum and liver of NAFLD-mSII (Fig. 4N, left). PKM2 can promote the transdifferentiation of M1 macrophages into M2 macrophages, producing fibrotic cytokines, such as TGFβ1 (*29*). Serum and liver PKM protein (Fig. 4N, right), namely PKM2, but not PKM1 (Fig. 4O), was significantly higher (*P* < 0.05) in NAFLD-mSII, as were serum lipids including diacylglycerols (DGs), phosphatidic acids (PAs), phosphatidylcholines (PCs), and phosphatidylethanolamines (PEs) (Fig. 4P). Furthermore, PO treatment stimulated CCL2 and CRP expression and their secretion in LO2 cells (Fig. 4, Q and R), and induced PKM2 in the THP1-derived M1 macrophages (fig. 4S). In addition, PO treatment also promoted the transdifferentiation of M1 macrophages into M2 macrophages (fig. S12). M2 macrophages can produce TGFβ1 and consequently stimulate hepatic stellate cell (HSC) activation, resulting in the development of liver fibrosis (*30, 31*). These data suggest that during NAFLD-mSII pathogenesis, an increase in serum lipids (such as DGs, PAs) stimulates hepatic CCL2 and CRP production, resulting in monocyte recruitment and M1 macrophage polarization leading to liver inflammation. Moreover, lipid-induced enzymatic activation of PKM2 by its tetramer formation promotes the transdifferentiation of hepatic M1 to M2 macrophages, aggravating liver fibrosis through increased fibrotic cytokine (TGFβ1) production and HSC activation (fig. S13).

**NAFLD-mSIII exhibits potential risk for HCC through EGF-EGFR/CHKA/PI3K-PDK1-AKT signaling-induced CEBPB- and ERCC3-dependent oncogenes**

KEGG analysis of liver proteomic data indicated that basal transcription factors and transcriptional misregulation in cancer pathways were enriched in NAFLD-mSIII (Fig. 5A). GTF2F1, GTF2F2, TAF15, H3-3A, CEBPB, and ERCC3 proteins were enriched in these pathways (Fig. 5B), with increased abundance in NAFLD-mSIII compared to the other groups (Fig. 5C). CEBPB and ERCC3 are associated with HCC development (*32, 33*), and were found to be significantly higher (*P* < 0.05) in liver in NAFLD-mSIII (Fig. 5D). PPI network analysis identified 8 potential canonical caner drivers for HCC development (Fig. 5E). Liver HNRNPA2B1, but not SF3B1 and CTNNB1 was up-regulated in the NAFLD-mSIII group, compared to those in the other two groups (Fig. 5F). Furthermore, induction of CEBPB and ERCC3 led to an increase of oncogenic *CCND1, HNRNPA2B1*, *KRAS*, *KMT2A*, *TERT* and *NFE2L2* expression in Huh7 cells and primary mouse hepatocytes (Fig. 5, G and H, and fig. S14) and promoted the migration and invasion of Huh7 cells (Fig. 5, I and J). In addition, serum concentrations of LY75, CHGA, CD55, and GPD1, proteins which have been associated with the development of cancers (*34-37*), were higher in NAFLD-mSIII group (Fig. 5K). Moreover, hepatic oncogenic HNRNPA2B1, KRAS, and TERT were also higher in NAFLD-mSIII than in the other two groups (fig. S15). Inducing HNRNPA2B1 promoted the migration and invasion of Huh7 cells (fig. S16), consistent with a previous report (*38*). Therefore, our findings support the notion that increased hepatic CEBPB and ERCC3 are associated with an increased risk for HCC development during NAFLD-mSIII pathogenesis.

KEGG analysis of liver phosphoproteomic data indicated that the ErbB signaling pathway was highly enriched in the NAFLD-mSIII group (Fig. 5L). PPI analysis revealed the enrichment of EGFR phosphorylation at a site Ser 1037 in the ErbB signaling pathway, likely enhancing CEBPB (Fig. 5M). Notably, EGFR phosphorylation at Ser 1037 positively correlated with CEBPB and ERCC3 protein abundance (Fig. 5N). Phosphoproteomic analysis showed that EGFR (Ser 1037) phosphorylation was higher in the NAFLD-mSIII group than in the NAFLD-SI and NAFLD-SII groups (Fig. 5O). These data suggest that phospho-EGFR (Ser 1037) may be crucial for inducing CEBPB and ERCC3. In addition, inducing both EGFR (Ser 1037)-WT (wild type) and EGFR-CAT (simulating a constantly phosphorylated status by changing Ser 1037 to Asp) markedly increased CEBPB and ERCC3 protein in Huh7 cells, with more protein upregulation in the EGFR (Ser 1037)-CAT amplification group than the WT overexpression group (Fig. 5P).

Furthermore, EGFR (Ser 1037) phosphorylation in Huh7 cells activated PI3K/AKT signaling and induced protein production of PKD1 (Fig. 5Q), a molecule upstream of AKT signaling, consistent with typical phosphorylated sites of EGFR (Tyr 992 and 1068) (*39*). Moreover, treatment with LY294002, a specific inhibitor of PI3K, or AKTi-1/2, a specific inhibitor of AKT, abolished the EGF-induced increase in CEBPB and ERCC3 in Huh7 cells (Fig. 5R). These data support the notion that EGFR phosphorylation at Ser 1037 enhances CEBPB and ERCC3 by activating PI3K-PDK1-AKT signaling in hepatocytes.

Serum lipidomic and metabolomic analyses revealed phosphocholine was significantly higher (*P* < 0.05) in the NAFLD-mSIII group (Fig. 5S), as was choline kinase α (CHKA), which catalyzes the conversion of choline to phosphocholine (*39*). Pull-down assays conducted in co-transfected 293T cells demonstrated a protein-protein interaction among EGFR (Ser 1037)-WT, CHKA, and PI3K proteins, known to activate AKT signaling by upregulating PKD1 (*39*), whereas EGFR (Ser 1037)-CAT induction markedly increased this protein interaction (fig. S17). In addition, hepatic PI3K (Tyr 467), AKT (Thr 308), and EGFR (Tyr 992 and 1068) phosphorylation and EGF and PKD1 protein abundance were significantly higher (*P* < 0.05) in the NAFLD-mSIII group than in the NAFLD-mSI and mSII groups (Fig. 5T and fig. S18).

To further validate the molecular subtypes identified as associated with HCC development, we collected 7 pairs of tumors and their para-tumor tissues from patients with HCC. Western-blotting analysis revealed a significant increase (*P* < 0.05) in hepatic EGF, CHKA, and PDK1 proteins, as well as phosphorylated forms of EGFR (Tyr 992), PI3K (Tyr 467), and AKT (Thr 308) in tumor tissues compared to their para-tumor counterparts (Fig. 6, A and B). Moreover, hepatic CEBPB, ERCC3 and oncogenic proteins HNRNPA2B1, KRAS, and TERT were elevated in tumor tissues compared to their para-tumoral counterparts (Fig. 6, C and D). These findings were consistent with identified the molecular mechanisms of NAFLD-mSIII.

Collectively, the upregulation of hepatic EGF-activated PI3K-PDK1-AKT signaling resulted in CEBPB and ERCC3 induction by enhancing EGFR phosphorylation and CHKA protein. This led to elevated serum phosphocholine and oncogene activation, which may raise the risk of HCC during NAFLD-mSIII pathogenesis (fig. S19).

In addition, we analyzed whether the increased risk of HCC in NAFLD-mSIII was associated with other potential triggers including fibrosis, obesity or diabetes. In our internal and external validation cohorts, 72% and 44% of patients in the NAFLD-mSIII group had non-severe fibrosis including patients with fibrosis score ≤1, respectively (table S6). The prevalence rates of lean NAFLD, overweight NAFLD, and obese NAFLD in the NAFLD-mSIII subtype accounted for 10%, 23%, and 67% of cases, respectively (table S7). In addition, diabetes-NAFLD, obese-NAFLD, diabetes/obese-NAFLD, and simple NAFLD accounted for 7%, 48%, 19% and 26% of cases, respectively (table S8). The results revealed that the high risk of HCC in NAFLD-mSIII subtype was independent of fibrosis, obesity and diabetes.

**Integrative multi-omics and comprehensive analyses characterize distinct liver histological and molecular characteristics of NAFLD molecular subtypes**

We further characterized the clinical features of these NAFLD molecular subtypes. Histological assessments of liver tissues (Fig. 7A) along with the transient elastography-controlled attenuation parameter (TE-CAP) (fig. S20A) indicated that steatosis and NAS scores in the NAFLD-mSI group were lower than the other groups. Moreover, hepatic fibrosis scores, transient elastography-based liver stiffness (TE-LSM), and the concentration of four serum liver fibrosis markers, including hyaluronidase (HA), procollagen III peptide (PCIII), type IV collagen (C-IV) and laminin (LN) were higher in NAFLD-mSII compared to the other groups, as were lobular inflammation scores (Fig. 7A and fig. S20, B to F). Of note, the proportion of patients with severe liver fibrosis and inflammation was markedly higher in NAFLD-mSII (table S9), indicating that these patients had a high risk of progression to cirrhosis. There was no significant difference (*P* >= 0.05) in hepatic ballooning scores and NAFLD severity degrees among these three groups (Fig. 7, A and B). Serum biochemistry analysis unveiled that the serum fasting insulin, HOMA-IR, C-peptide, ALT, and AST were significantly lower (*P* < 0.05) in the NAFLD-mSI group (fig. S21, A to C and table S10 and 11). Of note, BMI measurements were greater in the NAFLD-mSII group than the other groups (fig. S21D).

Next, we analyzed omics data from liver, serum, and urine. Analysis of the NAFLD-mSI group revealed serum proteins associated with lipid-related biological processes, including apolipoprotein L1 (APOL1) and apolipoprotein F (APOF) (Fig. 7C). There were various significant positive or negative correlations (*P* < 0.05) between serum ALPOL1 and APOF and metabolites or lipids, including TCA, GCA, triglycerides, 5α-cholest-7-en-3β-ol, PEs, lysophosphatidylinositol, and sphingomyelin (Fig. 7C), similar to a previous study in normolipidemic and hyperlipidemic populations (*40*). Urine omics of the NAFLD-mSI group detected 11 proteins related to lipid biological processes (Fig. 7C). The abundance of these proteins was positively or negatively correlated with the amount of urine metabolites (Fig. 7C). Hepatic CYP1A2 was negatively correlated with serum AOPL1 and urine GLRX; hepatic CYP3A4 was positively correlated with serum APOF, but negatively with serum AOPL1 and urine CALB1, CD300A, and GLRX. These results suggest that hepatic CYP1A2 and CYP3A4; serum APOL1, APOF, TGs, PEs, LPIs, SMs, 5α-cholest-7-en-3β-ol, TCA, and GCA; and urine LBP, PRDX3, PDGFRB, CD300A, AXL, KLK1, TYRO3, CALB1, TGFBR3, GLRX, TNR, PC(20:5(5Z,8Z,11Z,14Z,17Z)/20:5(5Z,8Z,11Z,14Z,17Z)), 27-hydroxycholesterol and 3,8-dihydroxy-6-methoxy-7(11)-eremophilen-12,8-olide may functionally interact to regulate hepatic lipid accumulation, contributing to NAFLD-mSI pathogenesis (Fig. 7C).

The NAFLD-mSII group was associated with liver inflammation and fibrosis. Serum omics analysis revealed associations between APOD and ITGA2 with inflammation, and identified 14 proteins linked with fibrosis. These proteins also correlated with 60 metabolites and 9 lipids (Fig. 7D). Urine omics analysis revealed associations between APMAP and SUSD2 with fibrosis. Urinary abundance of APMAP positively correlated with the urine metabolite 7alpha-hydroxy-5beta-cholestan-3-one (Fig. 7D). The above serum or urine proteins related to inflammation or fibrosis, except for PKM and SH3BGRL, were correlated with hepatic LAMA2, LAMA4, LAMB1, LAMC1, FLNA, FLNC, α-SMA, COL1A1, COL1A2, COL3A1, CRP, or PKM proteins (Fig. 7D). Conceivably, these molecules may play roles in the development of severe fibrosis and inflammation during NAFLD-mSII pathogenesis.

Integrative multi-omics data analysis in the NAFLD-mSIII group highlighted cancer pathways. FGFR1 in the urine was associated with cancer development and positively correlated with HCC-related hepatic ERCC3 and HNRNPA2B1 (Fig. 7E). Increased *FGFR1* copy number has been associated with high histopathological grade and high proliferation of breast cancer (*41*), and our findings suggest that urinary FGFR1 may be associated with an increased risk for the development of HCC in NAFLD-mSIII. The top 5 proteins, lipids, and metabolites in serum and urine across these molecular subtypes are shown in Fig. 7, F and G, respectively. It is noteworthy that there were no significant differences (*P* >= 0.05) in urine lipid contents among three NAFLD molecular subtypes.

**Machine learning** **identifies hepatic proteins and histological and molecular features for NAFLD molecular subtypes in internal and external patient cohorts**

We next aimed to identify candidate protein markers for each NAFLD molecular subtype. To do so, we examined hepatic proteins that were upregulated only in one of the three molecular NAFLD groups (n=19 proteins across all groups) (Fig. 8A). Machine learning revealed that hepatic LAMB1, FLNC, and ERCC3 achieved the best area under the curve (AUC) for NAFLD mSI to mSIII classification in 3-fold cross-validation (Fig. 8B). Next, we performed multiplex immunofluorescence (IF) analysis of LAMB1, FLNC, and ERCC3 in human liver sections from all NAFLD molecular subtypes (Fig. 8C). Confusion matrix and ROC analyses of multiplex IF data revealed that the NAFLD molecular subtypes well characterized by multiplex IF of hepatic LAMB1, FLNC, and ERCC3 with high accuracy (92.98%) and an AUC of 0.991 (Fig. 8, D and E). Similar classification efficiency for NAFLD molecular subtypes was obtained using proteomics of hepatic 1346 proteins and multiplex IF of 3 liver proteins (Fig. 8F).

To examine the accuracy of multiplex IF analysis of hepatic protein biomarkers for characterizing NAFLD molecular subtypes, 92 patients with biopsy-proven NAFLD were recruited from three hospitals and served as an independent validation cohort (Fig. 8G). Machine learning analysis of multiplex IF data showed that there were 6, 6, and 13 patients in NAFLD-mSI; 3, 7, and 9 patients in NAFLD-mSII; and 10, 4, and 34 patients in NAFLD-mSIII in Hospitals 1, 2, and 3, respectively (Fig. 8G). LAMB1, FLNC and ERCC3 protein abundances were consistent with observations in the internal cohort (Fig. 8H). In addition, histological assessments of liver specimens including steatosis, ballooning, lobular inflammation, fibrosis score or severity, and NAS score, were similar to those of the internal cohort (Fig. 8, I and J, and table S12). Moreover, there were similar alterations in serum fasting insulin, HOMA-IR, ALT, AST, and BMI among these NAFLD molecular subtypes in the external cohort (Fig. S22, table S13 and 14) compared to the alterations seen in the internal cohort. Notably, multiplex IF analysis showed increased hepatic CYP3A4 in the NAFLD-mSI group, fibrotic LAMB1 and α-SMA in the NAFLD-mSII group, and HCC-related proteins ERCC3 and HNRNPA2B1 in the NAFLD-mSIII group, when compared to the other two groups (Fig. 8K).

**Machine learning identifies serum and urine potential biomarkers specific for NAFLD molecular subtypes**

To investigate potential biomarkers for the non-invasive identification of NAFLD molecular subtypes, we utilized molecules (including proteins, lipids, and metabolites) enriched in serum and urine samples of each subtype as candidate biomarkers. Using a Random Forest algorithm to distinguish between the NAFLD molecular subtypes, we found that a serum multi-omics panel consisting of 78 potential biomarkers had diagnostic accuracy in distinguishing between NAFLD molecular subtypes (fig. S23). The urine multi-omics panel of 5 candidate biomarkers also hadpotential in differentiating NAFLD molecular subtypes (fig. S24). Both panels hold promise as potential biomarkers for the noninvasive diagnosis of NAFLD molecular subtypes.

**DISCUSSION**

Using an integrative multi-omics analysis, we identified three distinct NAFLD molecular subtypes. NAFLD-mSI had high hepatic CYP1A2 and CYP3A4, which was induced by activating F FFA/BAs-mTOR-FXR/PPARα signaling and alleviate hepatic steatosis. NAFLD-mSII had a high risk for liver cirrhosis and increased hepatic infiltration of M1 and M2 macrophages due to the lipid-triggered hepatic CCL2 production, and NAFLD-mSIII had increased risk for HCC development as well as increased CEBPB- and ERCC3-dependent induction of oncogenes due to the activation of the EGF-EGFR/CHKA/PI3K-PDK1-AKT cascade.

NAFLD/NASH is becoming one of the leading causes of HCC worldwide (*1-2*). However, the existing NAFLD clinical, genetic, and metabolic classifications unsatisfactorily identify the risk of HCC development in people with NAFLD. Here, our integrative multi-omics analysis revealed that NAFLD-SIII was associated with an increased risk for HCC development. Previous studies have suggested that NASH-driven HCC may be typically associated with cirrhosis (*1-2*), but a substantial proportion of patients with NAFLD who lack cirrhosis or fibrosis may develop HCC nonetheless (*17*). Similarly, in our internal and external validation cohorts, 79% and 21% of patients in the NAFLD-mSIII group had either fibrosis or no fibrosis, respectively. Moreover, patients with NAFLD who also have metabolic syndrome features (for example, obesity and type 2 diabetes) and simple steatosis, but no NASH or fibrosis, can develop HCC (*42*). Particularly, a long-term study has proven that lean NAFLD can progress to HCC (*12*). There were also patients without obesity or diabetes in NAFLD-mSIII. Therefore, our findings and those of others support the notion that patients with NAFLD may develop HCC regardless of the presence of liver fibrosis, obesity, or type 2 diabetes, but rather due to a strong association with enhanced EGF-EGFR/CHKA/PI3K-PDK1-AKT signaling that leads to upregulation of CEPBP- and ERCC3-induced oncogenes in liver. Accordingly, targeting this axis at an early stage may prevent NAFLD progression to HCC, especially in those patients in NAFLD-mSIII.

Recent studies have shown that most patients with NAFLD do not achieve satisfactory histological outcomes with OCA or fenofibrate treatment (*6, 7*). In our study, three distinct NAFLD molecular subtypes were characterized and validated without any significant genetic differences (*P* >= 0.05) between them. NAFLD-mSI accounted for 33% of the analyzed patients with NAFLD and had higher liver CYP1A2 and CYP3A4, but lower FXR and PPARα. CYP1A2 or CYP3A4 silencing markedly increased the accumulation of PO-induced lipid droplets in hepatocytes, and hepatic CYP1A2 or CYP3A4 was negatively correlated with serum TG and patient hepatic steatosis scores , indicating that CYP1A2 and CYP3A4 may counter hepatic steatosis. Therefore, induction of FXR and PPARα can increase lipid accumulation by suppressing CYP1A2 and CYP3A4, impairing their anti-steatotic effects in patients in NAFLD-mSII and NAFLD-mSIII. Logically, treatment with OCA or fenofibrate to maintain appropriate FXR and PPARα activation may benefit patients in NAFLD-mSI. In addition, patients in NAFLD-mSII accounted for 20% of all studied patients, and showed increased hepatic fibrosis and lobular inflammation. It is possible that hepatic DGs and PAs may stimulate the production of CCL2 and CRP in hepatocytes and induce PKM2 in macrophages to attract monocytes and promote their differentiation into M1 and M2 macrophages (*28*), leading to enhanced liver inflammation and fibrosis. A recent study has shown that treatment with OCA, an FXR agonist, improves fibrosis and other histological features in patients with NASH (*5*). This suggests that whether treatment with agonists to maintain appropriate FXR and PPARα activity can improve histological outcomes in NASH may depend on the molecular subtype of liver disease.

Our study has several limitations. First, all participants in our study were of Chinese ethnicity, meaning the results may not be directly applicable to other racial populations without further validation. Second, the inclusion of a healthy lean cohort may not accurately reflect the typical obesity status observed in individuals with NAFLD, potentially limiting the generalizability of our findings. Last, our study was constrained by a small number of patients with cirrhosis. It’s important to note that individuals with liver cirrhosis were not excluded from the analysis, which could impact the interpretation of our results in this subgroup. These limitations underscore the need for caution when extrapolating our findings to broader population and highlight areas for future research and validation.

In conclusion, integrative multi-omics analysis, coupled with extensive experimental validation, has enabled the identification of three distinct molecular subtypes of NAFLD in humans. By targeting these molecular subtypes of NAFLD through precision medicine strategies, it may be possible to reduce imprecision in NAFLD diagnosis and pharmacotherapy, potentially preventing its progressing into cirrhosis and HCC in the early stages. Our findings may lead to a better understanding of the heterogeneity of human NAFLD, as well as to the discovery of targets for the diagnosis and treatment of three NAFLD molecular subtypes, at least in the Chinese population.

**MATERIALS AND METHODS**

**Study design**

This exploratory study aimed to define discrete NAFLD molecular subtypes, the pathogenesis of each subtype, and potential therapeutic targets. We consecutively recruited 103 adult patients with NAFLD who underwent liver biopsy at the First Affiliated Hospital of Wenzhou Medical University (Wenzhou, China). Their liver tissue samples were examined by two experienced liver pathologists using the NASH Clinical Research Network Scoring System in a blinded manner. Sample size was estimated based on a previous similar study (*43*). We first identified three distinct NAFLD molecular subtypes based on liver proteomics from 103 NAFLD patients.. We then conducted integrative multi-omics including WGS, proteomic, phosphoproteomic, lipidomic and metabolomic analyses of the liver, serum, and urine specimens from some NAFLD patients in the first step analysis (n=103), to determine the distinguished clinical features, molecular mechanisms, and omics features of each NAFLD molecular subtype. Last, candidate markers for the NAFLD molecular subtypes were identified by machine learning and validated in internal and external NAFLD cohorts. The internal cohort included 57 patients from the First Affiliated Hospital of Wenzhou Medical University (Wenzhou, China). The external cohort included a total of 92 NAFLD patients from the Guangdong Province Traditional Chinese Medical Hospital (n=19, Guangzhou, China), the Affiliated Hospital of Hangzhou Normal University (n=17, Hangzhou, China) and the Nanfang Hospital Affiliated to Southern Medical University (n=56, Guangzhou, China).

All studies using human samples were approved by each participating center of the First Affiliated Hospital of Wenzhou Medical University (Approval No: 2016-246 ), the First Affiliated Hospital (Southwest Hospital) of Army Medical University (Approval No: (B)KY2023050), the Guangdong Province Traditional Chinese Medical Hospital (Approval No: Z2017-001-01), the Affiliated Hospital of Hangzhou Normal University (Approval No: 2021 (E2)-HS-040), the Nanfang Hospital Affiliated to Southern Medical University (Approval No: NFEC-2017-143) and the Chengdu Medical College (Approval No: 2023NO.100, AF-SOP-37-2.1).

**Statistical Analysis**

PCA was carried out using the R program. Continuous variables were analyzed by the independent-sample Student’s T-test, Mann-Whitney U test, one-way ANOVA and post hoc LSD or Tambane's T2 post hoc tests, or Kruskal-Wallis H-tests where applicable. Categorical variables were analyzed using Pearson's Chi-squared test or Fisher’s exact test, corrected by the Bonferroni method. Correlations were analyzed by Spearman correlation analysis, whereas associations between the continual variables and ordered categorical data were analyzed by the Kendall tau test. All statistical analyses were performed using SPSS. A two-tailed *P-value of* < 0.05 was considered statistically significant.

**Supplementary Materials**

This PDF file includes:

Materials and Methods

Figs. S1 to S24

Tables S1 to S14

References (*44–49*)

**Other Supplementary Material for this manuscript includes the following:**

Data file S1

Data file S2

MDAR Reproducibility Checklist

**REFERENCES AND NOTES**

1. Z. Younossi, Q. M. Anstee, M. Marietti, T. Hardy, L. Henry, M. Eslam, J. George, E. Bugianesi, Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol* **15**, 11-20 (2018).

2. G. Cholankeril, J. R. Kramer, J. Chu, X. Yu, M. Balakrishnan, L. Li, H. B. El-Serag, F. Kanwal, Longitudinal changes in fibrosis markers are associated with risk of cirrhosis and hepatocellular carcinoma in non-alcoholic fatty liver disease. *J Hepatol* **78**, 493-500 (2023).

3. H. Ledford, First US drug approved for a liver disease surging around the world. *Nature*, (2024).

4. S. A. Harrison, P. Bedossa, C. D. Guy, J. M. Schattenberg, R. Loomba, R. Taub, D. Labriola, S. E. Moussa, G. W. Neff, M. E. Rinella, Q. M. Anstee, M. F. Abdelmalek, Z. Younossi, S. J. Baum, S. Francque, M. R. Charlton, P. N. Newsome, N. Lanthier, I. Schiefke, A. Mangia, J. M. Pericas, R. Patil, A. J. Sanyal, M. Noureddin, M. B. Bansal, N. Alkhouri, L. Castera, M. Rudraraju, V. Ratziu, M.-N. Investigators, A Phase 3, Randomized, Controlled Trial of Resmetirom in NASH with Liver Fibrosis. *N Engl J Med* **390**, 497-509 (2024).

5. L. G. Di Pasqua, M. Cagna, G. Palladini, A. C. Croce, M. Cadamuro, L. Fabris, S. Perlini, L. Adorini, A. Ferrigno, M. Vairetti, FXR agonists INT-787 and OCA increase RECK and inhibit liver steatosis and inflammation in diet-induced ob/ob mouse model of NASH. *Liver Int* **44**, 214-227 (2024).

6. X. X. Wang, J. Luo, Z. H. Lu, S. Z. Fang, M. X. Sun, W. J. Luo, J. W. Shen, A. M. Liu, H. Ye, Therapeutic effect of fenofibrate for non-alcoholic steatohepatitis in mouse models is dependent on regime design. *Front Pharmacol* **14**, (2023).

7. K. Wang, Y. C. Zhang, G. J. Wang, H. P. Hao, H. Wang, FXR agonists for MASH therapy: Lessons and perspectives from obeticholic acid. *Med Res Rev* **44**, 568-586 (2024).

8. A. Mahmoudi, T. Jamialahmadi, T. P. Johnston, A. Sahebkar, Impact of fenofibrate on NAFLD/NASH: A genetic perspective. *Drug Discov Today* **27**, 2363-2372 (2022).

9. M. Eslam, A. J. Sanyal, J. George, P. International Consensus, MAFLD: A Consensus-Driven Proposed Nomenclature for Metabolic Associated Fatty Liver Disease. *Gastroenterology* **158**, 1999-2014 e1991 (2020).

10. M. T. Long, M. Noureddin, J. K. Lim, AGA Clinical Practice Update: Diagnosis and Management of Nonalcoholic Fatty Liver Disease in Lean Individuals: Expert Review. *Gastroenterology* **163**, 764-774 e761 (2022).

11. F. Chen, S. Esmaili, G. B. Rogers, E. Bugianesi, S. Petta, G. Marchesini, A. Bayoumi, M. Metwally, M. K. Azardaryany, S. Coulter, J. M. Choo, R. Younes, C. Rosso, C. Liddle, L. A. Adams, A. Craxi, J. George, M. Eslam, Lean NAFLD: A Distinct Entity Shaped by Differential Metabolic Adaptation. *Hepatology* **71**, 1213-1227 (2020).

12. A. M. C. Navarroza, S. N. Wong, Comparison of clinical and metabolic profiles of lean versus non-lean nonalcoholic fatty liver disease. *Indian J Gastroenterol* **40**, 380-388 (2021).

13. S. Lallukka, K. Sevastianova, J. Perttila, A. Hakkarainen, M. Orho-Melander, N. Lundbom, V. M. Olkkonen, H. Yki-Jarvinen, Adipose tissue is inflamed in NAFLD due to obesity but not in NAFLD due to genetic variation in PNPLA3. *Diabetologia* **56**, 886-892 (2013).

14. M. Longo, M. Meroni, E. Paolini, V. Erconi, F. Carli, F. Fortunato, D. Ronchi, R. Piciotti, S. Sabatini, C. Macchi, A. Alisi, L. Miele, G. Soardo, G. P. Comi, L. Valenti, M. Ruscica, A. L. Fracanzani, A. Gastaldelli, P. Dongiovanni, TM6SF2/PNPLA3/MBOAT7 Loss-of-Function Genetic Variants Impact on NAFLD Development and Progression Both in Patients and in In Vitro Models. *Cell Mol Gastroenterol Hepatol* **13**, 759-788 (2022).

15. M. Masoodi, A. Gastaldelli, T. Hyotylainen, E. Arretxe, C. Alonso, M. Gaggini, J. Brosnan, Q. M. Anstee, O. Millet, P. Ortiz, J. M. Mato, J. F. Dufour, M. Oresic, Metabolomics and lipidomics in NAFLD: biomarkers and non-invasive diagnostic tests. *Nat Rev Gastroenterol Hepatol* **18**, 835-856 (2021).

16. I. Martinez-Arranz, C. Bruzzone, M. Noureddin, R. Gil-Redondo, I. Minchole, M. Bizkarguenaga, E. Arretxe, M. Iruarrizaga-Lejarreta, D. Fernandez-Ramos, F. Lopitz-Otsoa, R. Mayo, N. Embade, E. Newberry, B. Mittendorf, L. Izquierdo-Sanchez, V. Smid, J. Arnold, P. Iruzubieta, Y. Perez Castano, M. Krawczyk, U. M. Marigorta, M. C. Morrison, R. Kleemann, A. Martin-Duce, L. Hayardeny, L. Vitek, R. Bruha, R. Aller de la Fuente, J. Crespo, M. Romero-Gomez, J. M. Banales, M. Arrese, K. Cusi, E. Bugianesi, S. Klein, S. C. Lu, Q. M. Anstee, O. Millet, N. O. Davidson, C. Alonso, J. M. Mato, Metabolic subtypes of patients with NAFLD exhibit distinctive cardiovascular risk profiles. *Hepatology* **76**, 1121-1134 (2022).

17. S. Mittal, H. B. El-Serag, Y. H. Sada, F. Kanwal, Z. Duan, S. Temple, S. B. May, J. R. Kramer, P. A. Richardson, J. A. Davila, Hepatocellular Carcinoma in the Absence of Cirrhosis in United States Veterans is Associated With Nonalcoholic Fatty Liver Disease. *Clin Gastroenterol Hepatol* **14**, 124-131 e121 (2016).

18. F. Gao, J. F. Huang, K. I. Zheng, X. Y. Pan, H. L. Ma, W. Y. Liu, C. D. Byrne, G. Targher, Y. Y. Li, Y. P. Chen, W. K. Chan, M. H. Zheng, Development and validation of a novel non-invasive test for diagnosing fibrotic non-alcoholic steatohepatitis in patients with biopsy-proven non-alcoholic fatty liver disease. *J Gastroenterol Hepatol* **35**, 1804-1812 (2020).

19. O. Govaere, S. Cockell, D. Tiniakos, R. Queen, R. Younes, M. Vacca, L. Alexander, F. Ravaioli, J. Palmer, S. Petta, J. Boursier, C. Rosso, K. Johnson, K. Wonders, C. P. Day, M. Ekstedt, M. Oresic, R. Darlay, H. J. Cordell, F. Marra, A. Vidal-Puig, P. Bedossa, J. M. Schattenberg, K. Clement, M. Allison, E. Bugianesi, V. Ratziu, A. K. Daly, Q. M. Anstee, Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci Transl Med* **12**, (2020).

20. K. Chella Krishnan, Z. Kurt, R. Barrere-Cain, S. Sabir, A. Das, R. Floyd, L. Vergnes, Y. Zhao, N. Che, S. Charugundla, H. Qi, Z. Zhou, Y. Meng, C. Pan, M. M. Seldin, F. Norheim, S. Hui, K. Reue, A. J. Lusis, X. Yang, Integration of Multi-omics Data from Mouse Diversity Panel Highlights Mitochondrial Dysfunction in Non-alcoholic Fatty Liver Disease. *Cell Syst* **6**, 103-115 e107 (2018).

21. S. Tsurusaki, Y. Tsuchiya, T. Koumura, M. Nakasone, T. Sakamoto, M. Matsuoka, H. Imai, C. Yuet-Yin Kok, H. Okochi, H. Nakano, A. Miyajima, M. Tanaka, Hepatic ferroptosis plays an important role as the trigger for initiating inflammation in nonalcoholic steatohepatitis. *Cell Death Dis* **10**, 449 (2019).

22. S. Huang, C. Sun, Y. Hou, Y. Tang, Z. Zhu, Z. Zhang, Y. Zhang, L. Wang, Q. Zhao, M. G. Chen, Z. Guo, D. Wang, W. Ju, Q. Zhou, L. Wu, X. He, A comprehensive bioinformatics analysis on multiple Gene Expression Omnibus datasets of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. *Sci Rep* **8**, 7630 (2018).

23. S. Zhang, X. Pan, H. Jeong, GW4064, an agonist of farnesoid X receptor, represses CYP3A4 expression in human hepatocytes by inducing small heterodimer partner expression. *Drug Metab Dispos* **43**, 743-748 (2015).

24. X. Hu, X. Jia, C. Xu, Y. Wei, Z. Wang, G. Liu, Q. You, G. Lu, W. Gong, Downregulation of NK cell activities in Apolipoprotein C-III-induced hyperlipidemia resulting from lipid-induced metabolic reprogramming and crosstalk with lipid-laden dendritic cells. *Metabolism* **120**, 154800 (2021).

25. W. Chen, X. Wu, X. Yan, A. Xu, A. Yang, H. You, Multitranscriptome analyses reveal prioritized genes specifically associated with liver fibrosis progression independent of etiology. *Am J Physiol Gastrointest Liver Physiol* **316**, G744-G754 (2019).

26. Z. Y. Wang, A. Keogh, A. Waldt, R. Cuttat, M. Neri, S. Zhu, S. Schuierer, A. Ruchti, C. Crochemore, J. Knehr, J. Bastien, I. Ksiazek, D. Sanchez-Taltavull, H. Ge, J. Wu, G. Roma, S. B. Helliwell, D. Stroka, F. Nigsch, Single-cell and bulk transcriptomics of the liver reveals potential targets of NASH with fibrosis. *Sci Rep* **11**, 19396 (2021).

27. H. Zhang, F. Chen, X. Fan, C. Lin, Y. Hao, H. Wei, W. Lin, Y. Jiang, F. He, Quantitative Proteomic analysis on Activated Hepatic Stellate Cells reversion Reveal STAT1 as a key regulator between Liver Fibrosis and recovery. *Sci Rep* **7**, 44910 (2017).

28. S. Devaraj, I. Jialal, C-reactive protein polarizes human macrophages to an M1 phenotype and inhibits transformation to the M2 phenotype. *Arterioscler Thromb Vasc Biol* **31**, 1397-1402 (2011).

29. F. Xu, M. Guo, W. Huang, L. Feng, J. Zhu, K. Luo, J. Gao, B. Zheng, L. D. Kong, T. Pang, X. Wu, Q. Xu, Annexin A5 regulates hepatic macrophage polarization via directly targeting PKM2 and ameliorates NASH. *Redox Biol* **36**, 101634 (2020).

30. J. Ahodantin, K. Nio, M. Funaki, X. Zhai, E. Wilson, S. Kottilil, L. Cheng, G. Li, L. Su, Type I interferons and TGF-beta cooperate to induce liver fibrosis during HIV-1 infection under antiretroviral therapy. *JCI Insight* **7**, (2022).

31. B. Saha, K. Kodys, G. Szabo, Hepatitis C Virus-Induced Monocyte Differentiation Into Polarized M2 Macrophages Promotes Stellate Cell Activation via TGF-beta. *Cell Mol Gastroenterol Hepatol* **2**, 302-316 e308 (2016).

32. J. Gao, C. Dai, X. Yu, X. B. Yin, F. Zhou, LncRNA LEF1-AS1 silencing diminishes EZH2 expression to delay hepatocellular carcinoma development by impairing CEBPB-interaction with CDCA7. *Cell Cycle* **19**, 870-883 (2020).

33. J. Zhang, J. Z. Huang, Y. Q. Zhang, X. Zhang, L. Y. Zhao, C. G. Li, Y. F. Zhou, H. Wei, J. Yu, Microtubule associated protein 9 inhibits liver tumorigenesis by suppressing ERCC3. *EBioMedicine* **53**, 102701 (2020).

34. S. Mehdi, M. Bachvarova, M. P. Scott-Boyer, A. Droit, D. Bachvarov, LY75 Ablation Mediates Mesenchymal-Epithelial Transition (MET) in Epithelial Ovarian Cancer (EOC) Cells Associated with DNA Methylation Alterations and Suppression of the Wnt/beta-Catenin Pathway. *Int J Mol Sci* **21**, (2020).

35. Z. Zeng, S. Lei, J. Wang, Y. Yang, J. Lan, Q. Tian, T. Chen, X. Hao, A novel hypoxia-driven gene signature that can predict the prognosis of hepatocellular carcinoma. *Bioengineered* **13**, 12193-12210 (2022).

36. Y. Yin, X. Xu, J. Tang, W. Zhang, G. Zhangyuan, J. Ji, L. Deng, S. Lu, H. Zhuo, B. Sun, CD97 Promotes Tumor Aggressiveness Through the Traditional G Protein-Coupled Receptor-Mediated Signaling in Hepatocellular Carcinoma. *Hepatology* **68**, 1865-1878 (2018).

37. K. Kumar, S. Malhotra, A. Sibal, Transient infantile hypertriglyceridaemia due to homozygous mutation in GPD1 presenting in childhood with hepatic adenoma. *BMJ Case Rep* **15**, (2022).

38. Y. Lu, X. Wang, Q. Gu, J. Wang, Y. Sui, J. Wu, J. Feng, Heterogeneous nuclear ribonucleoprotein A/B: an emerging group of cancer biomarkers and therapeutic targets. *Cell Death Discov* **8**, 337 (2022).

39. X. M. Lin, L. Hu, J. Gu, R. Y. Wang, L. Li, J. Tang, B. H. Zhang, X. Z. Yan, Y. J. Zhu, C. L. Hu, W. P. Zhou, S. Li, J. F. Liu, F. J. Gonzalez, M. C. Wu, H. Y. Wang, L. Chen, Choline Kinase alpha Mediates Interactions Between the Epidermal Growth Factor Receptor and Mechanistic Target of Rapamycin Complex 2 in Hepatocellular Carcinoma Cells to Promote Drug Resistance and Xenograft Tumor Progression. *Gastroenterology* **152**, 1187-1202 (2017).

40. P. N. Duchateau, I. Movsesyan, S. Yamashita, N. Sakai, K. Hirano, S. A. Schoenhaus, P. M. O'Connor-Kearns, S. J. Spencer, R. B. Jaffe, R. F. Redberg, B. Y. Ishida, Y. Matsuzawa, J. P. Kane, M. J. Malloy, Plasma apolipoprotein L concentrations correlate with plasma triglycerides and cholesterol levels in normolipidemic, hyperlipidemic, and diabetic subjects. *J Lipid Res* **41**, 1231-1236 (2000).

41. A. M. Bofin, B. Ytterhus, E. Klaestad, M. Valla, FGFR1 copy number in breast cancer: associations with proliferation, histopathological grade and molecular subtypes. *J Clin Pathol* **75**, 459-464 (2022).

42. M. Grohmann, F. Wiede, G. T. Dodd, E. N. Gurzov, G. J. Ooi, T. Butt, A. A. Rasmiena, S. Kaur, T. Gulati, P. K. Goh, A. E. Treloar, S. Archer, W. A. Brown, M. Muller, M. J. Watt, O. Ohara, C. A. McLean, T. Tiganis, Obesity Drives STAT-1-Dependent NASH and STAT-3-Dependent HCC. *Cell* **175**, 1289-1306 e1220 (2018).

43. J. Y. Xu, C. Zhang, X. Wang, L. Zhai, Y. Ma, Y. Mao, K. Qian, C. Sun, Z. Liu, S. Jiang, M. Wang, L. Feng, L. Zhao, P. Liu, B. Wang, X. Zhao, H. Xie, X. Yang, L. Zhao, Y. Chang, J. Jia, X. Wang, Y. Zhang, Y. Wang, Y. Yang, Z. Wu, L. Yang, B. Liu, T. Zhao, S. Ren, A. Sun, Y. Zhao, W. Ying, F. Wang, G. Wang, Y. Zhang, S. Cheng, J. Qin, X. Qian, Y. Wang, J. Li, F. He, T. Xiao, M. Tan, Integrative Proteomic Characterization of Human Lung Adenocarcinoma. *Cell* **182**, 245-261 e217 (2020).

44. Y. Jiang, A. Sun, Y. Zhao, W. Ying, H. Sun, X. Yang, B. Xing, W. Sun, L. Ren, B. Hu, C. Li, L. Zhang, G. Qin, M. Zhang, N. Chen, M. Zhang, Y. Huang, J. Zhou, Y. Zhao, M. Liu, X. Zhu, Y. Qiu, Y. Sun, C. Huang, M. Yan, M. Wang, W. Liu, F. Tian, H. Xu, J. Zhou, Z. Wu, T. Shi, W. Zhu, J. Qin, L. Xie, J. Fan, X. Qian, F. He, C. Chinese Human Proteome Project, Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* **567**, 257-261 (2019).

45. A. Borrelli, P. Bonelli, F. M. Tuccillo, I. D. Goldfine, J. L. Evans, F. M. Buonaguro, A. Mancini, Role of gut microbiota and oxidative stress in the progression of non-alcoholic fatty liver disease to hepatocarcinoma: Current and innovative therapeutic approaches. *Redox Biol* **15**, 467-479 (2018).

46. J. P. Brunet, P. Tamayo, T. R. Golub, J. P. Mesirov, Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A* **101**, 4164-4169 (2004).

47. Q. Pan, X. Zhang, L. Zhang, Y. Cheng, N. Zhao, F. Li, X. Zhou, S. Chen, J. Li, S. Xu, D. Huang, Y. Chen, L. Li, H. Wang, W. Chen, S. Y. Cai, J. L. Boyer, J. Chai, Solute Carrier Organic Anion Transporter Family Member 3A1 Is a Bile Acid Efflux Transporter in Cholestasis. *Gastroenterology* **155**, 1578-1592 e1516 (2018).

48. L. Zhang, Q. Pan, L. Zhang, H. Xia, J. Liao, X. Zhang, N. Zhao, Q. Xie, M. Liao, Y. Tan, Q. Li, J. Zhu, L. Li, S. Fan, J. Li, C. Zhang, S. Y. Cai, J. L. Boyer, J. Chai, Runt-related transcription factor-1 ameliorates bile acid-induced hepatic inflammation in cholestasis through JAK/STAT3 signaling. *Hepatology* **77**, 1866-1881 (2023).

49. N. Zhao, X. Zhang, J. Ding, Q. Pan, M. H. Zheng, W. Y. Liu, G. Luo, J. Qu, M. Li, L. Li, Y. Cheng, Y. Peng, Q. Xie, Q. Wei, Q. Li, L. Zou, X. Ouyang, S. Y. Cai, J. L. Boyer, J. Chai, SEMA7AR148W mutation promotes lipid accumulation and NAFLD progression via increased localization on the hepatocyte surface. *JCI Insight* **7**, (2022).

**Acknowledgements:** We thank Professor Na Liu (Sun Yat-sen University Cancer Center), Professor. Min-Dian Li (Army Medical University, China) and Hongming Miao (Army Medical University) and our team members (Army Medical University, China) for technical assistance. We also thank Beijing Genomics Institute (Shenzhen, China), Jingjie PTM BioLab (Hangzhou, China) and Metware Biotechnology (Wuhan, China) for their technical support regarding WGS, proteomics, phosphoproteomics, lipidomics, and metabolomics, respectively.

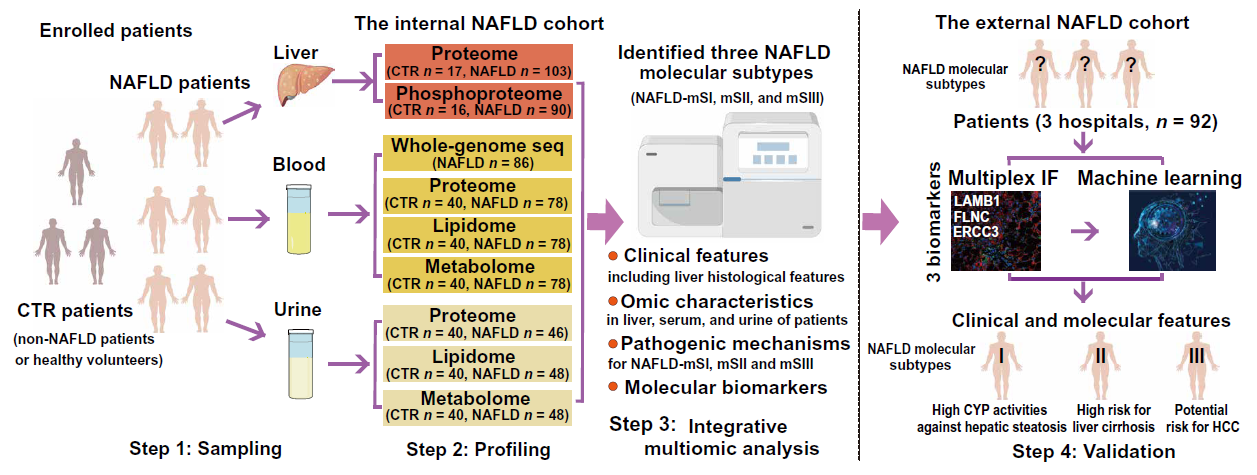
**Funding**: This work was supported by the grants from the National Natural Science Foundation of China (82325008, 92268110), the Outstanding Youth Foundation of Chongqing (cstc2021jcyj-jqX0005), the Project of Chongqing / University Innovation Research Group (2021cqspt01/414Z381), and the Scientific Foundation of Army Medical University (XZ-2019-505-001) to J.C.; the Sichuan Provincial Natural Science Foundation project (2022NSFSC0769) to Z.Y.J.; the School of Medicine, University of Verona, Verona, Italy to G.T.; the Southampton NIHR Biomedical Research Centre (NIHR203319) to C.D.B.; and the National Institutes of Health (NIH) R01 grant (R01DK134549) to X.O.

**Author contributions**: J.C. conceived and designed the study. J.J.D., X.X.Z., N.Z. and Q.Y. performed Western blot, Pull-down assay and data analysis. J.J.D and L.L. performed Multiplex IF staining and data analysis. M.N.Z., J.F.O.Y., and M.L. performed transwell migration and invasion assays. M.Q.L. and Z.Q.X. performed Oil Red O staining and Real-time qPCR. L.J.Z. and Y.T performed ELISA analysis. W.T., Q.P. and N.Z. performed preparation and culture of primary mouse hepatocytes and cell culture experiments. Y.P. collected clinical data of control individuals from the First Affiliated Hospital of Wenzhou Medical University (Wenzhou, China) and Southwest Hospital (Chongqing, China). X.L.C., H.Z.L., W.Y.L., and Y.L. collected clinical data of NAFLD patients from the First Affiliated Hospital of Wenzhou Medical University (Wenzhou, China), the Guangdong Province Traditional Chinese Medical Hospital (Guangzhou, China), the Affiliated Hospital of Hangzhou Normal University (Hangzhou, China) and the Nanfang Hospital Affiliated to Southern Medical University (Guangzhou, China), respectively. X.O., L.Q.Z, Z.Y.J, F.J.S, J.P.S, J.J.C, Y.X，S.Y.C, C.D.B, and G.T provided critical reagents/analysis tools; J.C, M.H.Z, and J.J.D wrote the manuscript and created the figures. C.M.X, X.O., S.Y.C, C.D.B, and G.T provided their critical revision.

**Competing interests**: The authors declare no conflicts of interest. J.C. and Z.Y.J are coinventors of patent pending titled “Application of a biomarker in the preparation of a product for the classification of NAFLD” (application number: 202410674429.2).

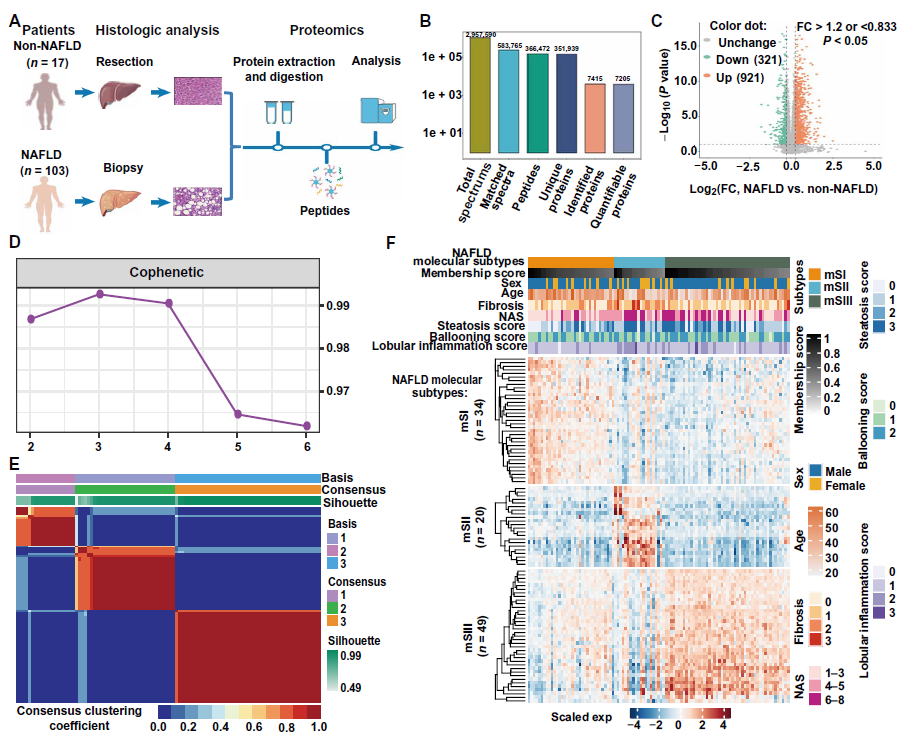
**Data and materials availability**: All data associated with this study are in the paper or supplementary materials. Source data are in data file S1. The raw data from proteomics, phosphoproteomics, lipidomics and metabolomics, and WGS data involved in this study have been deposited in the OMIX, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/omix) with accession codes OMIX003061, OMIX003063, OMIX002402, OMIX002389, OMIX002387, OMIX002423, OMIX002400, OMIX002401 and OMIX003087.

**Figure legends**

****

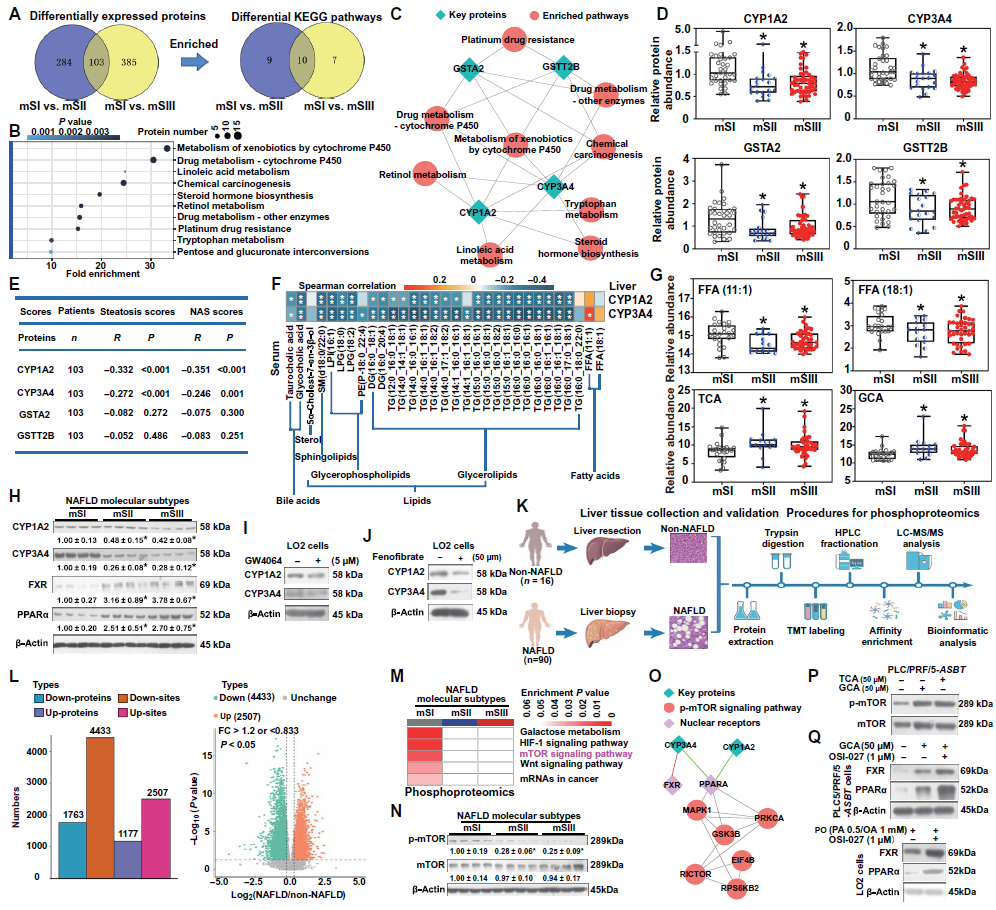
**Fig. 1. Experimental design of analysis in internal and external NAFLD cohorts**.

Multi-omics analysis was performed on the liver, serum, and urine of NAFLD patients in the internal cohort. The control (CTR) patients including non-NAFLD patients and healthy volunteers. Proteome and phosphoproteome analyses were performed on liver tissues of non-NAFLD patients in the internal cohort. Proteome, lipidome and metabolome analyses were performed on serum and urine samples of healthy volunteers in the internal cohort. Three distinct molecular subtypes of NAFLD were identified and their pathogenic mechanisms, clinical or multi-omic characteristics and molecular biomarkers were clarified. Multiplex IF was performed to validate the feasibility of molecular biomarkers in the diagnosis of three NAFLD molecular subtypes in internal and external cohorts.

****

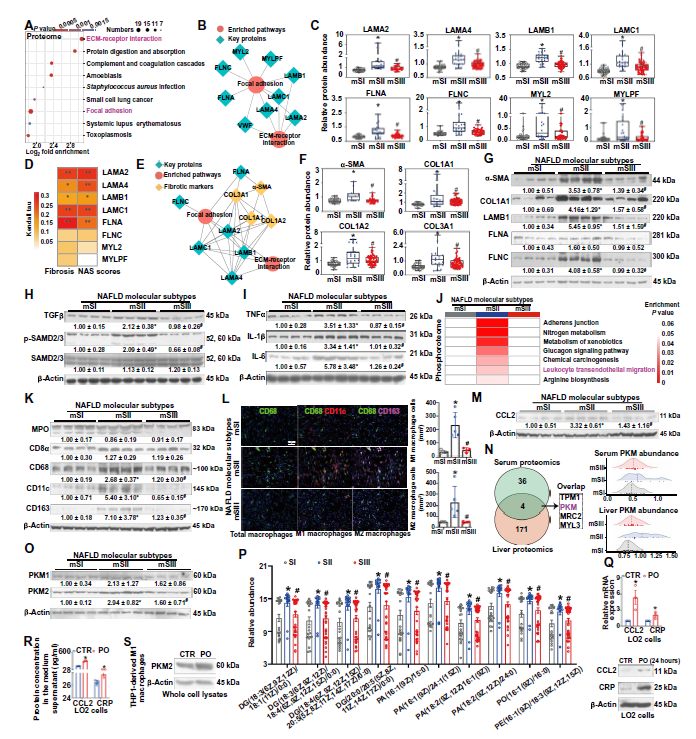
**Fig. 2. Proteomics of liver samples from patients with biopsy-proven NAFLD identifies three typical NAFLD molecular subtypes.**

(**A**) Illustration of liver proteomic analysis in Non-NAFLD patients (*n*=17) and NAFLD (*n*=103). (**B**) Basic information on hepatic proteomic data. (**C**) Volcano plot of all quantified proteins. (**D**) Relationship between cophenetic correlation coefficient and factorization ranks. (**E**) Heatmap of consensus matrix for rank=3. (**F**) Heatmap of differentially expressed proteins and clinical features among NAFLD molecular subtypes.

****

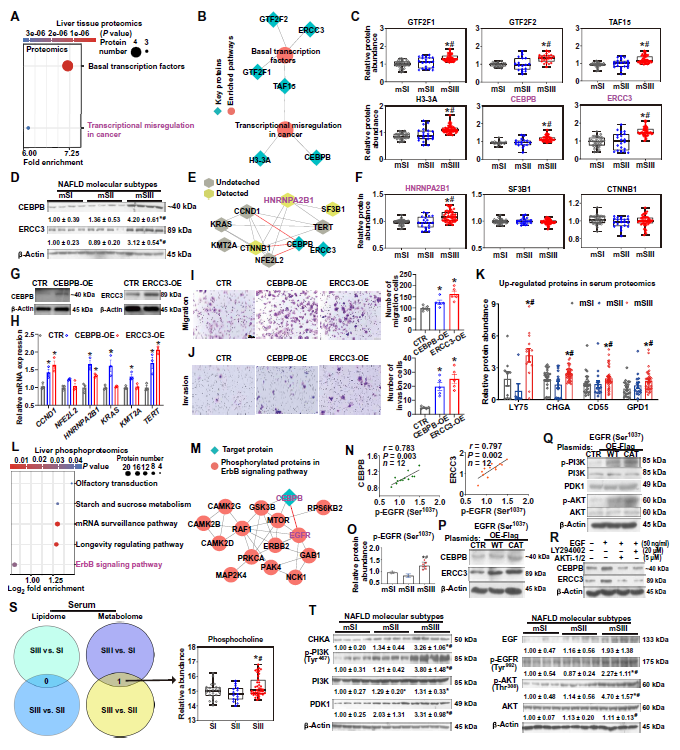
**Fig. 3. Molecular mechanisms underlying pathogenesis in NAFLD-mSI.**

(**A to B**) Differentially expressed proteins (DEPs; **A**, left) in NAFLD-mSI *vs.* NAFLD-mSII and mSIII (FC > 1.2; *P* < 0.05), and KEGG pathways enriched in the up-regulated DEPs (**A** right and **B**). (**C** and **D**) Four key proteins were enriched in 10 KEGG pathways, and their proteins were compared across the three NAFLD molecular subtypes. (**E**) Correlations between key hepatic proteins and liver steatosis or NAS scores for all patients. (**F**) Spearman correlation analyses among CYP1A2, CYP3A4, key serum lipids and metabolites in NAFLD-mSI (\**P* < 0.05). (**G**) Serum FFAs and conjugated BAs in the three NAFLD molecular subtypes. (**H**) Western-blotting analysis of hepatic CYP1A2, CYP3A4, FXR, and PPARα in the three NAFLD molecular subtypes, *n*=8 for each group. (**I and J**) Effect of treatment with GW4064 or fenofibrate on CYP1A2 and CYP3A4 in LO2 cells. (**K**) An illustration of liver phosphoproteomics. (**L**) Basic information of phosphoproteomic data. (**M**) KEGG pathways based on hepatic phosphoproteomics enriched in NAFLD-mSI (*P* < 0.05). (**N**) Hepatic phospho-mTOR/mTOR proteins in the three NAFLD molecular subtypes, *n*=8 for each group. (**O**) PPI network constructed using key proteins and phosphoproteins enriched in the mTOR signaling pathway, FXR, and PPARα. (**P**) TCA and GCA activation of mTOR signaling in PLC/RPF/5-*ASBT* cells. (**Q**) Effect of OSI-027 treatment on FXR and PPARα protein abundance in PLC/PRF/5-*ASBT* cells after 50 μM GCA treatment or in LO2 cells after PO treatment. Welch t-test was used in (D and G), and one-way ANOVA in (H and N). \**P* < 0.05 *vs*. NAFLD-mSI.



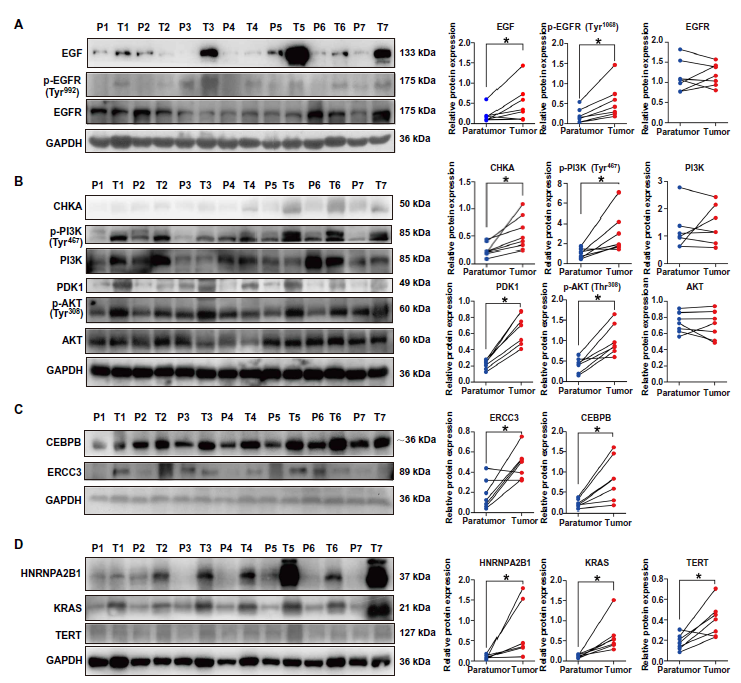
**Fig. 4. Molecular mechanisms underlying pathogenesis in NAFLD-mSII.**

(**A**) KEGG pathways enriched in NAFLD-mSII, based on the DEPs. (**B** and **C**) proteins enriched in the focal adhesion and ECM-receptor interaction pathways (B) and their protein abundance in three NAFLD molecular subtypes (C). (**D**) Correlations among hepatic 8 key proteins and fibrosis (left) or NAS score (right). (**E**) The protein-protein interaction (PPI) network of the key proteins of NAFLD-mSII (LAMA2, LAMA4, LAMB1, LAMC1, FLNA and FLNC) and fibrosis markers (α-SMA, COL1A1, COL1A2, and COL3A1) was analysed using STRING. (**F**) Hepatic α-SMA, COL1A1, COL1A2 and COL3A4 in the three NAFLD molecular subtypes. (**G** to **I**) Western blotting analysis of hepatic α-SMA, COL1A1, LAMB1, FLNA, FLNC, TGFβ, phospho-SMAD2/3, SMAD2/3, TNFα, IL1-β and IL-6 in three NAFLD molecular subtypes, *n*=8 per group. (**J**) KEGG pathways enriched in NAFLD-mSII based on phosphoproteomics (*P* < 0.05). (**K**) Hepatic markers for neutrophils, CD8+ T cells, and macrophages, *n*=8 per group. (**L**) Multiplex IF analysis of CD68, CD11c and CD163 in liver sections from patients in NAFLD-mSI-III (*n*=5, 5, 4, respectively). (**M**) Hepatic CCL2 proteins in each NAFLD molecular subtype, *n*=8 per group. (**N**) The overlap proteins of liver and serum which were differentially expressed in NAFLD-mSII vs. NAFLD-mSI and mSIII (left), and serum and liver PKM proteins in three NAFLD molecular subtype (right). (**O**) Hepatic PKM1 and PKM2 expression in the NAFLD molecular subtypes, *n*=8 per group. (**P**) Significantly (*P* < 0.05) elevated serum metabolites (Top10) in NAFLD-mSII *vs.* NAFLD-mSI and m-SIII. (**Q to R**) Abundance of *CCL2* and *CRP* transcripts and protein in LO2 cells (**Q**) and ELISA analysis of CCL2 and CRP supernatants of cultured LO2 cells (**R**) after PO treatment. (**S**) PKM2 protein in THP1-derived M1 macrophages after PO treatmentWelch t-test was used in (C, F, P, Q and R), and one-way ANOVA was used in (G, H, I, K, L, M and O). \**P* < 0.05 *vs*. CTR or NAFLD-mSI; #*P* < 0.05 *vs*. NAFLD-mSII.



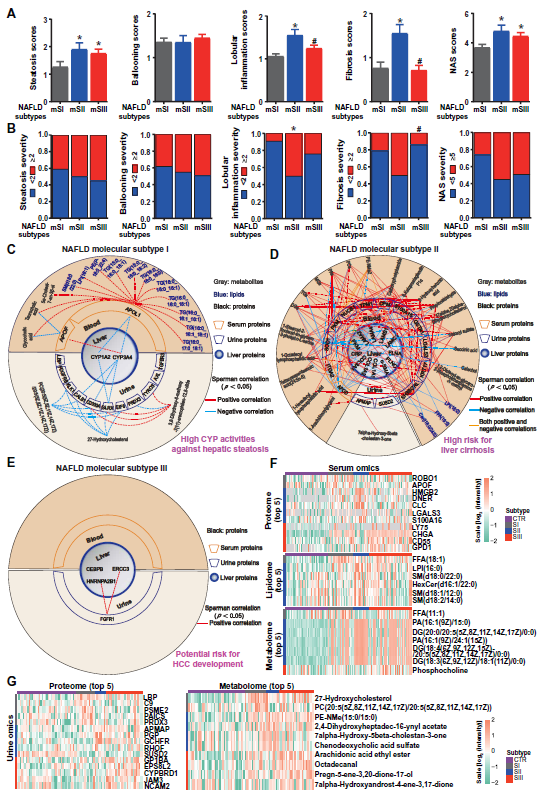
**Fig. 5. Molecular mechanisms underlying pathogenesis in NAFLD-mSIII**

(A) KEGG pathway analysis revealed that the differentially expressed proteins in hepatic proteomics were significantly (P < 0.05) enriched in NAFLD-mSIII vs NAFLD-mSI and mSII. (B) A network diagram of the differentially expressed proteins enriched in basal transcription factors and transcriptional misregulation in cancer pathways. (C) The levels of hepatic GTF2F1, GTF2F2, TAF15, H3-3A, CEBPB and ERCC3 proteins in three NAFLD molecular subtypes. (**D**) Hepatic CEBPB and ERCC3 in each NAFLD molecular subtyps, *n*=8 per group. (**E**) PPI analysis of interactions among CEBPB and ERCC3 and HCC-related oncogenes. (**F**) Hepatic oncogenes in each NAFLD molecular subtype. CEBPB or ERCC3 (**G**), oncogene transcripts (**H**), and transwell migration (**I**) and invasion assays (**J**) in Huh7 cells after induction of CEBPB or ERCC3. (**K**) Elevated proteins in serum proteomics of NAFLD-mSIII *vs.* NAFLD-mSI and mSII. (**L**) KEGG pathway enriched in NAFLD-mSIII based on phosphoproteomics. (**M**) The relationships between CEBPB and phosphorylated proteins enriched in ErbB signaling. (**N**) Hepatic phospho-EGFR (Ser 1037) positively correlated with CEBPB or ERCC3 in patients with NAFLD. (**O**) Hepatic phospho-EGFR (Ser 1037) in each NAFLD molecular subtype. (**P** and **Q**) Western-blotting analysis CEBPB, ERCC3, PI3K-AKT cascade in Huh7 cells after transfection with EGFR (Ser 1037)-WT or -CAT construct. (**R**) CEBPB and ERCC3 in Huh7 cells after treatment with LY294002, AKTi-1/2, or EGF treatment. (**S**) The differential metabolites or lipids between NAFLD-mSIII and the other two subtypes in serum were analyzed. Phosphatidylcholine was the only metabolite in serum that was different between NAFLD-mSIII and the other two subtypes. Serum phosphocholine abundance in each NAFLD molecular subtype. (**T**) Western blot analysis of the EGF-EGFR/CHKA/PI3K-PDK1-AKT cascade in the liver of each NAFLD subtype , *n*=8 per group. Welch t-test in (C, F, K, O and S), and one-way ANOVA in (D, H, I, J and T). \**P* < 0.05 *vs*. CTR or NAFLD-mSI; #*P* < 0.05 *vs*. NAFLD-mSII.



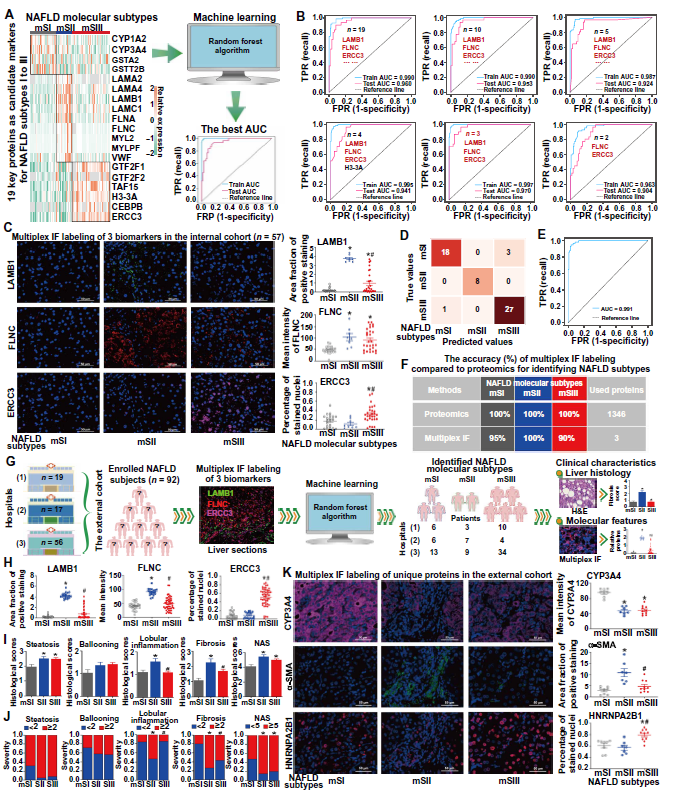
**Fig. 6. Molecular mechanisms underlying NAFLD-mSIII pathogenesis are verified in liver tissues of patients with HCC.**

Hepatic EGF, EGFR (Tyr 992), EGFR (**A**), CHKA, PI3K (Tyr 467), PI3K, PDK1, AKT (Thr 308), AKT (**B**), CEBPB, RECC3 (**C**), and HNRNPA2B1, KRAS, and TERT (**D**) in tumor and para-tumor tissues from patients HCC analyzed by Western blotting. Paired data were analyzed by paired t-test. \**P* < 0.05 Tumor *vs*. para-tumor.

****

**Fig. 7. Liver histological, molecular, and multi-omics characteristics of human NAFLD molecular subtypes.**

(**A**) Liver histological features of the three NAFLD molecular subtypes. (**B**) Percentages of patients with NAFLD with severe steatosis, ballooning, lobular inflammation, fibrosis, or NAS in each molecular subtype. (**C** to **E**) Integrative analysis of liver proteomics as well as proteomics, lipidomics, and metabolomics of serum and urine. (**F and G**) Heatmaps of different concentrations of serum proteins, lipids, and metabolites (F) and urine protein and metabolites (G),which was increased abundance in the specific NAFLD subtypes than CTR and the other two NAFLD subtypes, and ranked among the top 5 in terms of differences compared to the CTR group. There was no different concentrations of lipid in any subtype. The proteins, lipids, or metabolites One-way ANOVA in (A and B). \**P* < 0.05 *vs*. NAFLD-mSI; #*P* < 0.05 *vs*. NAFLD-mSII.

****

**Fig. 8. Machine learning identifies hepatic proteins associated with human NAFLD molecular subtypes in internal and external validattion cohorts.**

(**A**) A diagram for identifying potential hepatic-specific biomarkers. (**B**) ROC curve analyses using random forest. (**C**) Multiplex IF analysis of hepatic LAMB1, FLNC and ERCC3 in NAFLD-mSI-III (*n*=21, 8, 28). (**D**) A confusion matrix and ROC curve (**E**) for testing the accuracy of multiplex IF analysis of LAMB1, FLNC and ERCC3 in distinguishing the three NAFLD molecular subtypes. (**F**) Accuracy (%) comparison between multiplex IF analysis and liver proteomics in identifying NAFLD molecular subtypes. (**G**) A workflow diagram for external validation. (**H**) Multiplex IF analyses of hepatic LAMB1, FLNC and ERCC3 in the external cohort. (**I** and **J**) Liver histological features of patients in external validation cohort. (**K**) Multiplex IF analysis of hepatic CYP3A4, α-SMA and HNRNPA2B1 in individuals in NAFLD-mSI-III (*n*=8, 7, 10) in the external cohort. One-way ANOVA in (C, H, I, J and K). \**P* < 0.05 *vs*. NAFLD-mSI; #*P* < 0.05 *vs*. NAFLD-mSII.