Insights into Molecular Mechanisms Underlying NAFLD by Protein-Protein Interaction Network Analysis of the Differentially Expressed Genes in NASH Patients.

Aida Najari

Islamic Azad University Medical Branch of Tehran

Aminiyanfar Masoud

Islamic Azad University, Mashhad

Mobasheri Meysam (✉ meysam.mobasheri@gmail.com)

Iranian Institute of New-found Science (IINS)

Tabassom Sobati

Iranian Institute of New-found Science (IINS)

Research Article

Keywords: non-alcoholic fatty liver disease (NAFLD), non-alcoholic steatohepatitis (NASH), Systems biology, Transcriptomics Analysis, Hub genes, Network pharmacology, poly-pharmacology

Posted Date: February 23rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-220326/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The efficient treatment of the disease such as non-alcoholic fatty liver disease as a global concern is impeded by the limited understanding of the disease complexity. We applied a systems-biology framework for a functional analysis of disease development using candidate genes. We identified, the keys dysregulation, pathways, and Hub-genes leading the disease. Moreover, we constructed a molecular disease network for the disease via integrating the pathways, providing an intuitional view to understand the development. In addition, the druggability of genes and their new application in terms of marker discovery or new drug discovery were examined. Our analyses identify a series of highly overexpressed genes, also featuring high centrality, yet being less underlined in the context of NAFLD development in the previous research. We also found several relatively low overexpressed genes, having a central place in the PPI network topology, whose potential involvement in non-alcoholic steatohepatitis might be undermined otherwise using network biology. Our results reveal that the biological processes related to disease progression are complex. Indeed, our pathway studies identified Hub genes, which may reinforce poly-pharmacological cross-talks followed by disease progression to irreversible stages holistic approaches via systems biology tools help us identify more reliable targets and disease-mediated genes integrated into translational studies and drug development in this area.

Background

In the emergence of extensive data, disease-related target discoveries of complex disease intensively require a holistic view of disease ontology to address it adequately. To this end, biological networks can uncover complex systems level properties. It is systems biology that combines experimental and computational biology to decipher the complexity of biological systems to develop effective targeted therapeutic strategies for multi-faceted and poorly studied disorders like NAFLD. In this sense, biological networks represent valuable tools for understanding disease prognosis molecularly. The complexities of biology, however, made single-targeted drugs very rare. Instead, Omics' recent efforts have aimed to explore the possibility of utilizing the poly-pharmacology strategy to predict more practical multi-target therapeutics.

Generally, NAFLD is a liver manifestation of dysregulation of metabolism. The disease entails single steatosis (NAFL) to non-alcoholic steatohepatitis (NASH) with necro-inflammatory lesions. NASH development stems from complex interactions of metabolic reactions and stress pathways in liver cells, initiated by chronic excessive lipid accumulation, with inflammatory processes driven by various immune cell populations. Moreover, due to the disease's complicated nature, Pharmacotherapy of such ailments poses a great endeavor to a holistic multi-targeted approach. This emerging approach nowadays takes advantage of high-content transcriptomic analysis and network pharmacology.

During the past decade, rapid advances in high-throughput technologies have brought unprecedented opportunities for the large-scale analysis of NAFLD genes/proteins. These datasets are often heterogeneous and multi-dimensional; integrating and arranging such datasets to ascertain the key

molecular mechanisms and transform the data into meaningful biological phenomenon a major task and challenge. To meet the demand, pathway and network-based analyses have become an integral and influential approach to elucidate the biological implications underlying complex diseases. The recent attempt has been dedicated to understanding the molecular mechanisms underlying NAFLD, mainly due to the identification of genes or pathways associated with disease initiation and progression. For example, Pathways such as insulin resistance, lipid metabolism, apoptosis, endoplasmic reticulum stress, mitochondrial dysfunction, and immune response have been found to contribute to NASH pathogenesis.

Additionally, with the development of high throughput analysis technology, more and more genes/proteins have been suggested to be linked to NAFLD and provide a valuable resource to analyze candidate genes function, biochemical pathways, and networks related to NAFLD. Using a bioinformatics analysis also, many hub genes that are essential to steatohepatitis development have been identified. For example, RPL36A, RPL14, UBE2A, UBE2B, PRKCA, EGFR, CDC42, VEGFA, PRKCA, CDC42, FCGR2A, and LARP1B were identified in GSE89632, GSE 33814, and

GSE48452 datasets (1, 2). Hub genes such as UBQLN4, APP, SHBG, CTNNB1, COL1A1, COL1A2, CD24, COL1A2, COL3A1, CXCL6, DCN, EHF, FAP, LUM, PCOLCE2, and SOX9 were all identified in a NAFLD-related fibrosis dataset (3, 4). Recently, HNF4A was identified as central to NASH pathogenesis in NASH-related integrated datasets GSE17470, GSE24807, GSE37031, GSE89632 (5). Although NAFLD's previous transcriptomic signature provides identification of intuitive genes abnormally expressed or mutated and differentiate stages from healthy obesity to severe fibrosis, most of the studies did not precisely characterize a transcriptional pattern of the gene through disease transitions. And their interplay of genes toward the critical condition such as fibrosis and network structure of such complicated disorder was less focused.

Existing studies were based mainly on target-based genes, and there is a lack of framework and comprehensive examinations of driver genes. More importantly, stage-based studies of NAFLD etiology and pathology have lacked. It is due to limited diagnostic information in the context of transformation from severe steatohepatitis to an irreversible condition such as fibrosis and HCC in a more comprehensive approach. Given the recent analyses, In this study, we expected to make contributions to identify more novel gene markers to describe how the disease progresses to fibrosis or hepatocellular carcinoma. Thus, a comprehensive analysis of the NASH-related candidate genes within a systematic framework may provide us important insights into the molecular mechanisms underlying NASH disease.

In the present study, we initially retrieved NASH-related gene expression data sets using microarray-based analysis. We analyzed differentially expressed genes (DEGs) among healthy control (HC) and NASH groups. Then we conducted the Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analysis, protein-protein interaction (PPI) network analysis to unravel molecular and biochemical mechanisms underlying disease progression and recognize the potential therapeutic relevance. Also, we elucidate that the frame of the modulation method to study disease progression could be applied to other disease models. We identified a NASH transcriptomic signature strongly enriched in

genes controlling immune-inflammatory processes and Metabolism. Moreover, we constructed a schematic molecular network for NASH disease via integrating the pathways and network, providing an intuitional view to understand the development.

Methods

data source an identification of DEGs

We retrieved the gene expression dataset GSE37031(6) from Gene Expression Omnibus (7) (GEO)

(<http://www.ncbi.nlm.nih.gov/geo/>) with platform GPL14877 Affymetrix Human Genome U133 Plus 2.0 Array [Brainarray Version 13, HGU133Plus2_Hs_ENTREZG]. The dataset includes seven healthy (control) and eight nonalcoholic steatohepatitis (NASH) samples. The samples were collected from NASH cases in the Hospital ClínicIDIBAPS-Esther Koplowitz Center, Barcelona, Spain. The distribution of the data was visualized using a PCA mapping and boxplot (Figure1). A log fold change of >1.4 and a P-value of <0.05 was considered to indicate a statistically significant difference. The genes expression data were normalized by Robust Multichip Averaging (RMA)

(8) Method of the affy package in transcriptional analysis console (TAC) (4.0.1.36) and R (v.3.4.1) (<http://bioconductor.org/biocLite.R>), and the Linear Models for Microarray Data (affy, <https://www.bioconductor.org/packages/release/bioc/html/affy.html>). (Figure2)

PPI network construction and functional annotation

The list of highly differentially expressed genes (with p-value < 0.05 and fold change ≥ 1.4) was submitted to the STRING server (<https://string-db.org>; version 10.0) (9) to construct their protein-protein interaction (PPI) network. The centrality parameters of the PPI network were computed using social network analysis. Centrality parameters, including degree, betweenness, closeness, and eigenvector, was used to identify the network Hub genes. The STRING network interaction scores for each node were expressed as a joint probability derived from curated databases of experimental results, text mining, and computationally predicted by genetic proximity. STRING networks were calculated under the criteria for linkage only with experiments, databases, text mining, and co-expression with the default settings, i.e., average confidence score: 0.400.

Hub genes' PPI networks were reconstructed by the software Cytoscape (version 3.6.0), followed by the importing of results obtained from the STRING network analysis of significant genes. Hub genes' PPI network status described as follows: (Number of nodes: 131, number of edges: 1102, clustering coefficient: 0.443, PPI enrichment p-value $<1.0e16$).

We then submitted the predicted Hub genes to the STRING server to obtain the network of interactions between their proteins. We imported the resulting network to the Gephi package for visualization and clustering. The PPI network of the Hub genes was clustered using a community detection algorithm (10). The obtained modules were then analyzed using gene set enrichment analysis by Enrichr. A combined

score >7 was applied to identify pathways enriched in the gene set in each module. We validated the enriched pathways against the literature data in the context of NAFLD.

Healthy and NASH Network characterization

We made a comparison between seven healthy and eight patient samples. After data preprocessing, we summarized it in gene-level in Affymetrix Expression Console Software (version 1.4.1.46). Analyses were performed Based on extraction and comparison of the self-expression in healthy samples and the patient. The threshold for P-value as a critical component in identifying the hub-genes was considered 0.05. Moreover, we included centrality, network diameter, stress, and clustering for more accurate disease network characterization. The screening was done by investigating gene symbols that relate to common hub genes in each sample. Finally, we identified 510 genes with a universal symbol in a healthy sample, and those 511 genes of NASH groups with a universal symbol. All of this operation was done by MATLAB (Math Works) version (9.1.0.441655).

Results

Differentially expressed genes in NAFLD patients

Normalization was carried out using both RMA and Mas5 models; the former's results, which were more mediancentric, were accepted (Table 1) (Figure 2,3). Differential gene expression analysis was performed to identify highly overexpressed genes, the top 50 of which are listed in Table 1. This list shows significant overlap with the list of genes reported to exhibit high activity in NAFLD patients (1). For instance, some of our highest-ranked DEGs—including KLF6, CYR61, LEPR, JUN, and FOS—also rank highly in NAFLD relevance.

Table 1: Top50 differentially expressed genes (with p-value < 0.05 and fold change>= 1.4)

FOSB, NR4A1, NR4A2, FOS, LOC101927552, LRRC37A2, SIK1, SYT14, TAPT1, IGFBP1, EGR2, IBTK, CYR61, IGFBP2, DNAH6, TRDN, LOC100505938, EGR1, LINC01012, GBP1, COL12A1, NR4A3, MCL1, UGT2B15, UGT2B17, ABHD5, RIN2, MDM2, SPATA6L, PHLDA1, GABRG1, NAP1L4, GASIRR, CSRN1P1, TTLL5, SERPINE1, LOC100129917, DKFZP434L187, SOCS2, SHBG, TPM4, FAM110D, PRKCA-AS1, BRE-AS1, LOC100996624, PCDHB13, C5orf58, PHLDA2, ATF3, MYL12A, KLHDC4, LOC102724057, LRRC37A3, FST, KLF6, FOXP3, JUN, TRIB1, INHBE, PIGK, CTGF, MYO5B, SORBS2, TMEM45B, MALAT1, LOC100147773, KDELC1, DUSP1, ALDOC, PPM1K, NDUFA5, USP48, CETN2, U2AF1; U2AF1L5, CENPV, HIST1H2BC, ANKRD37, CPS1-IT1, TSPAN3, ZNF638, RPS27, RPS27P29, SLC44A1, NFX1, HIST1H2BE, KLF4, FCAMR, ANKRD10, TP53I3, ORMDL2, TMED10, THBS1, NFIL3, HINT3, LPP, APOL6, MIR21, VMP1, P29, NAMPT, IL1RN, MAFF, DLEU2, NPIPA2, ZFAND5, NR5A2, MYC, MYO1B, HSPD1, FN1, TGDS, LEPR, GPR88, SKIL, EIF4A1, SLC25A25, SMG1, B4GALT1, WDR1, PRO2852, KIAA1217, ACADSB, BTG2, IRS1, DNAJC21, IRS2, ZNF281, PALM2, ADAMTS1, RAKR, PS27P29, NUDT4, NUDT4P1, NUDT4P2, FAM126B, RBBP6, ALDH6A1, ADO, HNRNPU, MKLN1, SRSF4, SCAF11, FAM134B

Nina M. Pollak *et al.*'s findings concerning the metabolic role of transcription factor KLFs in metabolic manifestation suggest that hepatic KLF2, KLF5, KLF6, and KLF9 increase hepatic lipid accumulation. Additionally, their studies demonstrate that increased hepatic KLF6 expression directly activates TGFβ1 and its receptors; this may play a crucial role in the progression from steatosis to steatohepatitis with progressive fibrosis (11). In a systematic review of NAFLD's genetic associations, Kayleigh L. Wood *et al.* demonstrate the function of KLF6 as a tumor suppressor, which confers susceptibility to non-alcoholic fatty liver disease and non-alcoholic steatohepatitis (12). Another study by the same researchers considers LEPR as the gene best poised to act as a metabolic hormone receptor—for leptin, specifically—due to its adaptive responses related to insulin resistance and increased gluconeogenesis (12).

As a result of a broad review which is done by Petta, S. *et al.* on NAFLD's pathophysiology, leptin expression is induced by endothelin expression (EDN1) and the effect of angiotensin II. These events, which are generated by proinflammatory cytokine (TNF-α and IL-6) and adhesion molecules (VCAM-1 and ICAM-1)—are significantly higher in patients with NAFLD than in the control group and are therefore correlated with NAFLD severity (13). JUN is known as a proto-oncogene, the expression of which is reported to cause inflammatory responses and cancerous behavior in fatty liver contexts (13). Lou Y. *et al.* dissected the gene regulatory network while considering the transcription factor in various NAFLD stages. Based on

their study, the top 10 putative co-regulators in the gene expression data (GSE48452) included EGR1, JUN, FOS, JUNB, FOSB, and NR4A1, which have dominant significance in liver physiology (14).

Cysteine-rich angiogenic inducer 61 (CYR61) is a secreted, [extracellular matrix](#) (ECM)-associated signaling protein of the CCN family. In adults, CYR61 contributes to [inflammation](#) and tissue repair. It is associated with diseases that involve chronic inflammation—including rheumatoid arthritis, [atherosclerosis](#), [diabetes-related nephropathy](#) and [retinopathy](#), and many different forms of cancer. However, recent studies demonstrate that the presence of CYR61 is also associated with the hypomethylated and differentially methylated regions. It has even shown upregulated expression in the livers of patients with advanced NAFLD (15). Research by [Linling Ju et al.](#) revealed that the overexpression of CYR61 is related to the over-expression of fatty acid metabolism-associated genes. In addition, in vitro experiments conducted on murine primary hepatocytes indicate that CCN member family 1 is partially responsible for increasing intracellular TG content, pro-inflammatory cytokines, and the expression level of the apoptosis-associated proteins in steatosis (16).

Reconstruction and analysis of the PPI network of DEGs

Although differential expression analysis helps describe the contribution of genes to disease progression in terms of their expression level, the expression level is not the only factor in relevant diseases related to the importance of genes. Given that most biological phenotypes emerge from gene products' interaction, a more in-depth understanding of disease at the genetic level can be realized by analyzing the protein-protein interaction (PPI) of genes active in the patients. Herein, we constructed the PPI network of genes that are highly overexpressed in NAFLD patients. The network includes 623 proteins (nodes) interacting through 2933 edges. The obtained PPI network is characterized by a 5, average path length of 2.197, and a network centralization of 0.164.

We further used the social network analysis to identify the network central genes, degree, betweenness, and closeness centralities. We were concomitantly used as a combined filter to identify the Hub genes. As a result, we identified 131 genes as the network Hubs. The Hub genes were ranked based on their degree, betweenness, and closeness centralities. According to each parameter, the genes common in the top 50 genes were identified as the network's key Hubs

(table2). These genes may hold other aspects of importance that are not reflected in the expression level. Notably, PPARA, SOD2, IL1B, FOXO, PRKAA1 (AMPK), RAF1, INSR, NF- κ B, and CREBBP. they were absent from the list of our top 50 DEGs, are among the identified Hub genes. The significance of these genes in the context of NAFLD initiation/progression is well documented (3). These facts signify the network biology approach's potential in identifying important disease-driving genes that may not be uncovered by merely relying on expression level.

Table 2: Top50 hub genes (with p-value < 0.05 and fold change>= 1.4)

JUN, MYC, IGF1, PTEN, FOS, CTNNB1, INSR1, ACTB, NFKBIA, PTGS2, EGR1, MDM2, SMARCA2, IL-1B, PHLPP1, NR3C1, CREBBP, GART, SIRT1, EDN1, CCL2, RAF1, FOXO1., IRS1, ATM, CXCL12, DUSP1, MET, BCL6, SERPINE1, WDFY2, PPARA, AMAD4, SOD2, FOXO3, FGFR1, KLF2, KLF4, RHOB, XPO1, ICAM1, APP, HSPD1, ERN1, ATF3, EPHA2, NF1, MCL1, LDLR, FN1, INSR, NFKBIAFOS, EIF4A1, HNRNPU, SMAD4, NAMPT, CYP2B6, ZFP36, CSNK1A1, AP2A1, G6PC, ARF1, CTGF, FUS, HNRNPA2B1, POLR1B, H6PD, REL, AGTR1, SCARB2, JUND, THBS1, EGR2, NR4A1, IGF2, LATS2, ID2, NCOR1, TRIP12, DIP2A, DLG1, TAT, CUL4A

Structural clustering of Hub genes network

By using the STRING server, the integrated network of the Hub genes was reconstructed. As can be observed from Table 3, this network has a higher density, clustering coefficient, and average path length than the initial network using the community detection algorithm in Gephi 0.9.2 (www.gephi.com), an open-source network mapping tool (Figure 4).

Table 3: Network characteristic of significant genes compared with detected Hub genes

Network	Network diameter	Network density	Average clustering coefficient	Average path length	Network centralization
623nodes	10	0.015	0.287	2.197	0.164
131nodes	5	0.129	0.443	2.354	0.423

The network was clustered into three distinct and two small modules. For each module, the centrality parameters were calculated. Table 4 presents the lists of top-ranked genes in each module. Top-ranked genes attributed to module1 include proto-oncogenes (MET, MYC, RAF1, MDM2) and tumor suppressor (PTEN, ATM, ERN1, BCL6). In module 2, genes related to the metabolic aspect of the disease, for example, PPARA (with a critical role in the regulation of fatty acid uptake, beta-oxidation), INSR1 (a receptor tyrosine kinase which mediates the pleiotropic actions of insulin), Foxo1, G6PC, DUSP1 (may play an essential role in the human cellular response to environmental stress as well as in the negative regulation of cellular proliferation) constitute a particular characteristic of metabolic disease. These genes are ranked from those with an energy-sensing component to those that activate compensatory response modulators.

Along with module 3, there are some genes associated with inflammation and immune response such as IL-1B (a member of the [interleukin 1 family](#) of [cytokines](#) which is an important mediator of the [inflammatory response](#) and is involved in a variety of cellular activities, including cell proliferation, differentiation, and [apoptosis](#)), CXCL12 (The encoded protein functions as the ligand for the G-protein coupled receptor, chemokine (C-X-C motif) receptor 4, and plays a role in numerous diverse cellular

functions, including embryogenesis, immune surveillance, inflammation response, tissue homeostasis, and tumor growth and metastasis), ICAM1 (This gene encodes a cell surface glycoprotein which is typically expressed on endothelial cells and cells of the immune system) and PTGS2 (PTGS2 is responsible for the production of inflammatory prostaglandins). The third module contains liver cytochromes genes such as CYP2B6, CYP1A1, CYP3A4 and encodes a member of the cytochrome P450 superfamily of enzymes. They catalyze many reactions in drug and xenobiotic metabolism and the synthesis of cholesterol, steroids, and other lipids. The last module contains HNRNPM, HNRNPU, SNRPA1, and SRSF6, which is involved in Spliceosome, a family of proteins that binds nucleic acids and functions in the formation of ribonucleoprotein complexes. It is widely believed that changes in splicing may be an early event in cancer initiation. Moreover, several studies indicate the role of these changes in the manifestation of NAFLD disease (17). Hence, over-expression of Splicing factors might be considered an early event to detect the progression of NAFLD.

Table 4: Top ranked gene in each module				
Module1	module2	Module3	Module4	Module5
MYC	JUN	IGF1	NR3C1	OAT
ACTB	EGR1	EDN1	CYP2B6	H6PD
SMARCA2	FOXO1	SERPINE1	AHR	POLR1B
ATM	SMAD4	APP	CYP1A1	DBT
PTEN	FOS	PTGS2	CYP3A4	GART
NFKBIA	CREBBP	CCL2		GLS
PHLPP1	IRS1	WDFY2	TAT	HNRNPM
CTNNB1	INSR	IL-1B	DIP2A	HNRNPU
MDM2	SIRT1	CXCL2	ABCB11	SNRPA1
RAF1	DUSP1	ICAM1	VNN3	SRSF6
MET	G6PC	CTGF		
REL	ZFP36	NAMPT		
BCL6	PPARA	FN1		
ACTB1		AP2A1		

Gene set enrichment analysis

We conducted Functional enrichment on two sets of genes: significant genes and hub genes. We searched for enriched pathways of NASH-related genes and identified 212 significant enrichment pathways for hub genes, specifically. The KEGG items with the highest combined scores are represented in [Figure 5](#). The Foxo signaling, PI3k-Akt signaling, and AGE-RAGE signaling pathways scored highest among the hub gene sets presented in this study. Figure 6 portrays the distribution of hub genes in the highest-ranked pathways of the disease. We also compared pathway enrichment, both in the community of significant genes and our hub nodes, before modulation. There were negligible discrepancies, which implies that our selected hub nodes behaved in the same manner (though on a smaller scale) as the significant genes. However, we also found that some critical pathways present in our hub gene enrichment—such as “focal adhesion,” “glucagon signaling pathway,” and “Rheumatoid arthritis”—were absent or less significant in our significant gene enrichment. This presentation explains how the process

of finding hub nodes among significant genes can extrapolate subtle details related to disease characteristics.

Network characterization of the healthy and control subjects

The healthy subjects and patients' network parameter was individually calculated by Cytoscape version (3.6.0) (Table 5). We noticed that healthy subjects and patients' network indicators are close to each other. It is manifested that the disease's network structure has negligibly changed, and fatty liver disease and NFLD are considered a systematic disease.

Table 5: network parameter of the healthy subjects and patients by Cytoscape version (3.6.0)

	Network density	Network diameter	Network centralization	Shortest path
Control samples	0.031	10	0.107	20890(87%)
Nash samples	0.025	10	0.090	22960(90%)

Discussion

A combination of mechanisms—including persistent inflammation, metabolic dysfunction, and insulin resistance— are responsible for non-alcoholic steatohepatitis (NASH). Thus, the multifactorial nature of NASH requires a systems biology approach to the study of the disease to decipher the complexity associated with its initiation and progression. The present study aimed to gain a holistic PPI-level insight into biological processes driving the NASH progression. To this end, we first compared the gene expression profiles of NASH patients’ samples with the healthy subjects to identify the significantly overexpressed genes in the patients. Next, we reconstructed and analyzed the PPI network of DEGs to identify key NAFLD associated genes at the network level.

Although the number of our samples was limited, we used previous gene expression studies to gather gene signatures within the different stages of the disease, from obesity to progressive NASH.

Several genes related to immunity, oxidative stress responses, and metabolic functions were determined to play a role in NAFLD (12, 18-20). We found a satisfactory agreement between our results and previous findings on the role of genes in the disease's modulation. Also consistent with our results, biological processes such as focal adhesion, inflammatory response, fibrosis, and cellular response to chemical stimuli are highlighted in the literature (19-21).

The centrality analysis of DEGs' PPI networks identified the network's hub genes, including PPARA, CREBBP, CCL2, SERPINE1, ABCB11, SOD2, Foxo1, IL-1B, INSR, NF- κ B. These are both theoretical and experimental shreds of evidence to demonstrate that the PPI network hubs often play important roles in regulating the biological processes involving disease (4, 22) (23). Previous studies consistently show the significance of our identified hub genes as critical modulators of NAFLD initiation and progression (12, 18, 22, 24). Interestingly, several genes absent from the list of top DEGs, including PRS6KA5, CFLAR, TNFSF11, ADRB2, RAF1, RAGE, MEF2A, NFE2L2, ID2, GART, ZFP36 were found among the high-score hub genes. Among them, four proteins stand out as relatively unknown factors in NASH development.

Most notable among hubs is the Zfp36, which encodes tristetraprolin protein (a zinc finger transcription factor), which plays a significant role in negatively regulating [TNF \$\alpha\$](#) production by destabilizing its mRNA (25). A recent study has demonstrated that tristetraprolin post-transcriptionally regulates systemic insulin sensitivity and hepatic metabolism through the modulation of liver-derived FGF21 (26). Another study found that insulin resistance in obese mice was associated with enhanced Zfp36 expression in the hepatic macrophage (m1). The study of Caracciolo, V. *et al.*, revealed that the myeloid-specific deletion of Zfp36 protects against insulin resistance and fatty liver in mice whose obesity is diet-induced (27). Moreover, a significant correlation has been reported between TTP and hepatocarcinogenesis (28).

Another less explored gene is NFE2L2, which encodes nuclear factor erythroid 2-related factor 2 (NRF2), a transcription factor that mediates protection against [oxidative damage](#) triggered by injury or inflammation. Lu, C. *et al.* demonstrate the role of curcumin in increasing NRF2 expression (29). Due to its antioxidant and anti-inflammatory properties, curcumin's natural polyphenol has long been proposed as a potentially viable treatment for NAFLD. A recent systematic review and meta-analysis of a randomized controlled trial demonstrated that curcumin supplementation has favorable effects on metabolic markers and anthropometric parameters in patients with NAFLD (30). Chambel, S.S. has reported that NRF2 also mediates the impact of lipid metabolism on antioxidant defense, as observed in NAFLD experimental models (31). In another research, the pharmacological activation of NRF2 in obese and insulin-resistant mice was found to reverse insulin resistance, suppress hepatic steatosis, and mitigate NASH and liver fibrosis (19). Hence, the pharmacological induction of NRF2 appears to be a promising strategy for NAFLD prevention and treatment.

Another underscored genes less focused in NAFLD in the connective tissue growth factor (CTGF), a [multicellular component](#) of the [extracellular matrix](#)-associated [heparin-binding](#) proteins involved in many biological processes, including [cell adhesion](#), [migration](#), [proliferation](#), [angiogenesis](#), and [wound healing](#). The overexpression of CTGF has been considered a hallmark of fibrosis (32, 33). Besides, aberrant CTGF expression is associated with many types of malignancies, [diabetic nephropathy](#) and [retinopathy](#), arthritis, and cardiovascular diseases. Yoshino, J. showed that CTGF overexpression was associated with adipose tissue expansion and multi-organ insulin resistance in obese subjects (34). Colak, Y. *et al.* demonstrated that CTGF serum levels might be a clinical marker for distinguishing those NAFLD patients with advanced fibrosis from those who do not (35).

Lastly, CFLAR (a regulator of apoptosis that is structurally similar to caspase-8) functions as a suppressor of steatohepatitis and its metabolic disorders. According to Wang P.X. *et al.*, CFLAR attenuates steatohepatitis progression in both mice and monkeys (36). Liu Y. *et al.* demonstrate that Silibinin, a flavonolignan from milk thistle, performs its function by activating the CFLAR-JNK pathway (37). In so doing, it regulates downstream target genes involved in lipid metabolism (PPAR α , SREBP-1C, and PNPLA3), glucose uptake (PI3K-Akt), oxidative stress (NRF2, CYP2E1, and CYP4A), an inflammatory response. Analyses of treated HepG2 cells have confirmed its potential use in improving various symptoms of NASH (38). Silymarin is the extract of *Silybum marianum*, or milk thistle, which has been used to treat various liver disorders—particularly chronic liver diseases, cirrhosis, and hepatocellular carcinoma—because of its antioxidant and anti-inflammatory properties. (39)

The translational relevance of this research is accelerated by our authentication of the PPI networks of central genes, our interpretation of rationales, and the broad mechanisms we have drawn from the progression of the disease to an irreversible stage of non-alcoholic steatohepatitis in a previously healthy patient. Our biological function enrichment analysis enabled us to narrow the functional spectrum of NASH-related genes. Summarizing the results from our pathway and network analyses, we were able to pinpoint the vital pathways to metabolic perturbation, stress-related responses, and fibrosis initiation. The Foxo signaling pathway, the PI3k-Akt signaling pathway, and select pathways in cancer produced the highest combined scores. This result was consistent with previous research, thus offering valuable evidence to study the complicated connection underlying the NASH disease. Accordingly, greater emphasis is placed on the AMPK signaling pathway, the insulin resistance pathway, and the Foxo signaling pathway in several research lines on NAFLD progression (21, 40). Also included are several pathways involved in the cellular physiological processes linked to tumor cell proliferation, such as apoptosis, necroptosis, adherence junction, and cell cycle. Following previous research, some pathways are related to the HCC signaling pathway, such as the transforming growth factor-beta signaling pathway, the NF- κ B signaling pathway, and the Hippo and the JAK-STAT signaling pathway.

The clustering of the interactions between hub genes resulted in the PPI network's dissection into three distinct modules. According to previous studies, dense interactions between a particular set of proteins may underlie the biological functions coordinated by those proteins (41). Therefore, the clustering of the PPI network provides insights into the modules' biological functions and processes that might otherwise be challenging to uncover. Pathway enrichment analysis is broadly used to interpret PPIs in terms of their biological functions and processes. Discovering the pathways and processes most likely to be coordinated by each PPI module in a particular disease can reveal the molecular mechanisms that drive specific diseases. Owing to the complex and multifaceted nature of NASH, we could not attribute the individual clusters to specific aspects of the disease. Nevertheless, the groups of hub genes in each module constitute a mixture of specific characteristics. We can find genes that are metabolically related to oncogenic/apoptosis or inflammation/fibro-genesis (Figure 7,8).

We found the first cluster with members of the topologically remarkable gene (of the entire network of hub genes)(Table 4) receiving the highest enrichment scores through the apoptosis signaling pathway (P-

value: 3.000E-10), the PI3K-Akt signaling pathway (P-value: 9.799E-10), the Foxo signaling pathway (P-value: 2.46E-07), and other cancer-related pathways (P-value: 6.935E-08) (Figure 8). This cluster study confirms that the majority of the functional genes presented in this module are involved in the interaction between the tumor-suppressing and oncogenic signals through the PI3K-Akt signaling pathway. This cluster also explains the prognosis of severe liver fibrosis or hepatocellular carcinoma (42). Many studies have demonstrated that the PI3K-Akt signaling pathway balances oncogenesis and cell survival signals by regulating pro- and anti-apoptotic genes (40, 43). Conversely, the activation of the pathways such as apoptosis and Foxo signaling pathways in this module may indicate the activation of the cellular defensive mechanisms that counter the disease's progressive traits.

Moreover, enrichment of these pathways with a significant p-value can imply dysfunction in the Akt to regulate its downstream pathways in favor of maintaining balance. From this specific module, we found tumor suppressors (e.g., PTEN, PHLPP-1, ATM, BCL6, KLF2, ERN1) and protein oncogenes (e.g., MDM2, XIAP, ACTB, RAF1, CFLAR, and MCL1) among our top hub gene which are overexpressed. Furthermore, they may play a prominent role in exacerbating disease and insulin resistance as a hallmark of NASH progression.

It is worth noting that Akt is also considered a master regulator in insulin-mediated glucose homeostasis. Evidence indicates that Akt is negatively regulated by tumor-suppressing proteins whose activity interferes with insulin-mediated glucose homeostasis. Hub genes PHLPP1 and PTEN in this cluster are both tumor suppressors and negative regulators of Akt. Recently studies suggest that the overexpression of PHLPP1 may contribute to type 2 diabetes by interfering with Akt-mediated insulin signaling (44-46). Similarly, the PTEN overexpression is induced and mediated by high levels of free fatty acids and inflammatory cytokines (47-49) and can be involved in controlling cell division, whereas it is said that it could exacerbate insulin unresponsiveness. Consequently, despite the several findings that demonstrate inhibition of Akt's downstream signaling by these suppressors, it might decrease oxidative stress and DNA damage response, and this appears to contradict with glucose homeostasis function of Akt (43, 50). The clustering of all of these genes together in a single module may reflect the difficulties associated with treating insulin resistance with an antagonizing approach—a well-known challenge in managing cancers and diabetes.

A study of pathways enriched in the second highly populated cluster highlights insulin unresponsiveness and the impaired metabolic balance that justify another molecular aspect of NASH. This fact is reflected in the high enrichment scores of the AMPK signaling pathway (P-value=8.187E-14), the Foxo signaling pathway (P-value=8.187E-14), the insulin resistance pathway (P-value=6.149E-09), and the adipocytokine signaling pathway (P-value= 2.272E-6) (Fig. 8). Remarkably, we found that almost all of these enriched pathways might shed light on feedback against cells' disorganized bioenergetics. This disorganization can happen through nutrient-sensing signals and glucose homeostasis. The activation of these pathways reinforces the fact that several mechanisms, including elevated gluconeogenesis, lipolysis, and increased fatty acid metabolism oxidation, at the onset of NASH development. These pathways represent maintaining cell metabolism; nevertheless, the combination of prolonged insulin unresponsiveness,

accumulated stress, and metabolism perturbation can lead to pathologic conditions, making them "double-edged swords in the pathology of the disease.

Another sub-network constructed in this module elaborates on the complicated link between metabolic perturbation and inflammatory responses. This linkage is highlighted with the enrichment of pathways such as the MAPK signaling pathway (P-value=3.603E-08) and the AGE/RAGE signaling pathway (P-value=9.832E-06) in this specific module may account for the stress that may originate from insulin-mediated lipogenesis that subsequently gives rise to the progression of NASH. Meanwhile, several lines of evidence demonstrate a disruption in the balanced input and output of hepatic FFA manifested in active ROS generation (51). We can understand how hyperlipidemia due to obesity and hyperglycemia in the NASH cases and inflammation and oxidative stress may cause AGE products' formation through the glycosylation process (52). Generally, we can interpret this cluster's biological functions as a reflection of the mechanisms of transforming NAFLD characterized in the two-hit hypothesis. This NASH incidence theory implicates excessive fatty acid and insulin resistance as a "first hit," which are supposed to be connected to mitochondrial dysfunction and oxidative stress (36). A modular study of the third module explains the disease, which moves toward the steato-apoptotic stage. Our analysis indicates two pathways—the TNF signaling pathway (P value=4.908E-07) and the NF-kappa B signaling pathway (P value= 4.908E-07)—were also enriched with genes involved in NASH development. (Figure 8). These two pathways have been studied extensively for their role in inflammation and immunity responses [24]. The AGE/RAGE signaling pathway (P value=2.120E-07) and its downstream gene are prevalent in our hub gene enrichment; in fact, they appear to explain contribution to the progression of the disease. The enrichment of these pathways highlights the over-activity of pro-inflammatory components in this pathway, which triggers positive feedback mechanisms via AGE/RAGE. These inter-related activities—which are known to insulin resistance and systemic complications and the progression of NASH to fibrosis—are presented in our clustering analysis of the disease PPI network. RAGE ligands (AGEs), which are generated by the non-enzymatic glycation of accumulated lipids and glucose, trigger oxidative stress pathways and cause the over-activation of receptors through a positive feedback loop. This phenomenon has detrimental effects on hepatic insulin resistance, steatosis, fibrosis, ischemic and non-ischemic liver disease, and the growth and metastasis of HCC. Consequently, receptor blockages or restrictions in dietary AGEs appear to be an influential therapeutic target for these progressive hepatic disorders, as supported by numerous studies(53) (54) (55)(56).

The network-based analysis offers important insights about functional genes (e.g., RAF1, PTGS2, ATM, TNFSF11, CXCR4, ICAM-1, PTEN, SMARCA2, H6PD, SERPINE1, GLS, NAMPT), but it can also help to identify novel candidate targets and markers. On the one hand, some genes exhibit the pathological characteristics found in other disease contexts, including inflammatory diseases (e.g., rheumatoid arthritis, diabetes, metabolic syndrome, and various cancers) that can be repurposed as a pharmacological target in new disease. This complexity can cause challenges in determining the roles these genes playing in the context of fatty liver disease. On the other hand, some hub genes targeted by FDA-approved drugs—such as ADRB2, PTGS2, and AGTR1—are related to different diseases (e.g., cardiovascular disease) creates opportunities for repositioning and poly-pharmacological strategies

about NASH disease. As illustrated in the pathway enrichment, two interconnected mechanisms—including RAS activation and lipolysis regulation—appear to be stimulated by the sympathetic system. The beta-receptor (ADRB2) is among the top-ranked hub genes whose activity in the adrenergic system may lead to RAS activation through multiple signaling steps. RAS is a hallmark of several manifestations of NASH (including ROS formation and fibrosis initiation) (57).

Similarly, the receptor of angiotensin, AGTR1, is among the network hubs that cause aldosterone's biosynthesis through its function. This result can be reflected by a high score of the aldosterone synthesis pathway in the functional enrichments. Also, pathways pertaining to lipolysis regulation in adipocytes, which correlate with beta-receptor overactivity, are highly enriched. Thus, beta-receptor and angiotensin II receptor inhibitors, which are widely available in the market, might be considered for therapeutic use in the future to mitigate the devastating effects of fatty liver on the body.

In recent decades, growing evidence has highlighted the AGE/RAGE axis's pathological role in various diseases, such as diabetes and fatty liver disease. The mechanisms through which the AGE-RAGE pathway influences inflammatory reactions include rising oxidative stress generation and inflammatory responses in its downstream. Four proteins involved in this pathway (SMAD4, ICAM-1, EDN1, SERPINE) can be considered the key players in further translational applications in the context of NASH. The first key hub gene, SMAD4, is part of the protein complexes in this study. This hub gene encodes a protein that acts as a signal transducer in the activation of the fibrogenic pathway and apoptosis through HSC activation and can lead to organ damage such as diabetic nephropathy (58-60). One research has highlighted the role of SMAD4 as a risk factor for fibrosis in conjunction with BMI, TG, LDL-C, ALT, and AST (58). Studies in favor of the deletion of SMAD4 overexpression have shown that improvements to lipid metabolism, liver function, inflammation, or fibrosis could confirm its targeted role in improving fibrosis in the NASH context.

Similarly, the second Hub gene, ICAM-1, is associated with the increased adhesion of leukocytes via increased intercellular adhesion molecule-1 in the membrane of leukocyte and endothelial cells. As a result, more immune responses are activated, and more reactive oxygen is generated. Our data agree with previous findings explaining that hepato-steatosis is associated with increased hepatic ICAM-1 expression. Therefore, setting this glycoprotein as a target in graft protection in fatty liver subjects (and as a biomarker for susceptibility to organ injury, as discussed in previous studies) would be more promising.

The overexpression of the endothelin-1 gene (EDN1) due to the transcriptional regulator's hypoxia-induced activation could explain oxygen hemostasis irregularity in the NASH state (61). Previous studies have found higher levels of endothelin-1 in NASH patients correlated with the grade of their hepatic fibrosis. Such studies have also demonstrated the role of EDN1 as an angiogenic factor in tumor metastasis, and studies on zebrafishes have mentioned the liverspecific expression of EDN1-induced HCC. Hence, commercially available endothelin-1 receptor antagonists might be an excellent therapeutic target both for hypoxia-induced fibrosis and tumor growth that may occur in NASH disease. The activation of the

AGE/RAGE axis could also produce an extrahepatic reaction. Studies indicate that the overexpression of SERPINE1 downstream of the AGE/RAGE axis predisposes patients to severe fibrosis and systemic vascular complications in the fatty liver (including atherosclerosis), which further complicates the condition (62).

Additionally, microvascular dysfunction occurs due to the overexpression of ICAM-1 in the endothelial cells may facilitate atherosclerosis formation. As a result, these two genes might be considered CVD risk factors, which could direct drug development efforts toward treating the extrahepatic complications of NAFLD. In agreement with other studies, we confirmed the involvement of these four genes in NASH pathogenesis.

Studying the PPI network could also highlight genes with promising pharmacological targets for further investigation. For instance, recent literature has suggested that BCL6 (a master immune system regulator) could play a role in metabolic regulation, which is not related directly to its function through immunity response (63, 64). A study of knockout mice (2014) illustrated that BCL6 deletion decreased lipogenesis through alteration in SREBP1c Fasn and Scd1. It also reduced adipogenesis and fatty acid oxidation (via PPAR) (64). The BCL6 hub gene—which is coexpressed with hub genes like Foxo1, KLF2, and ATM in our study—supports new findings that BCL6 acts as an intermediary, relating stress response the metabolic irregularity. As the co-expression illustrates, it may form a bridge with another hub genes such as JUN, IL-1b, REL, CCI2, and SERPINE1. Likewise, considering the significance of BCL6 in cross-talk between apoptosis and metabolic regulation, the inhibition of BCL6 could be a plausible target for future treatment of fatty liver disease and insulin resistance (65, 66).

Our pathway analysis also revealed that the disease's progression through oxidative response ultimately leads to defects in cell bioenergetics. We can understand some changes as subordinate to metabolic reactions related to energy hemostasis (e.g., purine and glutamine metabolism), which are said to compensate for energy imbalance failures. In terms of dominant hub genes related to metabolic signaling pathways, those most strongly aligned with the metabolism were GART, GLS, and H6PD.

GART, which is involved in the de-novo biosynthesis of purine, causes noticeable changes in purine metabolism. Many studies have identified an elevation in uric acid in NAFLD patients (67, 68); this appears to occur after the upstream elevation of purine biosynthesis and after the activation of additional metabolic signals through purine metabolism. It produces uric acid as a byproduct. However, some studies reflected that uric acid's elevated formation could be a compensative mechanism against disease progression. Other studies established a high uric acid level in increasing insulin resistance and lipogenesis conditions (69). In addition, many studies point out that purine metabolism plays a significant role in cancer prognosis (70, 71). Based on this evidence, it is likely that alterations to this metabolic pathway in favor of cell proliferation contribute to the disease's cancer-like progression. Ultimately, we can conclude that GART-expressed enzymes, not uric acid, can probably act as promising biomarkers for earlier detection of NAFLD progression.

The GLS gene overexpression represents a metabolic alteration in glutamine catabolism. A recent study by J. Simón *et al.* revealed that glutaminase (GLS), regulated by c-myc proto-oncogene, is overexpressed in the late stages of NASH and the early stages of HCC (72). They concluded that the enzyme's inhibition could deactivate the Krebs cycle and the electron transport chain (ETC) and decrease β -oxidation, which contributes to reduced ROS formation (73). Another study by *Miller RA et al.* shows that increased activity of haptic glutaminase (GLS2) is closely related to carbon generation. This carbon is needed for gluconeogenesis through mitochondrial stimulation anaplerotic reactions in response to glucagon (74). This study indicates the role of glutamine degradation in the gluconeogenic pathway under insulin resistance conditions. Afterward, Hyper-ammonia can also occur during increased glutaminase activity. *Javier Ampuero et al.* demonstrated the role of metformin as a hepatic glutaminase inhibitor. So this medicine can primarily be proposed to treat encephalopathic cirrhosis caused by excess ammonia in the brain (75). A more consistent occurrence is the dysregulated expression of liver aminotransferase, such as ornithine transaminase produced through OAT expression. So, we may find overexpression of OAT in our hub sets of genes as a scavenger that responds to the hyper-ammonia state (76). Our study hypothesizes that GLS2 can present promise not only for the treatment of hyperglycemic diabetes but also for the termination of hyper-gluconeogenic activity due to insulin resistance in fatty liver disease (74).

Hub gene H6PD—the proteins of which are known to activate 11- β HSD1 by generating NADPH (77)—influences the rise in corticosteroid-associated metabolic activity (e.g., gluconeogenesis, biosynthesis, insulin resistance, accumulation of visceral fat, vascular reactivity, vascular remodeling, and sodium reabsorption). Since excess glucocorticoids promote obesity, hyperlipidemia, insulin resistance, and antagonizing approach in the therapeutic strategies for all metabolic syndromes could be more promising in the treatment of fatty liver disease.

Despite decades of research, NAFLD remains one of the most complex diseases with no efficient cure. While previous studies have demonstrated this disease's multifactorial nature, our study shed more light on the complexity associated with it at the molecular level. Our study confirms many genes well associated with NASH from previous reports. It was able to identify several genes less studied in the NAFLD context, an inspection of which in future research may help better characterization and develop a more effective treatment. Our gene expression, PPI network, and enrichment analysis also provide useful insights into NASH's development at the pathway level. Indeed an extensive body of the current knowledge of pathways contributing to NASH was reflected in our network and enrichment analysis. Clustering analysis of DEGs' PPI network also revealed more in-depth associations among genes that support disease progression, being otherwise challenging to identify. Our research demonstrates the usefulness of network and systems biology in facilitating the integration of separate NAFLD molecular data in a coherent framework, which enables consistent interpretation of the data in the context of molecular disease mechanisms and therapeutic target identification.

Abbreviations

NAFLD: Non-alcoholic fatty liver disease; HCC: Hepatocellular Carcinoma; NASH: Non-alcoholic Steatohepatitis; KEGG: Kyoto Encyclopedia of Genes and Genomes; GEO: Gene Expression Omnibus; TAC: Transcriptomics Analysis Console; PCA: Principal Component Analysis; DEGs: Differentially Expressed Genes; PPI: Protein-Protein interaction; AGEs: Advanced glycation End-products; RAGE: The receptor for advanced glycation end products; HSC: Hepatic stellate cell; CVD: Cardiovascular disease. HepG2 cells: liver hepatocellular carcinoma cell line; NADPH: Nicotinamide adenine dinucleotide phosphate; RAS: Renin-angiotensin system; FFA: Free fatty acid; T2DM: Type 2 diabetes; BMI: body mass index; TG: Triglycerides; LDL-C: Low-density lipoprotein; ALT: alanine transaminase; and AST: aspartate aminotransferase; FDA: The United States Food and Drug Administration; ECM extracellular matrix.

Declarations

Availability of data and materials

All relevant raw data on which the conclusions of the paper rely are available

Ethics approval and consent to participate

Not applicable

Competing interests

The Authors declare that they have no competing interests

Authors' contributions

MM was responsible for conceiving, designing, and conducting the study and contributed to interpreting the results and drafting and revising the manuscript. NA made a significant contribution in analyzing the data, interpretation, and validation of the draft manuscript's data and preparation. AM contributed to some parts of analysis and visualization. TS was involved in the verification of the results.

References

1. Wang R, Wang X, Zhuang L. Gene expression profiling reveals key genes and pathways related to the development of non-alcoholic fatty liver disease. *Annals of Hepatology*. 2016;15(2):190-9.
2. Li L, Liu H, Hu X, Huang Y, Wang Y, He Y, et al. Identification of key genes in non-alcoholic fatty liver disease progression based on bioinformatics analysis. *Molecular Medicine Reports*. 2018;17.
3. Qi S, Wang C, Li C, Wang P, Liu M. Candidate genes investigation for severe nonalcoholic fatty liver disease based on bioinformatics analysis. *Medicine (Baltimore)*. 2017;96(32):e7743-e.
4. Ryaboshapkina M, Hammar M. Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Scientific Reports*. 2017;7.

5. Baciú C, Pasini E, Angeli M, Schwenger K, Afrin J, Humar A, et al. Systematic integrative analysis of gene expression identifies HNF4A as the central gene in pathogenesis of non-alcoholic steatohepatitis. *PLoS One*. 2017;12(12):e0189223e.
6. Lopez-Vicario C, Gonzalez-Periz A, Rius B, Moran-Salvador E, Garcia-Alonso V, Lozano JJ, et al. Molecular interplay between Delta5/Delta6 desaturases and long-chain fatty acids in the pathogenesis of non-alcoholic steatohepatitis. *Gut*. 2014;63(2):344-55.
7. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res*. 2013;41(Database issue):D991-5.
8. Kim C, Hwang S, Zhang S-D. RMA with quantile normalization mixes biological signals between different sample groups in microarray data analysis. *Proceedings - 2014 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2014*. 2014:139-43.
9. 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-d13.
10. Blondel V, Guillaume J-L, Lambiotte R, Lefebvre E. Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics Theory and Experiment*. 2008;2008.
11. Pollak NM, Hoffman M, Goldberg IJ, Drosatos K. Kruppel-like factors: Crippling and un-crippling metabolic pathways. *JACC Basic Transl Sci*. 2018;3(1):132-56.
12. Wood KL, Miller MH, Dillon JF. Systematic review of genetic association studies involving histologically confirmed non-alcoholic fatty liver disease. *BMJ Open Gastroenterol*. 2015;2(1):e000019.
13. Petta S, Gastaldelli A, Rebelos E, Bugianesi E, Messa P, Miele L, et al. Pathophysiology of Non Alcoholic Fatty Liver Disease. *International journal of molecular sciences*. 2016;17(12):2082.
14. Lou Y, Chen YD, Sun FR, Shi JP, Song Y, Yang J. Potential Regulators Driving the Transition in Nonalcoholic Fatty Liver Disease: a Stage-Based View. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology*. 2017;41(1):239-51.
15. Hotta K, Kitamoto T, Kitamoto A, Ogawa Y, Honda Y, Kessoku T, et al. Identification of the genomic region under epigenetic regulation during non-alcoholic fatty liver disease progression. *Hepatology research : the official journal of the Japan Society of Hepatology*. 2018;48(3):E320-e34.
16. Ju L, Sun Y, Xue H, Chen L, Gu C, Shao J, et al. CCN1 promotes hepatic steatosis and inflammation in non-alcoholic steatohepatitis. *Scientific Reports*. 2020;10(1):3201.
17. Webster NJG. Alternative RNA Splicing in the Pathogenesis of Liver Disease. *Front Endocrinol (Lausanne)*. 2017;8:133-.
18. Ryaboshapkina M, Hammar M. Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Scientific reports*. 2017;7(1):12361-.
19. Wang R, Wang X, Zhuang L. Gene expression profiling reveals key genes and pathways related to the development of non-alcoholic fatty liver disease. *Annals of hepatology*. 2016;15(2):190-9.

20. Qi S, Wang C, Li C, Wang P, Liu M. Candidate genes investigation for severe nonalcoholic fatty liver disease based on bioinformatics analysis. *Medicine (Baltimore)*. 2017;96(32):e7743.
21. Li L, Liu H, Hu X, Huang Y, Wang Y, He Y, et al. Identification of key genes in nonalcoholic fatty liver disease progression based on bioinformatics analysis. *Mol Med Rep*. 2018;17(6):7708-20.
22. Liss KH, Finck BN. PPARs and nonalcoholic fatty liver disease. *Biochimie*. 2017;136:65-74.
23. Wood KL, Miller MH, Dillon JF. Systematic review of genetic association studies involving histologically confirmed non-alcoholic fatty liver disease. *BMJ Open Gastroenterol*. 2015;2(1):e000019-e.
24. Marra F, Tacke F. Roles for Chemokines in Liver Disease. *Gastroenterology*. 2014;147(3):577-94.e1.
25. Bouchard L, Vohl M-C, Deshaies Y, Rhéaume C, Daris M, Tchernof A. Visceral adipose tissue zinc finger protein 36 mRNA levels are correlated with insulin, insulin resistance index, and adiponectinemia in women. *European journal of endocrinology*. 2007;157(4):451-7.
26. Sawicki KT, Chang HC, Shapiro JS, Bayeva M, De Jesus A, Finck BN, et al. Hepatic tristetraprolin promotes insulin resistance through RNA destabilization of FGF21. *JCI insight*. 2018;3(13).
27. Caracciolo V, Young J, Gonzales D, Ni Y, Flowers SJ, Summer R, et al. Myeloid-specific deletion of Zfp36 protects against insulin resistance and fatty liver in diet-induced obese mice. *American journal of physiology Endocrinology and metabolism*. 2018;315(4):E676-e93.
28. Krohler T, Kessler SM. The mRNA-binding Protein TTP/ZFP36 in Hepatocarcinogenesis and Hepatocellular Carcinoma. 2019;11(11).
29. Lu C, Zhang F, Xu W, Wu X, Lian N, Jin H, et al. Curcumin attenuates ethanol-induced hepatic steatosis through modulating Nrf2/FXR signaling in hepatocytes. *IUBMB Life*. 2015;67(8):645-58.
30. Jalali M, Mahmoodi M, Mosallanezhad Z, Jalali R, Imanieh MH, Moosavian SP. The effects of curcumin supplementation on liver function, metabolic profile and body composition in patients with non-alcoholic fatty liver disease: A systematic review and meta-analysis of randomized controlled trials. *Complementary therapies in medicine*. 2020;48:102283.
31. Chambel SS, Santos-Goncalves A, Duarte TL. The Dual Role of Nrf2 in Nonalcoholic Fatty Liver Disease: Regulation of Antioxidant Defenses and Hepatic Lipid Metabolism. *BioMed research international*. 2015;2015:597134.
32. Kodama T, Takehara T, Hikita H, Shimizu S, Shigekawa M, Tsunematsu H, et al. Increases in p53 expression induce CTGF synthesis by mouse and human hepatocytes and result in liver fibrosis in mice. *J Clin Invest*. 2011;121(8):334356.
33. Toda N, Mori K, Kasahara M, Koga K, Ishii A, Mori KP, et al. Deletion of connective tissue growth factor ameliorates peritoneal fibrosis by inhibiting angiogenesis and inflammation. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*. 2018;33(6):943-53.
34. Yoshino J, Patterson BW, Klein S. Adipose Tissue CTGF Expression is Associated with Adiposity and Insulin Resistance in Humans. *Obesity (Silver Spring, Md)*. 2019;27(6):957-62.

35. Colak Y, Senates E, Coskunpinar E, Oltulu YM, Zemheri E, Ozturk O, et al. Concentrations of connective tissue growth factor in patients with nonalcoholic fatty liver disease: association with liver fibrosis. *Dis Markers*. 2012;33(2):7783.
36. Wang PX, Ji YX, Zhang XJ, Zhao LP, Yan ZZ, Zhang P, et al. Targeting CASP8 and FADD-like apoptosis regulator ameliorates nonalcoholic steatohepatitis in mice and nonhuman primates. *Nature medicine*. 2017;23(4):439-49.
37. Liu Y, Xu W, Zhai T, You J, Chen Y. Silibinin ameliorates hepatic lipid accumulation and oxidative stress in mice with non-alcoholic steatohepatitis by regulating CFLAR-JNK pathway. *Acta pharmaceutica Sinica B*. 2019;9(4):745-57.
38. Liu Y, Yu Q, Chen Y. Effect of silibinin on CFLAR-JNK pathway in oleic acid-treated HepG2 cells. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*. 2018;108:716-23.
39. Federico A, Dallio M, Loguercio C. Silymarin/Silybin and Chronic Liver Disease: A Marriage of Many Years. *Molecules*. 2017;22(2).
40. Zeng L, Tang WJ, Yin JJ, Zhou BJ. Signal transductions and nonalcoholic fatty liver: a mini-review. *Int J Clin Exp Med*. 2014;7(7):1624-31.
41. Zhong Q, Pevzner SJ, Hao T, Wang Y, Mosca R, Menche J, et al. An inter-species protein-protein interaction network across vast evolutionary distance. *Molecular systems biology*. 2016;12(4):865-.
42. Li X, Stevens PD, Liu J, Yang H, Wang W, Wang C, et al. PHLPP is a negative regulator of RAF1, which reduces colorectal cancer cell motility and prevents tumor progression in mice. *Gastroenterology*. 2014;146(5):1301-12.e10.
43. Zhuravleva E, Tschopp O, Hemmings BA. Role of PKB/Akt in Liver Diseases. In: Dufour J-F, Clavien P-A, editors. *Signaling Pathways in Liver Diseases*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 243-59.
44. Manning BD, Cantley LC. AKT/PKB signaling: navigating downstream. *Cell*. 2007;129(7):1261-74.
45. Andreozzi F, Procopio C, Greco A, Mannino GC, Miele C, Raciti GA, et al. Increased levels of the Akt-specific phosphatase PH domain leucine-rich repeat protein phosphatase (PHLPP)-1 in obese participants are associated with insulin resistance. *Diabetologia*. 2011;54(7):1879-87.
46. Brognard J, Newton AC. PHLiPPing the switch on Akt and protein kinase C signaling. *Trends Endocrinol Metab*. 2008;19(6):223-30.
47. Matsuda S, Kobayashi M, Kitagishi Y. Roles for PI3K/AKT/PTEN Pathway in Cell Signaling of Nonalcoholic Fatty Liver Disease. *ISRN Endocrinol*. 2013;2013:472432-.
48. Stiles B, Wang Y, Stahl A, Bassilian S, Lee WP, Kim Y-J, et al. Liver-specific deletion of negative regulator Pten results in fatty liver and insulin hypersensitivity. *Proceedings of the National Academy of Sciences*. 2004;101(7):20827.
49. Peyrou M, Bourgoin L, Poher AL, Altirriba J, Maeder C, Caillon A, et al. Hepatic PTEN deficiency improves muscle insulin sensitivity and decreases adiposity in mice. *J Hepatol*. 2015;62(2):421-9.

50. Bankoglu EE, Tschopp O, Schmitt J, Burkard P, Jahn D, Geier A, et al. Role of PTEN in Oxidative Stress and DNA Damage in the Liver of Whole-Body Pten Haplodeficient Mice. *PLoS One*. 2016;11(11):e0166956.
51. Berlanga A, Guiu-Jurado E, Porras JA, Auguet T. Molecular pathways in non-alcoholic fatty liver disease. *Clin Exp Gastroenterol*. 2014;7:221-39.
52. Leung C, Herath CB, Jia Z, Andrikopoulos S, Brown BE, Davies MJ, et al. Dietary advanced glycation end-products aggravate non-alcoholic fatty liver disease. *World J Gastroenterol*. 2016;22(35):8026-40.
53. Hollenbach M. The Role of Glyoxalase-I (Glo-I), Advanced Glycation Endproducts (AGEs), and Their Receptor (RAGE) in Chronic Liver Disease and Hepatocellular Carcinoma (HCC). *Int J Mol Sci*. 2017;18(11):2466.
54. Ramasamy R, Yan SF, Schmidt AM. Receptor for AGE (RAGE): signaling mechanisms in the pathogenesis of diabetes and its complications. *Ann N Y Acad Sci*. 2011;1243:88-102.
55. Yamagishi S-i, Matsui T. Role of receptor for advanced glycation end products (RAGE) in liver disease. *Eur J Med Res*. 2015;20(1):15-.
56. Bongarzone S, Savickas V, Luzi F, Gee AD. Targeting the Receptor for Advanced Glycation Endproducts (RAGE): A Medicinal Chemistry Perspective. *Journal of Medicinal Chemistry*. 2017;60(17):7213-32.
57. Ramalingam L, Menikdiwela K, LeMieux M, Dufour JM, Kaur G, Kalupahana N, et al. The renin angiotensin system, oxidative stress and mitochondrial function in obesity and insulin resistance. *Biochimica et biophysica acta Molecular basis of disease*. 2017;1863(5):1106-14.
58. Xu F, Liu C, Zhou D, Zhang L. TGF- β /SMAD Pathway and Its Regulation in Hepatic Fibrosis. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*. 2016;64(3):157-67.
59. Qin G, Wang GZ, Guo DD, Bai RX, Wang M, Du SY. Deletion of Smad4 reduces hepatic inflammation and fibrogenesis during nonalcoholic steatohepatitis progression. 2018;19(5):301-13.
60. Yang L, Roh YS, Song J, Zhang B, Liu C, Loomba R, et al. Transforming growth factor beta signaling in hepatocytes participates in steatohepatitis through regulation of cell death and lipid metabolism in mice. *Hepatology (Baltimore, Md)*. 2014;59(2):483-95.
61. Lu JW, Liao CY, Yang WY, Lin YM, Jin SL, Wang HD, et al. Overexpression of endothelin 1 triggers hepatocarcinogenesis in zebrafish and promotes cell proliferation and migration through the AKT pathway. *PLoS One*. 2014;9(1):e85318.
62. Jin R, Krasinskas A, Le NA, Konomi J, Holzberg J, Romero R, et al. Association between plasminogen activator inhibitor-1 and severity of liver injury and cardiovascular risk in children with non-alcoholic fatty liver disease: PAI-1 and pediatric NAFLD. *Pediatric Obesity*. 2016;13.
63. Senagolage MD, Sommars MA, Ramachandran K, Futtner CR, Omura Y, Allred AL, et al. Loss of Transcriptional Repression by BCL6 Confers Insulin Sensitivity in the Setting of Obesity. *Cell Rep*. 2018;25(12):3283-98.e6.

64. LaPensee CR, Lin G, Dent AL, Schwartz J. Deficiency of the transcriptional repressor B cell lymphoma 6 (Bcl6) is accompanied by dysregulated lipid metabolism. *PloS one*. 2014;9(6):e97090-e.
65. Cardenas MG, Yu W, Beguelin W, Teater MR, Geng H, Goldstein RL, et al. Rationally designed BCL6 inhibitors target activated B cell diffuse large B cell lymphoma. *J Clin Invest*. 2016;126(9):3351-62.
66. Sommars MA, Ramachandran K, Senagolage MD, Futtner CR, Germain DM, Allred AL, et al. Dynamic repression by BCL6 controls the genome-wide liver response to fasting and steatosis. 2019;8.
67. Suzuki I, Yamauchi T, Onuma M, Nozaki S. Allopurinol, an inhibitor of uric acid synthesis—can it be used for the treatment of metabolic syndrome and related disorders? *Drugs Today (Barc)*. 2009;45(5):363-78.
68. Mokhles M, Salama S, Hossary Y, Badary O, Foad S. Impact of Treating Hyperuricemic NAFLD Patients with Allopurinol on Cytokeratin 18: A Pilot Study 2018.
69. Lombardi R, Pisano G, Fargion S. Role of Serum Uric Acid and Ferritin in the Development and Progression of NAFLD. *International journal of molecular sciences*. 2016;17(4):548-.
70. Cong X, Lu C, Huang X, Yang D, Cui X, Cai J, et al. Increased expression of glycinamide ribonucleotide transformylase is associated with a poor prognosis in hepatocellular carcinoma, and it promotes liver cancer cell proliferation. *Hum Pathol*. 2014;45(7):1370-8.
71. Liu X, Ding Z, Liu Y, Zhang J, Liu F, Wang X, et al. Glycinamide ribonucleotide formyl transferase is frequently overexpressed in glioma and critically regulates the proliferation of glioma cells. *Pathology, research and practice*. 2014;210(4):256-63.
72. Sookoian S, Pirola CJ. Alanine and aspartate aminotransferase and glutamine-cycling pathway: their roles in pathogenesis of metabolic syndrome. *World J Gastroenterol*. 2012;18(29):3775-81.
73. Simón J, Nunez M, Tussy PF, Barbier-Torres L, Ramos DF, Santos BG, et al. Glutaminase 1 targeting in nonalcoholic steatohepatitis. *J Hepatol*. 2018;68:S358-S9.
74. Miller RA, Shi Y, Lu W, Pirman DA, Jatkar A, Blatnik M, et al. Targeting hepatic glutaminase activity to ameliorate hyperglycemia. 2018;24(4):518-24.
75. Ampuero J, Ranchal I, Nunez D, Diaz-Herrero Mdel M, Maraver M, del Campo JA, et al. Metformin inhibits glutaminase activity and protects against hepatic encephalopathy. *PLoS One*. 2012;7(11):e49279.
76. Seiler N. Ornithine aminotransferase, a potential target for the treatment of hyperammonemias. *Current drug targets*. 2000;1(2):119-53.
77. Hewitt KN, Walker EA, Stewart PM. Minireview: hexose-6-phosphate dehydrogenase and redox control of 11 β -hydroxysteroid dehydrogenase type 1 activity. *Endocrinology*. 2005;146(6):2539-43.
78. Pacifico L, Andreoli GM, D'Avanzo M, De Mitri D, Pierimarchi P. Role of osteoprotegerin/receptor activator of nuclear factor kappa B/receptor activator of nuclear factor kappa B ligand axis in nonalcoholic fatty liver disease. *World J Gastroenterol*. 2018;24(19):2073-82.
79. Gruben N, Shiri-Sverdlov R, Koonen DP, Hofker MH. Nonalcoholic fatty liver disease: A main driver of insulin resistance or a dangerous liaison? *Biochimica et biophysica acta*. 2014;1842(11):2329-43.

80. Zhang RN, Pan Q, Zheng RD, Mi YQ, Shen F, Zhou D, et al. Genome-wide analysis of DNA methylation in human peripheral leukocytes identifies potential biomarkers of nonalcoholic fatty liver disease. *Int J Mol Med*. 2018;42(1):443-52.
81. Filip R, Radzki RP, Bieńko M. Novel insights into the relationship between nonalcoholic fatty liver disease and osteoporosis. *Clin Interv Aging*. 2018;13:1879-91.

Figures

Figure 1

PCA of Sample data (Black balls are NASH and Gray balls are Control samples) PCA mapping illustrates differential expression between normal and disease samples and confirms validity of chosen samples datasets for further transcription analysis.

Figure 2

Heat map of the 12 differentially expressed genes with $|\log_2\text{foldchange}| > 2$. Red represents upregulated DEGs and yellow represents down-regulated DEGs

Figure 3

Box plot for RMA versus MASS5 Normalization. Normalization was carried out using both RMA and Mas5 models; the results of the former, which were more median-centric, were accepted.

Figure 4

Visualization for Hub genes network by using the Gephi software. Color represents degree of nodes. The greener nodes are the more degree between nodes exists and the darker pink the less degree nodes the network has. The size of nodes shows closeness centrality, label color also indicates eigenvector centrality and label size represents betweenness centrality of nodes in PPI network.

Figure 5

Top KEGG pathway enrichment of total significant nodes of PPI network. Significant gene sets were submitted to Enrichr and the top 10 KEG pathways were retrieved according to higher combined score

Figure 6

KEGG pathway enrichment of hub nodes. Hub genes of each module were submitted to Enrichr and the top KEG pathways associated with hubgens were retrieved according to higher combined score

Figure 7

Functional enrichment analysis of hubs of each module. KEGG 2016 analysis were used to obtain significantly enriched KEGG complications. The X axis represents combined scores of each pathways and the Y axis presents top pathways which is enriched in 3 significant modules.

Figure 8

Protein-protein interaction of clusters. The figure shows how hub genes distributed in each module. The PPI network of hub genes consists of 3 significant and 2 small modules. Blue nodes are hub genes of module 1, green nodes are hub genes of Module 2, Purple nodes serves as hub genes of Module 3, dark green and orange are nodes of module 4 and module 5 in order.