



Comprehensive meta-analysis reveals distinct gene expression signatures of MASLD progression

Ignazio S Piras¹ , Johanna K DiStefano² 

Metabolic dysfunction-associated steatotic liver disease (MASLD) and its progressive form, metabolic dysfunction-associated steatohepatitis (MASH), pose significant risks of severe fibrosis, cirrhosis, and hepatocellular carcinoma. Despite their widespread prevalence, the molecular mechanisms underlying the development and progression of these common chronic hepatic conditions are not fully understood. Here, we conducted the most extensive meta-analysis of hepatic gene expression datasets from liver biopsy samples to date, integrating 10 RNA-sequencing and microarray datasets (1,058 samples). Using a random-effects meta-analysis model, we compared over 12,000 shared genes across datasets. We identified 685 genes differentially expressed in MASLD versus normal liver, 1,870 in MASH versus normal liver, and 3,284 in MASLD versus MASH. Integrating these results with genome-wide association studies and coexpression networks, we identified two functionally relevant, validated coexpression modules mainly driven by SMOC2, ITGBL1, LOXL1, MGP, SOD3, and TAT, HGD, SLC25A15, respectively, the latter not previously associated with MASLD and MASH. Our findings provide a comprehensive and robust analysis of hepatic gene expression alterations associated with MASLD and MASH and identify novel key drivers of MASLD progression.

DOI [10.26508/lsa.202302517](https://doi.org/10.26508/lsa.202302517) | Received 8 December 2023 | Revised 25 March 2024 | Accepted 26 March 2024 | Published online 2 April 2024

Introduction

Metabolic dysfunction-associated steatotic liver disease (MASLD), formerly known as nonalcoholic fatty liver disease (NAFLD), is a condition characterized by the accumulation of fat in the liver that can progress to metabolic dysfunction-associated steatohepatitis (MASH), a more severe form of the disease (1, 2). In general, hepatic steatosis attributable to metabolic dysfunction is considered relatively benign, but MASH is linked with greater morbidity and mortality, diminished quality of life, and substantial healthcare costs (3). At present, MASLD is estimated to affect ~30% of the global population (4, 5, 6), and this figure is projected to rise to 56% by the

year 2040 (7). Despite its substantial impact on public health, there are currently no approved pharmacological treatments for MASLD.

Intensive efforts to develop therapeutic agents for MASLD and MASH are presently underway. A comprehensive understanding of the molecular mechanisms driving MASH progression is an important component of this process; however, our current knowledge in this area remains incomplete. One approach to identify potential mechanisms driving the development of MASLD and progression to MASH involves the investigation of hepatic gene expression patterns, and many groups have conducted transcriptomic profiling experiments using liver biopsies obtained from patients spanning the MASLD spectrum (8, 9, 10, 11, 12, 13, 14, 15, 16, 17). Differences in gene expression have identified a core set of genes uniquely associated with advanced fibrosis (8) and discriminated among histological stages on the MASLD spectrum (16, 18, 19, 20). However, findings among these studies have not been consistent, which may reflect differences in methods used to measure gene expression, patient population characteristics, and comparator groups. Furthermore, most reports have been limited by modest sample sizes.

Meta-analysis of gene expression data can provide a more comprehensive and robust view of gene expression patterns than any individual study alone. In addition, this approach can help identify consistent and reproducible gene expression signatures across different studies, potentially leading to new insights into the underlying biology of complex diseases or other biological processes. Thus, to identify characteristics of hepatic gene expression changes associated with MASLD and obtain a deeper understanding of progression to MASH, we performed a meta-analysis of relevant gene expression datasets obtained from the Gene Expression Omnibus (GEO) data repository (21, 22) and sequencing reads archives. We included a total of 1,058 samples from 10 RNA-sequencing (RNA-Seq) and microarray datasets derived from liver tissue, applying strict quality controls and accounting for confounding factors. The integration of the different datasets allowed comparison of more than 12,000 shared genes that were analyzed using a random-effects model implemented in the *GeneMeta* algorithm, a meta-analytical workflow specific for RNA profiling data (23). Furthermore, we conducted a follow-up analysis integrating

¹Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ, USA

²Diabetes and Metabolic Disease Research Unit, Translational Genomics Research Institute, Phoenix, AZ, USA

Correspondence: ipiras@tgen.org; jdistefano@tgen.org

Table 1. Final sample sizes of all the datasets used in the meta-analysis.

Dataset	Original			After QC			Final sample size	Platform	Reference
	MASH	MASLD	CTL	MASH	MASLD	CTL			
GSE48452	17	9	28	16	8	28	52	Affymetrix microarray	(9)
GSE61260	24	23	62	24	23	61	108	Affymetrix microarray	(13)
GSE83452	104	44	0	104	43	0	147	Affymetrix microarray	(15)
GSE167523	47	51	0	47	50	0	97	Affymetrix microarray	(14)
GSE33814	12	19	13	12	19	13	44	Illumina microarray	(16)
GSE89632	19	20	24	19	20	24	63	Illumina microarray	(10)
GSE135251	155	51	10	155	51	10	216	RNA-Seq	(11)
GSE126848	16	15	26	16	15	26	57	RNA-Seq	(17)
GSE130970	42	36	0	41	34	0	75	RNA-Seq	(12)
PRJNA512027	105	50	36	105	50	36	191	RNA-Seq	(8)

the meta-analysis results with publicly available datasets from genome-wide association studies (GWAS) and generating coexpression networks from the same datasets included in the meta-analysis. Finally, we projected the relevant coexpression modules in a liver-specific Bayesian regulatory causal network to identify key drivers perturbing the disease-associated networks.

Results

Dataset selection

We performed an extensive search of the Gene Expression Omnibus database (GEO) using “NAFLD,” “NASH,” “steatosis,” “liver fibrosis,” “liver inflammation,” or “fatty liver” as key terms. Twelve datasets were retrieved. We then applied the following inclusion criteria: (1) expression data derived from microarray or RNA-seq; (2) use of RNA that was sourced from human liver tissue; (3) a case-control study design in which MASLD, MASH, and/or a control group (CTL: normal liver) were used; and (4) a total sample size >20. We identified nine datasets meeting these criteria (GEO IDs: [GSE135251](#), [GSE126848](#), [GSE130970](#), [GSE167523](#), [GSE83452](#), [GSE61260](#), [GSE48452](#), [GSE33814](#), and [GSE89632](#)), to which we added our previously published RNA-seq dataset (NCBI Bioproject Accession [PRJNA512027](#)) (8). Characteristics of these datasets, including sample sizes and post-quality control (QC) checks, are shown in [Table 1](#). These datasets formed the basis for the subsequent meta-analyses and coexpression network analysis ([Fig S1A](#)).

Differentially expressed genes (DEGs) in MASLD

The final dataset for the MASLD analysis comprised 516 total samples: 317 MASLD and 199 CTL. Through data harmonization across datasets, we identified 13,376 shared genes. Using a random-effects meta-analysis model, we observed 685 genes showing statistically significant differences in hepatic expression between individuals with MASLD and those with normal liver ([Table S1](#)). Of these DEGs, 360 were down-regulated and 325 were up-regulated in

those with MASLD. The 30 genes showing the strongest evidence for differential expression are shown in [Fig 1A](#). The five most significant genes were prolyl 4-hydroxylase subunit alpha 1 (*P4HA1*), Ras Association Domain Family Member 4 (*RASSF4*), acyl-CoA Dehydrogenase Short/Branched Chain (*ACADSB*), chromosome 11 open reading frame 54 (*C11orf54*), and transmembrane protein 45B (*TMEM45B*). To gain insights into the biological processes, molecular functions, and cellular components associated with this set of DEGs, we performed pathway analysis using the GO database. Our analysis identified 77 enriched functional classes ([Table S2](#)), and the top 10 biological processes for all DEGs, including those that were up- or down-regulated relative to the control group, are shown in [Fig 1B](#). Genes showing reduced expression in MASLD were predominantly associated with activin binding, activin receptor activity, and glycogen biosynthetic processes, whereas those with up-regulated expression in MASLD were associated with processes of exocytosis and ion transmembrane transporter activity.

DEGs in MASH

To determine gene expression patterns that may be specific to MASH, we compared gene expression between individuals with MASH ($n = 541$) and those with normal liver ($n = 199$). In a comparison of 13,476 shared genes between these two groups, we identified 1,870 genes exhibiting statistically significant differential expression. Among these genes, 1,005 were found to be up-regulated in MASH, whereas 865 were down-regulated ([Table S3](#); [Fig 2A](#)). The top DEGs were *P4HA1*, PRAME family member 10 (*PRAMEF10*), protein kinase AMP-activated catalytic subunit alpha 2 (*PRKAA2*), denticleless E3 ubiquitin protein ligase homolog (*DTL*), and exonuclease 1 (*EXO1*), as well as genes such as methionine adenosyltransferase 1A (*MAT1A*) (24), integrin subunit beta like 1 (*ITGBL1*) (25), and insulin-like growth factor binding protein 2 (*IGFBP2*) previously linked with MASLD (26). Pathway analysis revealed 213 significantly enriched GO classes ([Table S4](#)). The most significant classes were associated with sulfur and alpha-amino metabolic processes for the up-regulated genes in MASH, whereas the down-regulated genes were primarily associated with extracellular matrix organization ([Fig 2B](#)).

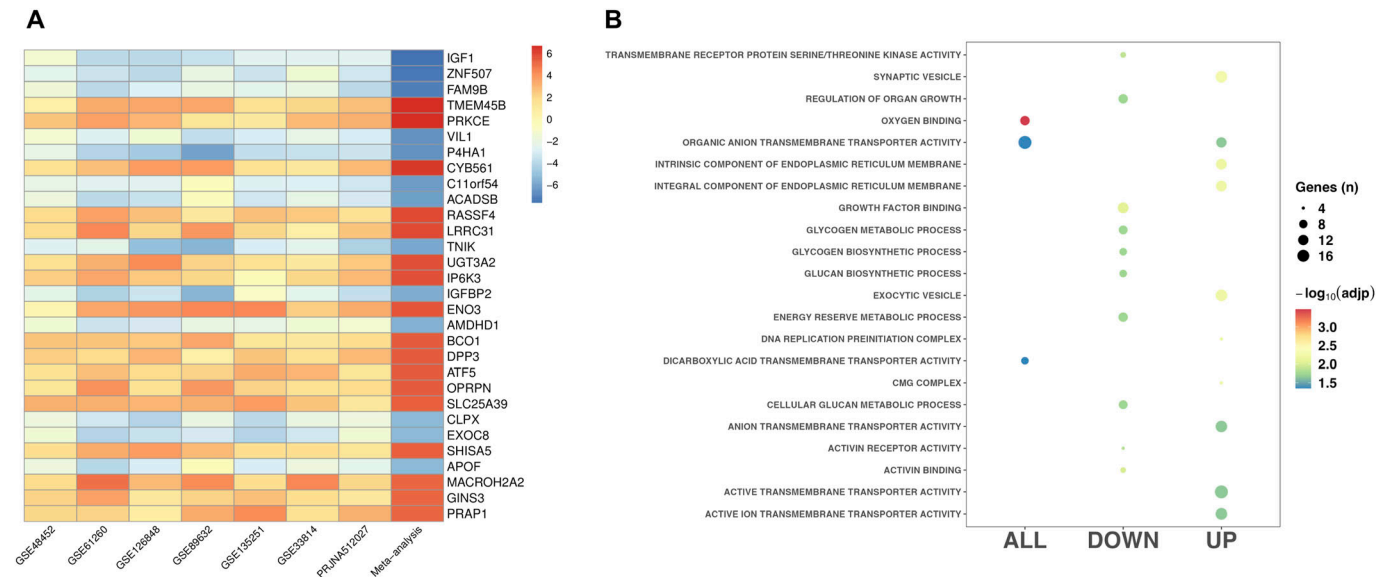


Figure 1. Differentially expressed genes in MASLD. (A) The most significant differentially expressed genes identified in the MASLD versus CTL meta-analysis. (B) Top GO functional classes (FDR < 0.05) for the differentially expressed genes from the MASLD versus CTL meta-analysis.

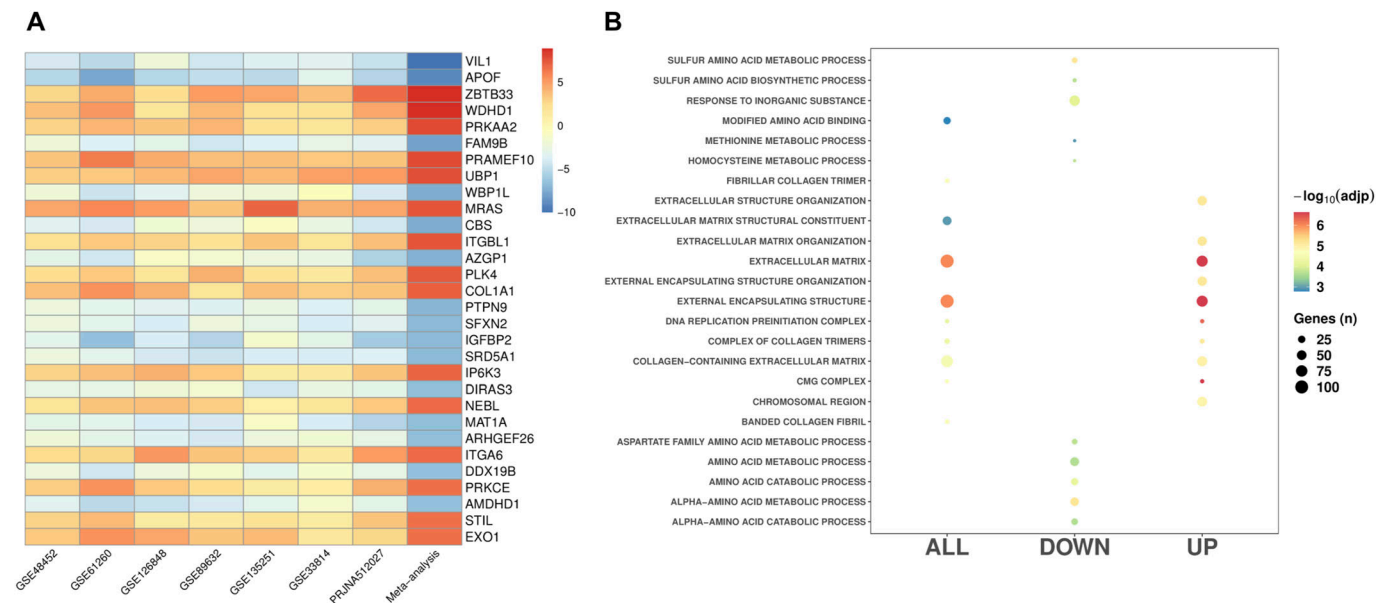


Figure 2. Differentially expressed genes in MASH. (A) The most significant differentially expressed genes identified in the MASH versus CTL meta-analysis. (B) Top GO functional classes (FDR < 0.05) for the differentially expressed genes from the MASH versus CTL meta-analysis.

Gene expression patterns associated with MASLD progression

We next sought to identify genes involved in MASLD progression by comparing hepatic gene expression values between MASLD and MASH samples. As shown in Table 1, 317 MASLD and 541 MASH samples comprised the final dataset. We observed 12,817 genes shared among these datasets, similar to the number of shared genes identified in the MASLD analysis. However, the random-effects meta-analysis revealed 3,284 genes that were differentially expressed between MASH and MASLD (1,850 up- and 1,434

down-regulated in MASH) (Table S5), which is substantially greater than the MASLD analysis. The genes showing the most significant evidence for differential expression are shown in Fig 3A. Pathway analysis of the 3,284 DEGs revealed 684 significantly enriched functional classes (Table S6). Genes showing up-regulation in MASH relative to MASLD were associated with the extracellular matrix and chemotaxis, whereas those down-regulated in MASH were mainly linked to the catabolic processes of organic compounds, carboxylic acids, and alpha-amino acids (Fig 3B).

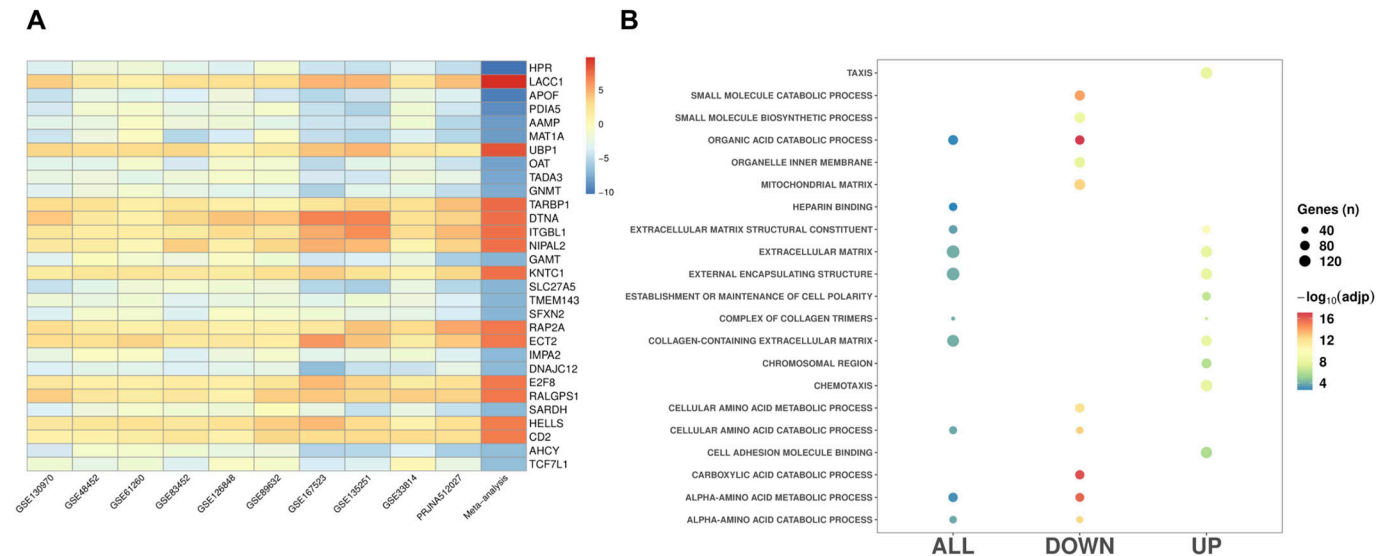


Figure 3. Gene expression patterns associated with MASLD progression. (A) The most significant differentially expressed genes identified in the MASH versus MASLD meta-analysis. (B) Top GO functional classes (FDR < 0.05) for the differentially expressed genes from the MASH versus MASLD meta-analysis.

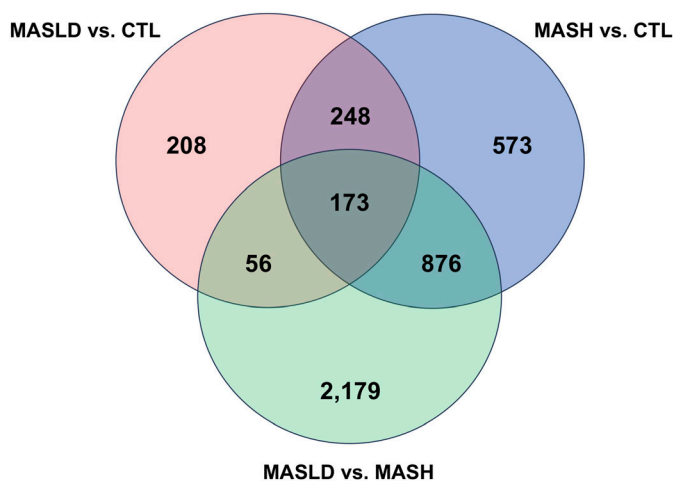


Figure 4. Overlap of differentially expressed genes detected across the three meta-analyses.

Overlap and shared genes among meta-analyses

When comparing the results from the three meta-analyses, we observed an overlap of 173 genes. In pairwise comparisons of the meta-analysis data, we identified 421 shared genes between the MASLD versus CTL and MASH versus CTL analyses, 1,049 shared genes between the MASH versus MASLD and the MASH versus CTL meta-analyses, and 229 shared genes between MASH versus MASLD and the MASLD versus CTL meta-analyses (Fig 4). The lists of these overlapping genes are presented in Table S7.

Enrichment of GWAS and transcriptome-wide association study (TWAS) genes in the MASLD and MASH meta-analyses

We used four large MASLD GWAS (see the Materials and Methods section) for the MAGMA analysis and detected 141 significant signals encompassing a total of 127 unique genes (Table S8). Manhattan plots are shown in Figs S2, S3, S4, and S5. We used this list of genes to run hypergeometric enrichment. Although no significant enrichment was detected (Table S9), we did observe overlap of 47 signals with 36 unique genes identified in the meta-analysis (Table S10). To identify genetic variants associated with gene expression, we conducted a TWAS analysis referencing liver tissue based on four large GWAS studies comprising MASLD patients and unaffected controls (27, 28, 29, 30). These GWAS had a total of 5,187, 439, 82, and 306 SNPs that were significant at the genome-wide level ($P < 5.0 \times 10^{-8}$). Significant findings from the TWAS were obtained only from the study by reference (29), with 83 genes showing significant associations (Table S11, Fig S6) and one from reference (27) (Table S12, Fig S7). To further investigate the relevance of these 84 significant TWAS genes ($\text{adj-}p < 0.05$), we conducted a hypergeometric enrichment analysis across the meta-analysis. We observed an enrichment of TWAS genes in the comparison between MASH and CTL samples, approaching significance ($P = 0.034$; $P\text{-adj} = 0.103$) (Table S13). Overall, five genes overlapped with the MASLD versus CTL meta-analysis, 15 genes with the MASH versus CTL meta-analysis, and 18 genes with the MASLD versus MASH meta-analysis. All the genes were from the TWAS conducted on the reference (29) GWAS data. The Venn diagram in Fig S8 depicts this overlap, and the complete list of overlapping genes can be found in Table S14. Finally, eight of the TWAS genes overlapping with the

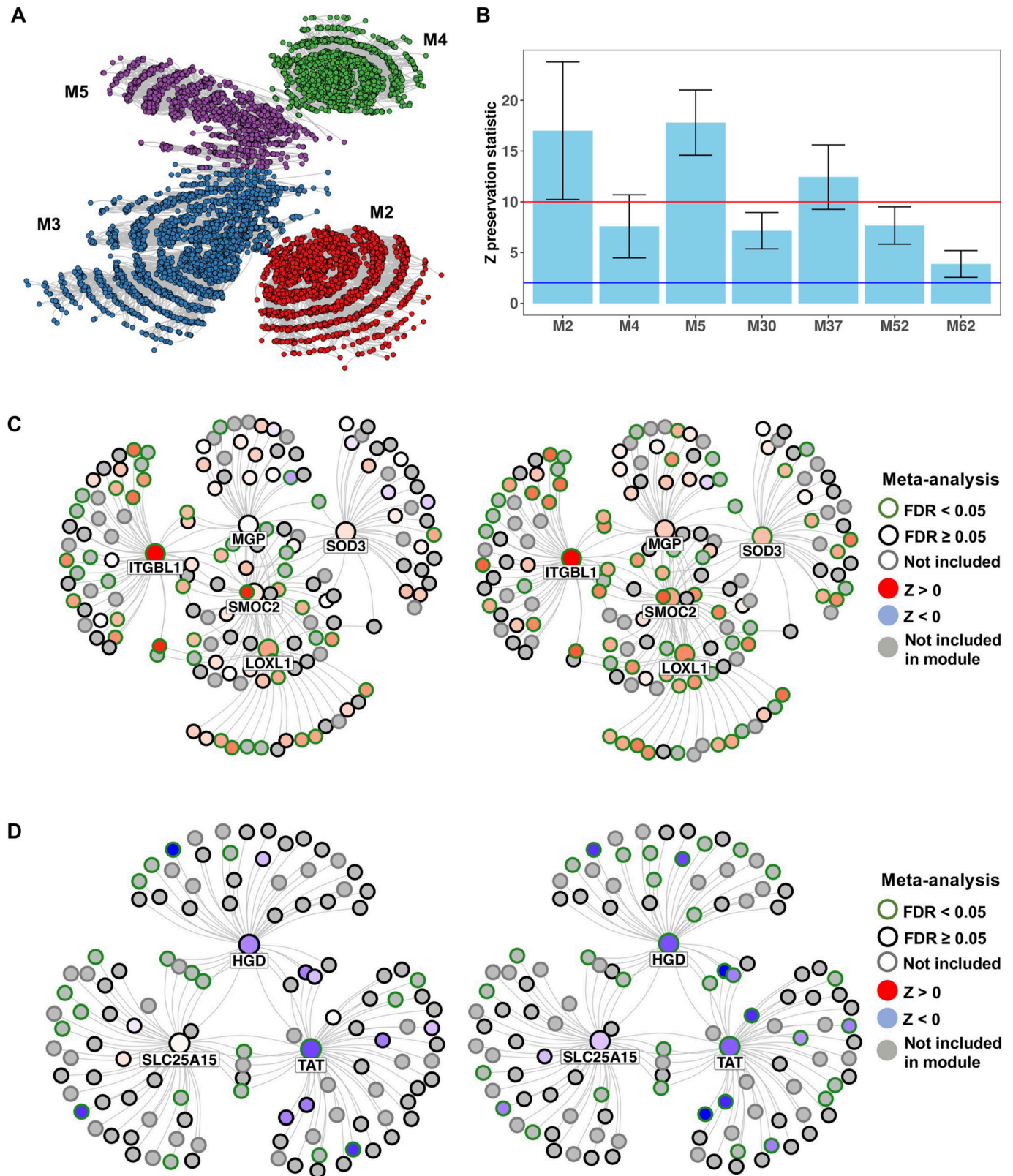


Figure 5. Coexpression, module preservation, and key driver analyses.

(A) Comprehensive coexpression network color-coded by top-level modules M2, M3, M4, and M5. (B) Average preservation statistics across six validation datasets for the seven functionally relevant modules associated with MASH and MASLD. The blue and red lines indicate the preservation statistics cutoffs for moderate ($2 \leq Z < 10$) and strong ($Z \geq 10$) module preservation, respectively. (C) Bayesian causal subnetwork showcasing the top five key drivers along with their neighbors in module M5: MASH

meta-analysis were also significant in the MAGMA analysis (*L3MBTL3*, *GGT1*, *GSTT2B*, *SLC25A19*, *EPHA2*, *DLG5*, *MRPS7*, and *KLHL18*) (Table S14).

Coexpression analysis reveals differentially expressed network modules in MASLD and MASH

We first analyzed [PRJNA512027](#) as a discovery dataset and included 8,262 variable genes, which led to the detection of 103 significantly coexpressed modules ($P < 0.01$ and comprising 50 or more genes). The entire coexpression network with the four top-level modules (M2, M3, M4, and M5) highlighted is depicted in [Fig 5A](#). We extracted the eigengenes and conducted a differential expression analysis comparing MASLD versus CTL, MASH versus CTL, and MASH versus MASLD. In the comparison of MASH versus CTL, we identified 49 significant differentially expressed modules, and in MASH versus MASLD, we found 69 significant differentially expressed modules (Table S15). A total of 49 modules were differentially expressed in both MASH versus CTL and MASH versus MASLD comparisons. However, no differentially expressed modules were detected in the MASLD versus CTL comparison.

M2 and M5 were among the top-level modules with the highest significance (down-regulated and up-regulated, respectively, in MASH versus CTL). In addition, M2 was significantly down-regulated in MASH versus MASLD, whereas M4 and M5 were significantly up-regulated in MASH versus MASLD. In the analysis of modules other than the top-level ones (i.e., M2–M5), we detected M52, M53, M37, and M62 in MASH versus CTL, and M52, M53, M37, M62, and M30 in MASH versus MASLD. The GO analysis of the significantly expressed top modules revealed significant GO enrichment in 55 modules, accounting for a total of 3,859 GO functional classes (Table S16). Among the top-level modules, M2 was enriched for ribosomal, oxidative phosphorylation, and ATP metabolic processes (hub gene: *HPN*); M4 for ion and gate channel activity (hub gene: *KRT77*) and M5 for extracellular matrix (hub gene: *SDC4*).

All functionally relevant modules were preserved across datasets

We assessed the preservation of the seven functionally relevant modules M2, M4, M5, M30, M37, M52, and M62 across datasets [GSE48452](#), [GSE61260](#), [GSE126848](#), [GSE89632](#), [GSE135251](#), and [GSE33814](#). After averaging the preservation statistics across these datasets, we observed that all modules exhibited moderate to strong preservation in all datasets, with standard errors also falling in this range ([Fig 5B](#)). Specifically, modules M2, M5, and M37 showed strong average preservation, whereas all the other modules exhibited moderate preservation. Modules M2 and M4 showed no preservation in [GSE33814](#), whereas module M62 showed no preservation in datasets [GSE33814](#) and [GSE126848](#) ([Fig S9](#)).

For validating module differential expression, we conducted a meta-analysis across the replication datasets. We designated modules as “fully validated” if they were statistically significant and

exhibited concordant $\log_2FC$ direction, and as “partially validated” if they were not statistically significant but showed concordant $\log_2FC$ direction. We fully validated the differential expression of M52 and M5 in MASH versus CTL and partially validated module M62 (Table S17 (A) and [Fig S10A](#)). In addition, we fully validated the differential expression of modules M52, M5, M62, and M2 in MASH versus MASLD and partially validated module M30 in MASH versus MASLD (Table S17 (B) and [Fig S10B](#)). These results indicate that modules M2, M5, M52, and M62 are coexpression networks that are consistently and significantly associated with MASH and MASLD progression.

Identification of key drivers in the coexpression modules

We conducted a key driver analysis on the five fully and partially validated modules (M52, M5, M62, M2, and M30), identifying a total of 181 significant key drivers located within the coexpression modules (false discovery rate [FDR] < 0.05) (Table S18). The most significant key drivers (FDR $< 2.2 \times 10^{-26}$) were all located in module M5 and included SPARC-related modular calcium binding 2 (*SMOC2*), *ITGBL1*, lysyl oxidase like 1 (*LOXL1*), matrix Gla protein (*MGP*), and superoxide dismutase 3 (*SOD3*) ([Fig 5C](#)). Notably, *ITGBL1* and *LOXL1* were significantly overexpressed in the MASH versus CTL meta-analysis, whereas *SMOC2*, *ITGBL1*, *LOXL1*, and *SOD3* were significantly overexpressed in the MASH versus MASLD meta-analysis ([Fig 5C](#)). Module M52, which was fully validated in both comparisons, yielded three significant key drivers: tyrosine aminotransferase (*TAT*), homogentisate 1,2-dioxygenase (*HGD*), and solute carrier family 25 member 15 (*SLC25A15*) ([Fig 5D](#)). *TAT* was significantly down-regulated in MASH versus CTL and significantly down-regulated in MASH versus MASLD, whereas *HGD* was significantly down-regulated in MASH versus MASLD. In contrast, the other fully validated module, M62, did not exhibit any significant key drivers. The causal networks for the partially validated modules M2 (top key driver: ISG15 ubiquitin like modifier [*ISG15*]) and M30 (top key driver: CCAAT enhancer binding protein delta [*CEBPD*]) are depicted in [Figs S11A and B](#) and [S12A and B](#), respectively. *ISG15* was overexpressed in MASH versus CTL and in MASH versus MASLD, albeit not significantly ([Fig S11A and B](#)). Lastly, *CEBPD* was significantly down-regulated in MASH versus CTL and MASH versus MASLD ([Fig S12A and B](#)).

Discussion

In this study, we present the outcomes of an extensive meta-analysis of hepatic gene expression data for MASLD and MASH, representing the most comprehensive analysis conducted to date. Our key discoveries include the identification of distinct gene expression patterns that distinguish MASLD, MASH, and healthy liver tissue. In addition, we validated two differentially expressed coexpression modules and identified novel key drivers with significant regulatory potential within these coexpression networks. Collectively, these findings provide new insight into potential liver-

versus CTL and MASH versus MASLD. (D) Bayesian causal subnetwork showcasing the significant key drivers along with their neighbors in module M52: MASH versus CTL and MASH versus MASLD.

specific molecular mechanisms underlying the development and progression of MASLD.

To the best of our knowledge, there have only been two previously published meta-analyses investigating hepatic gene expression in MASLD. The first study used microarray-based data from seven GEO datasets, encompassing 137 MASLD samples, and identified a 218-gene signature associated with MASLD (31). Unlike the current work, this analysis correlated gene expression with clinical variables rather than disease status, analyzed data individually without merging for statistical analysis, and did not investigate hub genes or key drivers within coexpression networks. Consequently, there was minimal methodological overlap between the two investigations. The results from the current work thus build upon these findings by incorporating a significantly larger sample size—nearly 10 times greater—merging all datasets for statistical analysis and conducting association analyses for both MASLD and MASH diagnosis.

In the second study (32), a comprehensive analysis of 12 datasets comprising a total of 812 samples was performed to identify a gene expression signature specific to MASH. Although there were seven datasets that overlapped with our study, several notable differences emerged between the two analyses. First, the primary objective of the published analysis was to identify biomarkers associated with MASH, whereas our focus was on unraveling gene expression networks that could provide insights into key pathological mechanisms. Second, the published meta-analysis included datasets from pediatric MASLD studies, which may introduce potential confounding factors when combined with adult samples. Third, the previous analytical design involved segregating the datasets into Discovery (N = 309) and Validation (N = 503) groups, whereas we opted to integrate all datasets for our analysis. It is worth mentioning that most datasets within the Discovery group comprised fewer than 20 MASLD or MASH samples, which could impact the statistical power of the analysis.

Interestingly, some notable distinctions emerged in our results when comparing the DEGs identified in the MASLD and MASH meta-analyses. First, there was a substantial contrast in the number of DEGs between the MASH and CTL meta-analysis (N = 1,870) relative to the MASLD versus CTL analyses (N = 685). Moreover, the MASH versus MASLD meta-analysis identified 3,284 DEGs. The difference in the number of DEGs may reflect underlying biological variability between MASLD and MASH, suggesting that these two conditions, while related, have distinct transcriptomic signatures and potentially different underlying mechanisms. Alternatively, the greater number of DEGs in the MASH analysis might indicate that MASH is a more complex and heterogeneous condition with a wider range of gene expression changes.

Second, the affected pathways differed depending on the analysis. In the MASLD analysis, the most significantly altered pathways included activin binding and receptor activity, glycogen biosynthesis, exocytosis, and ion transmembrane transporter activity. These results corroborate previous studies. For instance, dysregulated glycogen metabolism is recognized as a potential risk factor for MASLD (33), as are abnormal serum levels of activin (34). Exocytosis processes may potentially be linked to extracellular vesicles, which have been associated with the development of both whole-body and hepatic insulin resistance, as well as steatosis in

the context of MASLD (35). In contrast, the MASH analysis detected functional classes associated with sulfur and alpha-amino metabolic processes and extracellular matrix organization. The DEGs from the MASH versus MASLD comparison were associated with the extracellular matrix, chemotaxis, catabolic processes of organic compounds, carboxylic acids, and alpha-amino acids. Interestingly, the same MASH-associated pathways were also identified in the coexpression analysis. The dysregulation in affected pathways again suggests that MASH and MASLD have different underlying disease mechanisms. For example, the involvement of pathways related to sulfur and alpha-amino metabolic processes in MASH might indicate a role for oxidative stress and amino acid metabolism in the pathogenesis of MASH, which is typically characterized by inflammation and fibrosis. In contrast, the pathways related to activin binding and receptor activity in MASLD may highlight the importance of certain signaling pathways in the earlier stages of liver fat accumulation.

We conducted a key driver analysis on the functionally relevant coexpression modules to identify genes with a disproportionately significant impact on the regulation of other genes within the fully validated M5, M52, and M62 modules. Whereas M62 exhibited no significant key drivers, both the M5 and M52 modules yielded noteworthy results. The M5 module was found to be significantly up-regulated in comparisons of MASH versus CTL and MASH versus MASLD and was enriched for “extracellular matrix” GO functional classes, consistent with the main gene meta-analysis. This finding is consistent with the development of hepatic fibrosis, primarily as a result of abnormal expression and accumulation of extracellular matrix proteins in the liver (36). In addition, we identified five significant key drivers with potentially critical regulatory roles. One of these genes, *SMOC2*, a member of the SPARC family of matricellular proteins, displayed increased hepatic expression in individuals with MASLD and mice fed a high-fat diet (HFD) (37). Notably, *SMOC2*-knockout mice exhibited protection against liver fibrosis and reduced hepatic inflammation induced by HFD (37). Larsen et al (38) recently observed elevated hepatic and plasma levels of *SMOC2* in individuals with MASH compared to those without MASLD and showed that *SMOC2* is primarily expressed by hepatic stellate cells, which play a pivotal role in fibrogenesis. *SMOC2* levels were also elevated in hepatocellular carcinoma tissue relative to normal tissue, and *SMOC2* overexpression promoted hepatocellular carcinoma cell proliferation (39).

In addition to *SMOC2*, the remaining key drivers identified in the M5 module have previously been linked with MASLD. In individuals with chronic hepatitis B, *ITGBL1*, which promotes cell migration (40), has been shown to regulate fibrogenesis (25) and is part of a six-gene signature predictive of cirrhosis (41). In a weighted gene coexpression network analysis (WGCNA) focused on susceptibility gene modules related to immune cells in MASLD, *ITGBL1* was among the hub genes linked to immune infiltration, fibrosis progression, and activity score (42). *LOXL1*, a member of the lysyl oxidase (LOX) family of enzymes involved in collagen and elastin crosslinking, has also been linked to liver cirrhosis (43, 44). Studies have demonstrated that down-regulation of *LOXL1* can slow disease progression (44), and HSC-specific *LOXL1* knockout in a mouse model of non-obese NASH attenuated liver steatosis, inflammation, and fibrosis, suggesting that *LOXL1* may be involved in HSC activation and

fibrogenesis (45). *SOD3*, which encodes an antioxidative enzyme with various functions (46), was found to promote HSC activation and fibrogenesis when deficient (47), whereas its overexpression in HFD-fed mice blocked the obesity, hepatic steatosis, and insulin resistance typically induced by this diet (48). MGP, an extracellular matrix protein that inhibits calcification (49), has only recently been implicated in MASLD. Research by Hui et al (50) showed that hepatic *Mgp* expression increased in tandem with MASH progression in a mouse model and was significantly correlated with fibrosis severity in humans with MASH (50). *Mgp* was highly expressed in HSC and identified as a key driver for liver fibrosis through network modeling, associated with the regulators of fibrosis, *Loxl1*, *Smoc2*, and *Itgb11* (50).

The M52 module was significantly down-regulated in the MASH versus CTL and MASH versus MASLD comparisons and was enriched for alpha-amino acid metabolic process, consistent with the main gene meta-analysis. Altered levels of circulating amino acid levels have been observed in MASLD and MASH patients, along with differences in fatty acids and vitamins (51). *TAT*, *HGD*, and *SLC25A15* were identified as significant key drivers for this module. *TAT*, significantly down-regulated in both MASH meta-analyses, encodes a hepatic enzyme that catalyzes the conversion of tyrosine to 4-hydroxyphenylpyruvate. Mutations in *HGD* and *SLC25A15* lead to the development of alkaptonuria (52) and hyperornithinaemia-hyperammonaemia-homocitrullinuria syndrome (53), respectively. Unlike the key drivers identified for the M5 module, there is currently no evidence connecting these genes to the development or progression of MASLD. These genes might therefore represent novel candidates for functional studies investigating the molecular features of MASH and MASLD associated with amino acid-related metabolic dysfunction. The relevance of these results is strengthened by the consistent patterns found across several datasets.

Our comprehensive study design and bioinformatics workflow enabled us to identify robust and consistent signals at the network level across different datasets, pinpointing key genes that might provide insights into new therapeutic targets and further define the molecular landscape of MASLD and MASH. Despite the strengths of our rigorous approach, we acknowledge certain limitations. First, while RNA-sequencing provides a comprehensive assessment of a tissue's transcriptome, integrating microarray studies into the meta-analysis, with their more limited genome coverage, presents a risk of missing significant genes. This occurs because the analysis depends on genes that are common to all datasets. Nevertheless, we adopted this approach to include a wider range of datasets, thereby enhancing the robustness of our findings. Second, it's important to note that we did not conduct any experimental validation of the detected key drivers and their perturbed networks. Although consistent signals were observed, further investigation is necessary to confirm the regulatory role of these key drivers in both in vitro and in vivo models, as well as determine their potential effect on MASLD and MASH pathology.

Our study provides a comprehensive and robust view of hepatic gene expression changes associated with MASLD and MASH. The variations in the number of DEGs detected in the individual MASH and MASLD analyses suggest that these two conditions are not only phenotypically different, but also exhibit distinct molecular

profiles. Further research into these differences can potentially lead to a deeper understanding of the conditions and improved clinical management strategies. Our findings underscore the importance of meta-analysis in elucidating complex disease processes and highlight the need for additional investigations to validate and expand upon these results.

Materials and Methods

Data preprocessing

The data selection workflow is illustrated in Fig S1A. Criteria and results of the search are provided in the results section. We executed the preprocessing of raw data using separate pipelines for RNA-seq and microarray data. The details of the workflow are illustrated in Fig S2B. For RNA-seq data (GSE135251, GSE126848, GSE130970), we obtained raw data files from sequencing read archives and performed alignment with the GRCh38 human genome reference using *kallisto* 0.46.1 (54). We imported transcript-level abundance using Entrez Gene annotations (55) and estimated counts and transcript lengths with the *tximport* package (56) to obtain a matrix of average transcript length, weighted by sample-specific transcript abundance estimates, to counterbalance different expressions of gene-level counts. We then excluded genes with fewer than five total counts across all samples and applied a variance-stabilizing transformation via the *vsd* function from DESeq2 (57), which transforms the count data and provides an approximately homoscedastic matrix of values. Principal Component Analysis was used to identify and remove outlier samples, defined as those exceeding the cutoff of ± 4 standard deviations from the mean on at least one of the top two principal components. Raw counts were normalized with the *voom* method (58), and surrogate variable analysis was conducted via the *sva* function with the *leek* method from the *sva* R package (59). The derived surrogate variables were used to adjust the *voom*-normalized expression values using the *removeBatchEffect* function from *limma* (60). Gene annotations were standardized across datasets to HGNC (HUGO Gene Nomenclature Committee) symbols using the *R-BiomaRt* package (61).

For the preprocessing of Affymetrix microarray data (GSE167523, GSE83452, GSE61260, GSE48452), we obtained raw data (CEL files) from GEO and normalized the data using the Robust Multi-Array Average algorithm (62), as implemented in the *oligo* R package (63). We used the *arrayQualityMetrics* R package (64) to evaluate data quality with respect to reproducibility, outliers, and signal-to-noise ratio. We generated and analyzed several visualizations, including heatmaps depicting inter-array expression distances, Principal Component Analysis plots, and MA-plots (using log-intensity ratios and log-intensity averages), which enabled us to identify patterns, trends, and potential outliers within the data. Samples were classified as outliers in two of the three metrics in the initial round of QC runs from the dataset and were subsequently removed from further analysis. After outlier removal, data were once again normalized starting from the raw data, excluding genes in the lower 25th percentile of average expression and adjusting for surrogate

variables, thus mirroring the workflow applied to the RNA-seq data. Gene annotations were standardized as described above. For Illumina microarray datasets (GSE33814, GSE89632), we downloaded the matrix of normalized expression values from GEO and applied the same workflow as described for Affymetrix data, excluding the normalization step (Fig S1B).

Gene expression meta-analysis

We used the methodology of Choi et al (23) to generate a comprehensive ranked gene list based on the FDR associated with each gene. The input consisted of the *voom*-normalized (for RNA-seq) and RMA-normalized matrices (for microarrays), adjusted for surrogate variables. We executed the analysis using the R package *GeneMeta* setting a random-effects meta-analysis model to account for heterogeneity across studies and standardizing data using Z-score transformation (function: “Zscore”). The FDR for each gene was generated using the “ZscoreFDR” function, with 50,000 permutations factored into the calculations. We considered an FDR < 0.05 to be statistically significant evidence for differential gene expression. Subsequently, we performed pathway analysis with the DEGs, with reference to the Gene Ontology (GO) database using the *enrichGO* function, as implemented in the *clusterProfiler* package. P-values were adjusted using the FDR method, and GO processes with an FDR < 0.05 were considered statistically significant.

Enrichment with GWAS and TWAS results

We used four large GWAS to investigate the enrichment of MASLD-associated genes across our meta-analysis results. In addition, using the same GWAS, we investigated the relationship between genetically regulated gene expression and MASLD, conducting a TWAS imputing expression values from the GTEx v8 liver expression data. The first GWAS involved the UK Biobank (UKB) cohort, which comprised 28,396 individuals with MASLD and 108,652 healthy individuals (29). MASLD status was determined using ICD codes, whereas a liver fat percent <5% by abdominal MRI defined controls. The second GWAS (28), also using the UKB cohort, included 4,761 MASLD cases and 373,227 unaffected controls. In this study, the authors assigned MASLD status using the diagnostic codes recommended by recent consensus guidelines. The third GWAS (30) included 1,106 MASLD cases and 8,571 controls and histological data from liver tissue in 235 available participants from the electronic medical records and genomics network. The samples were selected using a natural language processing algorithm billing codes, text queries, laboratory values, and medication records. Finally, the fourth GWAS (27) included 1,483 MASLD cases, and 17,781 controls histologically characterized. We obtained GWAS summary statistics for all the studies from the GWAS catalogue (Accession numbers: GCST90094908, GCST90054782, GCST008471, and GCST90011885, respectively). To ensure data quality, we performed data cleaning using the “*munge_sumstat*” function from *Linkage Disequilibrium Score Regression (ldsc)* software (65). During the data cleaning process, we conducted quality control checks and applied filters to include only relevant SNPs. Specifically, we retained SNPs with an imputation quality (INFO) greater than 0.9, a minimum allele

frequency higher than 1%, and P-association values between 0 and 1. We also removed insertion/deletion variants (INDELs) and SNPs with duplicated “rs” numbers.

Subsequently, GWAS summary statistics were analyzed using the Multi-marker Analysis of GenoMic Annotation (MAGMA) method, which provides gene-level statistics using a multiple regression approach to incorporate linkage disequilibrium (LD) information between markers and to detect multi-marker effects (66). P-values were adjusted for multiple testing using the Bonferroni method, accounting for the number of genes tested. We then used the *FUSION* software to conduct the TWAS (67). We imputed expression values using liver data models from GTEx v8 (European reference data), which provide a relationship measure between an individual’s genotype and gene expression levels, thereby capturing the *cis*-acting genetic effects on gene expression. Using *FUSION*, we computed gene weights that estimate the effects of individual SNPs on gene expression. These gene expression models were then used to evaluate whether the predicted gene expression levels were associated with the phenotype. To account for multiple testing, we adjusted TWAS P-values using the FDR method.

We also performed an enrichment analysis to assess the functional relevance of the genes identified through MAGMA and TWAS (FDR < 0.05). In this analysis, we used the TWAS and MAGMA gene lists as the gene sets of interest and the DEGs from our meta-analysis as the candidate list. The reference set was the comprehensive list of genes included in each meta-analysis. To perform the enrichment analysis, we used the “enrichment” function from the bc3net R package. This function allows for the calculation of hypergeometric statistics, which assess the enrichment of the MAGMA and TWAS genes within the candidate list of DEGs.

Coexpression analysis, key driver analysis, and module validation

Coexpression analysis was conducted using the Multiscale Embedded Gene Expression Network Analysis (MEGENA) algorithm, which offers a robust alternative to existing coexpression network clustering methods, including WGCNA (68). We selected the dataset PRJNA512027 as a discovery cohort as it encompasses MASH (n = 104), MASLD (n = 49), and CTL (n = 36) samples, and includes the largest sample sizes among the datasets available. We used other datasets that also included MASH, MASLD, and CTL for module preservation analysis. The matrix of SV-adjusted expression values was filtered to include only the top 50% of genes with the highest median absolute deviation. Network generation was executed using the MEGENA R package (68) based on the following workflow. Initially, we calculated signed pairwise gene correlations using Pearson’s method with 1,000 permutations, retaining correlations that were significant at the 5% FDR level (function: *calculate_correlation*). Significantly correlated gene pairs (FDR < 0.05) were ranked and iteratively tested for planarity, leading to the development of a planar filtered network using the planar maximally filtered graph technique (function: *calculate_PFN*). Subsequently, we conducted a multiscale clustering analysis to identify coexpression modules at varying network scale topologies and their respective hub genes (function: *do.MEGENA*). Coexpression modules deemed significant (with a permuted *P* < 0.01 and module of 50 genes or more) were carried forward for further analysis. Next, we

extracted module eigengenes (the first principal component of the gene module) using the function *moduleEigengenes* from the WGCNA R package (69). Pairwise differential expression between diagnostic groups was computed using the *limma* R package (60). Modules with significant associations were annotated for GO functional classes, following the same workflow we adopted for DEGs.

To validate our findings, we conducted a module preservation analysis on the functionally relevant modules. We used the “*modulePreservation*” function from the WGCNA R package to derive the Z-score preservation statistics. Modules with a Z-score greater than 10 were deemed to have strong preservation, those with a Z-score >2 and <10 indicated moderate preservation, and those with a Z-score <2 showed no preservation. This preservation analysis was applied across datasets that included MASLD, MASH, and CTL samples (GSE48452, GSE61260, GSE126848, GSE89632, GSE135251, and GSE33814). Subsequently, we extracted the eigenvalues of the functionally relevant modules from the six validation datasets and then evaluated their differential expression, focusing on the comparisons: MASH versus CTL and MASH versus MASLD. The *P*-values and the log₂ FC obtained from these validation datasets were combined through a meta-analytical approach based on the Fisher’s weighted test (70). The input was the unadjusted *P*-values derived from the differential module expression analysis. Given that the Fisher Z weighted test requires one-tailed *P*-values, we converted the two-tailed nominal *P*-values to one-tailed *P*-value. If the Log₂FC was greater than zero, the formula used was: $p1Tailed = p2Tailed / 2$. Otherwise, the formula was: $p1Tailed = 1 - (p2Tailed / 2)$. The uncorrected *P*-values were then weighted by the sample sizes of the datasets and combined with the *combine.test* function with the “*z.transform*” option, as part of the *survcomp* R package (71). We calculated the average log₂ FC by weighting for sample size for each study using the “*weighted.mean*” R functions.

We applied weighted Key Driver Analysis to identify central hub genes within functionally relevant modules using the mergeomics R package. Hub genes are characterized by a high number of strong correlations with other genes in the network; however, this does not necessarily indicate a causal relationship. For this analysis, we used a liver-specific Bayesian regulatory network as a reference (72). The parameters set for the analysis were as follows: “search depth” was set to 1, “edge type” was specified as undirected, “minhuboverlap” was set at 0.33 and “edge.factor” was 0. This algorithm prioritizes module genes based on their ability to regulate other genes in the module. It uses a pre-built causal regulatory network, identifying genes whose neighbors predominantly belong to the same coexpression module. These “key driver” genes, predicted as top regulators within a coexpression module, are targets for potential novel treatment designed to prevent the progression.

Supplementary Information

Supplementary Information is available at <https://doi.org/10.26508/lsa.202302517>.

Acknowledgements

This research was supported by a grant from the National Institute of Diabetes and Digestive and Kidney Diseases (R01DK127015).

Author Contributions

IS Piras: formal analysis, investigation, methodology, and writing—original draft, review, and editing.

JK DiStefano: conceptualization, data curation, supervision, investigation, project administration, and writing—original draft, review, and editing.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

References

- Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, Harrison SA, Brunt EM, Sanyal AJ (2018) The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American association for the study of liver diseases. *Hepatology* 67: 328–357. doi:10.1002/hep.29367
- Vernon G, Baranova A, Younossi ZM (2011) Systematic review: The epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. *Aliment Pharmacol Ther* 34: 274–285. doi:10.1111/j.1365-2036.2011.04724.x
- Younossi ZM, Blissett D, Blissett R, Henry L, Stepanova M, Younossi Y, Racila A, Hunt S, Beckerman R (2016) The economic and clinical burden of nonalcoholic fatty liver disease in the United States and Europe. *Hepatology* 64: 1577–1586. doi:10.1002/hep.28785
- Le MH, Yeo YH, Li X, Li J, Zou B, Wu Y, Ye Q, Huang DQ, Zhao C, Zhang J, et al (2022) 2019 global NAFLD prevalence: A systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 20: 2809–2817.e28. doi:10.1016/j.cgh.2021.12.002
- Riazi K, Azhari H, Charette JH, Underwood FE, King JA, Afshar EE, Swain MG, Congly SE, Kaplan GG, Shaheen AA (2022) The prevalence and incidence of NAFLD worldwide: A systematic review and meta-analysis. *Lancet Gastroenterol Hepatol* 7: 851–861. doi:10.1016/S2468-1253(22)00165-0
- Younossi ZM, Golabi P, Paik JM, Henry A, Van Dongen C, Henry L (2023) The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): A systematic review. *Hepatology* 77: 1335–1347. doi:10.1097/HEP.0000000000000004
- Le MH, Yeo YH, Zou B, Barnett S, Henry L, Cheung R, Nguyen MH (2022) Forecasted 2040 global prevalence of nonalcoholic fatty liver disease using hierarchical bayesian approach. *Clin Mol Hepatol* 28: 841–850. doi:10.3350/cmh.2022.0239
- Gerhard GS, Legendre C, Still CD, Chu X, Petrick A, DiStefano JK (2018) Transcriptomic profiling of obesity-related nonalcoholic steatohepatitis reveals a core set of fibrosis-specific genes. *J Endocr Soc* 2: 710–726. doi:10.1210/js.2018-00122
- Ahrens M, Ammerpohl O, von Schonfels W, Kolarova J, Bens S, Itzel T, Teufel A, Herrmann A, Brosch M, Hinrichsen H, et al (2013) DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab* 18: 296–302. doi:10.1016/j.cmet.2013.07.004
- Arendt BM, Comelli EM, Ma DW, Lou W, Teterina A, Kim T, Fung SK, Wong DKH, McGilvray I, Fischer SE, et al (2015) Altered hepatic gene expression

- in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. *Hepatology* 61: 1565–1578. doi:[10.1002/hep.27695](https://doi.org/10.1002/hep.27695)
11. Govaere O, Cockell S, Tiniakos D, Queen R, Younes R, Vacca M, Alexander L, Ravaioli F, Palmer J, Petta S, et al (2020) Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci Transl Med* 12: eaba4448. doi:[10.1126/scitranslmed.aba4448](https://doi.org/10.1126/scitranslmed.aba4448)
 12. Hoang SA, Oseini A, Feaver RE, Cole BK, Asgharpour A, Vincent R, Siddiqui M, Lawson MJ, Day NC, Taylor JM, et al (2019) Gene expression predicts histological severity and reveals distinct molecular profiles of nonalcoholic fatty liver disease. *Sci Rep* 9: 12541. doi:[10.1038/s41598-019-48746-5](https://doi.org/10.1038/s41598-019-48746-5)
 13. Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schonfels W, Ahrens M, Heits N, Bell JT, Tsai PC, Spector TD, et al (2014) Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci U S A* 111: 15538–15543. doi:[10.1073/pnas.1412759111](https://doi.org/10.1073/pnas.1412759111)
 14. Kozumi K, Kodama T, Murai H, Sakane S, Govaere O, Cockell S, Motooka D, Kakita N, Yamada Y, Kondo Y, et al (2021) Transcriptomics identify thrombospondin-2 as a biomarker for NASH and advanced liver fibrosis. *Hepatology* 74: 2452–2466. doi:[10.1002/hep.31995](https://doi.org/10.1002/hep.31995)
 15. Lefebvre P, Lalloyer F, Bauge E, Pawlak M, Gheeraert C, Dehondt H, Vanhoutte J, Woitrain E, Hennuyer N, Mazuy C, et al (2017) Interspecies NASH disease activity whole-genome profiling identifies a fibrogenic role of PPAR α -regulated dermatopontin. *JCI Insight* 2: e92264. doi:[10.1172/jci.insight.92264](https://doi.org/10.1172/jci.insight.92264)
 16. Starmann J, Falth M, Spindelbock W, Lanz KL, Lackner C, Zatloukal K, Trauner M, Sultmann H (2012) Gene expression profiling unravels cancer-related hepatic molecular signatures in steatohepatitis but not in steatosis. *PLoS One* 7: e46584. doi:[10.1371/journal.pone.0046584](https://doi.org/10.1371/journal.pone.0046584)
 17. Suppli MP, Rigbolt KTG, Veidal SS, Heebøll S, Eriksen PL, Demant M, Bagger JI, Nielsen JC, Oró D, Thrane SW, et al (2019) Hepatic transcriptome signatures in patients with varying degrees of nonalcoholic fatty liver disease compared with healthy normal-weight individuals. *Am J Physiol Gastrointest Liver Physiol* 316: G462–G472. doi:[10.1152/ajpgi.00358.2018](https://doi.org/10.1152/ajpgi.00358.2018)
 18. Wruck W, Kashofer K, Rehman S, Daskalaki A, Berg D, Gralka E, Jozefczuk J, Drews K, Pandey V, Regenbrecht C, et al (2015) Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes. *Sci Data* 2: 150068. doi:[10.1038/sdata.2015.68](https://doi.org/10.1038/sdata.2015.68)
 19. Moylan CA, Pang H, Dellinger A, Suzuki A, Garrett ME, Guy CD, Murphy SK, Ashley-Koch AE, Choi SS, Michelotti GA, et al (2014) Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology* 59: 471–482. doi:[10.1002/hep.26661](https://doi.org/10.1002/hep.26661)
 20. Asselah T, Bieche I, Laurendeau I, Paradis V, Vidaud D, Degott C, Martinot M, Bedossa P, Valla D, Vidaud M, et al (2005) Liver gene expression signature of mild fibrosis in patients with chronic hepatitis C. *Gastroenterology* 129: 2064–2075. doi:[10.1053/j.gastro.2005.09.010](https://doi.org/10.1053/j.gastro.2005.09.010)
 21. Edgar R, Domrachev M, Lash AE (2002) Gene expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30: 207–210. doi:[10.1093/nar/30.1.207](https://doi.org/10.1093/nar/30.1.207)
 22. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al (2013) NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Res* 41: D991–D995. doi:[10.1093/nar/gks1193](https://doi.org/10.1093/nar/gks1193)
 23. Choi JK, Yu U, Kim S, Yoo OJ (2003) Combining multiple microarray studies and modeling interstudy variation. *Bioinformatics* 19: i84–i90. doi:[10.1093/bioinformatics/btg1010](https://doi.org/10.1093/bioinformatics/btg1010)
 24. Saenz de Urturi D, Buque X, Porteiro B, Folguez C, Mora A, Delgado TC, Prieto-Fernández E, Olaizola P, Gómez-Santos B, Apodaka-Biguri M, et al (2022) Methionine adenosyltransferase 1a antisense oligonucleotides activate the liver-brown adipose tissue axis preventing obesity and associated hepatosteatosis. *Nat Commun* 13: 1096. doi:[10.1038/s41467-022-28749-z](https://doi.org/10.1038/s41467-022-28749-z)
 25. Wang M, Gong Q, Zhang J, Chen L, Zhang Z, Lu L, Yu D, Han Y, Zhang D, Chen P, et al (2017) Characterization of gene expression profiles in HBV-related liver fibrosis patients and identification of ITGBL1 as a key regulator of fibrogenesis. *Sci Rep* 7: 43446. doi:[10.1038/srep43446](https://doi.org/10.1038/srep43446)
 26. Chen X, Tang Y, Chen S, Ling W, Wang Q (2021) IGFBP-2 as a biomarker in NAFLD improves hepatic steatosis: An integrated bioinformatics and experimental study. *Endocr Connect* 10: 1315–1325. doi:[10.1530/EC-21-0353](https://doi.org/10.1530/EC-21-0353)
 27. Anstee QM, Darlay R, Cockell S, Meroni M, Govaere O, Tiniakos D, Burt AD, Bedossa P, Palmer J, Liu YL, et al (2020) Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort(☆). *J Hepatol* 73: 505–515. doi:[10.1016/j.jhep.2020.04.003](https://doi.org/10.1016/j.jhep.2020.04.003)
 28. Fairfield CJ, Drake TM, Pius R, Bretherick AD, Campbell A, Clark DW, Fallowfield JA, Hayward C, Henderson NC, Joshi PK, et al (2022) Genome-wide association study of NAFLD using electronic health records. *Hepatol Commun* 6: 297–308. doi:[10.1002/hep4.1805](https://doi.org/10.1002/hep4.1805)
 29. Miao Z, Garske KM, Pan DZ, Koka A, Kaminska D, Mannisto V, Sinsheimer JS, Pihlajamäki J, Pajukanta P (2022) Identification of 90 NAFLD GWAS loci and establishment of NAFLD PRS and causal role of NAFLD in coronary artery disease. *HGG Adv* 3: 100056. doi:[10.1016/j.xhgg.2021.100056](https://doi.org/10.1016/j.xhgg.2021.100056)
 30. Namjou B, Lingren T, Huang Y, Parameswaran S, Cobb BL, Stanaway IB, Connolly JJ, Mentch FD, Benoit B, Niu X, et al (2019) GWAS and enrichment analyses of non-alcoholic fatty liver disease identify new trait-associated genes and pathways across eMERGE Network. *BMC Med* 17: 135. doi:[10.1186/s12916-019-1364-z](https://doi.org/10.1186/s12916-019-1364-z)
 31. Ryaboshapkina M, Hammar M (2017) Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Sci Rep* 7: 12361. doi:[10.1038/s41598-017-10930-w](https://doi.org/10.1038/s41598-017-10930-w)
 32. Hasin-Brumshtein Y, Sakaram S, Khatri P, He YD, Sweeney TE (2022) A robust gene expression signature for NASH in liver expression data. *Sci Rep* 12: 2571. doi:[10.1038/s41598-022-06512-0](https://doi.org/10.1038/s41598-022-06512-0)
 33. Allende DS, Gawrieh S, Cummings OW, Belt P, Wilson L, Van Natta M, Behling CA, Carpenter D, Gill RM, Kleiner DE, et al (2021) Glycogenosis is common in nonalcoholic fatty liver disease and is independently associated with ballooning, but lower steatosis and lower fibrosis. *Liver Int* 41: 996–1011. doi:[10.1111/liv.14773](https://doi.org/10.1111/liv.14773)
 34. Yndestad A, Haukeland JW, Dahl TB, Bjørø K, Gladhaug IP, Berge C, Damås JK, Haaland T, Løberg EM, Linnestad P, et al (2009) A complex role of activin A in non-alcoholic fatty liver disease. *Am J Gastroenterol* 104: 2196–2205. doi:[10.1038/ajg.2009.318](https://doi.org/10.1038/ajg.2009.318)
 35. Dorairaj V, Sulaiman SA, Abu N, Abdul Murad NA (2020) Extracellular vesicles in the development of the non-alcoholic fatty liver disease: An update. *Biomolecules* 10: 1494. doi:[10.3390/biom10111494](https://doi.org/10.3390/biom10111494)
 36. Tacke F, Weiskirchen R (2021) Non-alcoholic fatty liver disease (NAFLD)/non-alcoholic steatohepatitis (NASH)-related liver fibrosis: Mechanisms, treatment and prevention. *Ann Transl Med* 9: 729. doi:[10.21037/atm-20-4354](https://doi.org/10.21037/atm-20-4354)
 37. Yuting Y, Lifeng F, Qiwei H (2019) Secreted modular calcium-binding protein 2 promotes high fat diet (HFD)-induced hepatic steatosis through enhancing lipid deposition, fibrosis and inflammation via targeting TGF- β 1. *Biochem Biophys Res Commun* 509: 48–55. doi:[10.1016/j.bbrc.2018.12.006](https://doi.org/10.1016/j.bbrc.2018.12.006)
 38. Larsen FT, Hansen D, Terkelsen MK, Bendixen SM, Avolio F, Wernberg CW, Lauridsen MM, Grønkaer LL, Jacobsen BG, Klinggaard EG, et al (2023) Stellate cell expression of SPARC-related modular calcium-binding protein 2 is associated with human non-alcoholic fatty liver disease severity. *JHEP Rep* 5: 100615. doi:[10.1016/j.jhep.2022.100615](https://doi.org/10.1016/j.jhep.2022.100615)
 39. Su JR, Kuai JH, Li YQ (2016) Smoc2 potentiates proliferation of hepatocellular carcinoma cells via promotion of cell cycle progression. *World J Gastroenterol* 22: 10053–10063. doi:[10.3748/wjg.v22.i45.10053](https://doi.org/10.3748/wjg.v22.i45.10053)
 40. Jang DG, Kwon KY, Song EK, Park TJ (2022) Integrin β -like 1 protein (ITGBL1) promotes cell migration by preferentially inhibiting integrin-

- ECM binding at the trailing edge. *Genes Genomics* 44: 405–413. doi:[10.1007/s13258-021-01204-x](https://doi.org/10.1007/s13258-021-01204-x)
41. Xu MY, Qu Y, Li Z, Li F, Xiao CY, Lu LG (2016) A 6 gene signature identifies the risk of developing cirrhosis in patients with chronic hepatitis B. *Front Biosci (Landmark Ed)* 21: 479–486. doi:[10.2741/4403](https://doi.org/10.2741/4403)
 42. Wang Z, Xia Y, Pan Y, Zhang L, Tang F, Yu X, Yang Z, Wang D, Yang L, Jia J, et al (2023) Weighted gene co-expression network analysis of immune infiltration in nonalcoholic fatty liver disease. *Endocr Metab Immune Disord Drug Targets* 23: 1173–1185. doi:[10.2174/1871530323666221208105720](https://doi.org/10.2174/1871530323666221208105720)
 43. Greene AG, Eivers SB, Dervan EWJ, O'Brien CJ, Wallace DM (2020) Lysyl oxidase like 1: Biological roles and regulation. *Exp Eye Res* 193: 107975. doi:[10.1016/j.exer.2020.107975](https://doi.org/10.1016/j.exer.2020.107975)
 44. Zhao W, Yang A, Chen W, Wang P, Liu T, Cong M, Xu A, Yan X, Jia J, You H (2018) Inhibition of lysyl oxidase-like 1 (LOXL1) expression arrests liver fibrosis progression in cirrhosis by reducing elastin crosslinking. *Biochim Biophys Acta Mol Basis Dis* 1864: 1129–1137. doi:[10.1016/j.bbadis.2018.01.019](https://doi.org/10.1016/j.bbadis.2018.01.019)
 45. Yang A, Yan X, Fan X, Shi Y, Huang T, Li W, Chen W, Jia J, You H (2021) Hepatic stellate cells-specific LOXL1 deficiency abrogates hepatic inflammation, fibrosis, and corrects lipid metabolic abnormalities in non-obese NASH mice. *Hepatol Int* 15: 1122–1135. doi:[10.1007/s12072-021-10210-w](https://doi.org/10.1007/s12072-021-10210-w)
 46. Sah SK, Agrahari G, Kim TY (2020) Insights into superoxide dismutase 3 in regulating biological and functional properties of mesenchymal stem cells. *Cell Biosci* 10: 22. doi:[10.1186/s13578-020-00386-3](https://doi.org/10.1186/s13578-020-00386-3)
 47. Sun YL, Bai T, Zhou L, Zhu RT, Wang WJ, Liang RP, Li J, Zhang CX, Gou JJ (2021) SOD3 deficiency induces liver fibrosis by promoting hepatic stellate cell activation and epithelial-mesenchymal transition. *J Cell Physiol* 236: 4313–4329. doi:[10.1002/jcp.30174](https://doi.org/10.1002/jcp.30174)
 48. Cui R, Gao M, Qu S, Liu D (2014) Overexpression of superoxide dismutase 3 gene blocks high-fat diet-induced obesity, fatty liver and insulin resistance. *Gene Ther* 21: 840–848. doi:[10.1038/gt.2014.64](https://doi.org/10.1038/gt.2014.64)
 49. Cranenburg ECM, Vermeer C, Schurgers LJ (2008) Matrix Gla-protein: The calcification inhibitor in need of vitamin K. *Thromb Haemost* 100: 593–603. doi:[10.1160/th08-02-0087](https://doi.org/10.1160/th08-02-0087)
 50. Hui ST, Gong L, Swichkow C, Blencowe M, Kaminska D, Diamante G, Pan C, Dalsania M, French SW, Magyar CE, et al (2023) Role of matrix Gla protein in transforming growth factor- β signaling and nonalcoholic steatohepatitis in mice. *Cell Mol Gastroenterol Hepatol* 16: 943–960. doi:[10.1016/j.jcmgh.2023.08.007](https://doi.org/10.1016/j.jcmgh.2023.08.007)
 51. Piras C, Noto A, Ibba L, Deidda M, Fanos V, Muntoni S, Leoni VP, Atzori L (2021) Contribution of metabolomics to the understanding of NAFLD and NASH syndromes: A systematic review. *Metabolites* 11: 694. doi:[10.3390/metabo11100694](https://doi.org/10.3390/metabo11100694)
 52. Vilboux T, Kayser M, Introne W, Suwannarat P, Bernardini I, Fischer R, O'Brien K, Kleta R, Huizing M, Gahl WA (2009) Mutation spectrum of homogentisic acid oxidase (HGD) in alkaptonuria. *Hum Mutat* 30: 1611–1619. doi:[10.1002/humu.21120](https://doi.org/10.1002/humu.21120)
 53. Camacho JA, Obie C, Biery B, Goodman BK, Hu CA, Almashanu S, Steel G, Casey R, Lambert M, Mitchell GA, et al (1999) Hyperornithinaemia-hyperammonaemia-homocitrullinuria syndrome is caused by mutations in a gene encoding a mitochondrial ornithine transporter. *Nat Genet* 22: 151–158. doi:[10.1038/9658](https://doi.org/10.1038/9658)
 54. Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34: 525–527. doi:[10.1038/nbt.3519](https://doi.org/10.1038/nbt.3519)
 55. Maglott D, Ostell J, Pruitt KD, Tatusova T (2007) Entrez gene: Gene-centered information at NCBI. *Nucleic Acids Res* 35: D26–D31. doi:[10.1093/nar/gkl993](https://doi.org/10.1093/nar/gkl993)
 56. Sonesson C, Love MI, Robinson MD (2015) Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Res* 4: 1521. doi:[10.12688/f1000research.7563.2](https://doi.org/10.12688/f1000research.7563.2)
 57. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550. doi:[10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8)
 58. Law CW, Chen Y, Shi W, Smyth GK (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 15: R29. doi:[10.1186/gb-2014-15-2-r29](https://doi.org/10.1186/gb-2014-15-2-r29)
 59. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28: 882–883. doi:[10.1093/bioinformatics/bts034](https://doi.org/10.1093/bioinformatics/bts034)
 60. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47. doi:[10.1093/nar/gkv007](https://doi.org/10.1093/nar/gkv007)
 61. Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* 4: 1184–1191. doi:[10.1038/nprot.2009.97](https://doi.org/10.1038/nprot.2009.97)
 62. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4: 249–264. doi:[10.1093/biostatistics/4.2.249](https://doi.org/10.1093/biostatistics/4.2.249)
 63. Carvalho BS, Irizarry RA (2010) A framework for oligonucleotide microarray preprocessing. *Bioinformatics* 26: 2363–2367. doi:[10.1093/bioinformatics/btq431](https://doi.org/10.1093/bioinformatics/btq431)
 64. Kauffmann A, Gentleman R, Huber W (2009) arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics* 25: 415–416. doi:[10.1093/bioinformatics/btn647](https://doi.org/10.1093/bioinformatics/btn647)
 65. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BMSchizophrenia Working Group of the Psychiatric Genomics Consortium, (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 47: 291–295. doi:[10.1038/ng.3211](https://doi.org/10.1038/ng.3211)
 66. de Leeuw CA, Mooij JM, Heskes T, Posthuma D (2015) MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 11: e1004219. doi:[10.1371/journal.pcbi.1004219](https://doi.org/10.1371/journal.pcbi.1004219)
 67. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, Jansen R, de Geus EJC, Boomsma DI, Wright FA, et al (2016) Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* 48: 245–252. doi:[10.1038/ng.3506](https://doi.org/10.1038/ng.3506)
 68. Song WM, Zhang B (2015) Multiscale embedded gene co-expression network analysis. *PLoS Comput Biol* 11: e1004574. doi:[10.1371/journal.pcbi.1004574](https://doi.org/10.1371/journal.pcbi.1004574)
 69. Langfelder P, Horvath S (2008) WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9: 559. doi:[10.1186/1471-2105-9-559](https://doi.org/10.1186/1471-2105-9-559)
 70. Zaykin DV (2011) Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *J Evol Biol* 24: 1836–1841. doi:[10.1111/j.1420-9101.2011.02297.x](https://doi.org/10.1111/j.1420-9101.2011.02297.x)
 71. Schroder MS, Culhane AC, Quackenbush J, Haibe-Kains B (2011) survcomp: An R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics* 27: 3206–3208. doi:[10.1093/bioinformatics/btr511](https://doi.org/10.1093/bioinformatics/btr511)
 72. Ding J, Blencowe M, Nghiem T, Ha SM, Chen YW, Li G, Yang X (2021) Mergeomics 2.0: A web server for multi-omics data integration to elucidate disease networks and predict therapeutics. *Nucleic Acids Res* 49: W375–W387. doi:[10.1093/nar/gkab405](https://doi.org/10.1093/nar/gkab405)



License: This article is available under a Creative Commons License (Attribution 4.0 International, as described at <https://creativecommons.org/licenses/by/4.0/>).