Combined stem cell and predictive models reveal flavin cofactors as targets in metabolic liver dysfunction

Julian Weihs[&1], Fatima Baldo[&2,3,4], Alessandra Cardinali[&2,5], Gehad Youssef[&2,3,4], Katarzyna Ludwik[6], Nils Haep[7,8], Peter Tang[7], Pavitra Kumar[9], Cornelius Engelmann[8,9], Susanna Quach[1], Mijuna Meindl[1], Martin Kucklick[10,11], Susanne Engelmann[10,11], Bruno Chillian[12], Michael Rothe[13], David Meierhofer[14], Isabella Lurje[9,15], Linda Hammerich[9], Prakash Ramachandran[16], Timothy J. Kendall[16], Jonathan A. Fallowfield[16], Harald Stachelscheid[6], Igor Sauer[7], Frank Tacke[9], Philip Bufler[1], Christian Hudert[1], Namshik Han[2,3,4*], Milad Rezvani[1,8,17,18, 19*]

1. Charité Universitätsmedizin Berlin, Department of Pediatrics, Division of Gastroenterology, Nephrology and Metabolic Medicine, Augustenburger Platz 1, 13353 Berlin
2. Milner Therapeutics Institute, University of Cambridge
3. Cambridge Centre for AI in Medicine, University of Cambridge
4. Cambridge Stem Cell Institute, University of Cambridge
5. LifeArc, Lynton House, 7-12 Tavistock Square, London, WC1H 9LT
6. Berlin Institute of Health (BIH) at Charité – Universitätsmedizin Berlin, BIH Core Unit pluripotent Stem Cells and Organoids (CUSCO), Berlin, Germany
7. Department of Surgery, Campus Virchow-Klinikum, Experimental Surgery, CCM|CVK, Charité – Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany
8. Clinician Scientist Program, Berlin Institute of Health at Charité (BIH), Anna-Louisa-Karsch-Str. 2, 10178 Berlin, Germany
9. Department of Hepatology and Gastroenterology, Charité, Universitätsmedizin Berlin, Campus Virchow-Klinikum (CVK) and Campus Charité Mitte (CCM), Berlin, Germany
10. Institute for Microbiology, Technische Universität Braunschweig, Braunschweig, Germany
11. Microbial Proteomics, Helmholtz Center for Infection Research, Braunschweig, Germany
12. TRI Thinking Research Instruments GmbH
13. Lipidomix, 13125 Berlin, Germany.
14. Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, 14195 Berlin, Germany
15. Department of General, Visceral and Transplantation Surgery, Heidelberg University Hospital, Heidelberg, Germany
16. Centre for Inflammation Research, Institute for Regeneration and Repair, University of Edinburgh, Edinburgh, UK
17. Berlin Institute of Health (BIH), 10178 Berlin, Germany
18. Berlin Institute of Health (BIH) at Charité-Universitätsmedizin Berlin, BIH Center for Regenerative Therapies (BCRT), Berlin, Germany
19. Division of Gastroenterology, Hepatology and Nutrition, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA

[&]These authors contributed equally
*Corresponding authors, milad.rezvani@charite.de; nh417@cam.ac.uk

Drug discovery for multifactorial diseases like metabolic dysfunction-associated steatotic liver disease (MASLD) remains challenging due to inadequate models and untargeted drug screenings. We combined stem-cell-based modeling with computational drug predictions identifying flavin pathways as therapeutic targets in MASLD. For disease stage-specific discovery, we established a MASLD testing model, compounding metabolic triggers to intensify mitochondrial crisis. *In vitro* injuries included adipo- and myokines, immune cell co-culture, and genomic risk factors. Benchmarking experiments revealed similarities with advanced human MASLD. To query therapeutic compounds, protein-protein-interaction networks, weighted gene co-expression, and knowledge graph-based analyses independently predicted flavin adenine dinucleotide (FAD) as an anti-MASLD factor. Dysregulated flavoproteomes *in vitro* and *in vivo*–in pediatric and adult MASLD patients– supported our flavin network-focused strategy. We established therapeutic FAD concentrations to mitigate metabolic injury and fibro-inflammation in human multicellular liver organoids and other assays. We enhanced therapeutic FAD effects through genetic mitochondrial biogenic augmentation and identified orally available flavo-active compounds—including Aspirin—restoring mitochondrial respiration. Our study demonstrates how integrating stem cell-derived disease modeling with computed drug predictions can expedite therapeutic discovery.

Disease modeling of multifactorial diseases remains challenging, and the target discovery process is inefficient[1]. For example, cancer cells are suitable for large-scale drug screenings but lack physiological

relevance. In contrast, induced pluripotent stem cells (iPSCs) can generate versatile disease models that clarify pharmacological targets in difficult-to-culture cell types[2]. However, the limited scalability of iPSC-based systems poses challenges to effectively screen large compound libraries, which themselves have restricted read-out dimensions. Computationally enhanced drug discovery aims to overcome these limitations by prioritizing compounds for experimental validation. Machine learning (ML)-based algorithms harness clinical and molecular data, to predict therapeutic efficacy of drug compounds, enabling more focused candidate lists. However, the potential of ML- and iPSC-based systems, along with integrating clinical cohort data to expedite experimental validation feedback loops, has not been fully realized.
Metabolic dysfunction-associated steatotic liver disease (MASLD) is a prime example of a multifactorial disease with high unmet need. It affects ~30% of the global population[3] and is projected to increase[4]. In MASLD, nutrient overload, extrahepatic signaling, and genetic susceptibility (e.g. in *PNPLA3*) cause hepatic steatosis and lipotoxic cell stress. Hepatocyte injury activates both innate and adaptive immunity[5], progressively leading to fibrogenesis, potentially developing to cirrhosis and liver failure. Although the disease sequence is understood[6], the multiple injury mechanisms hamper drug discovery. MASLD models typically expose hepatocytes to free fatty acids (FFAs), emphasizing early-stage steatosis[7]. While relevant, this method mainly tests anti-steatotic drugs, overlooking subsequent pathological drivers like metabolic crisis and mitochondrial dysfunction. Additionally, MASLD is systemic, involving signals from adipose or muscle tissue, not captured by FFAs. Notably, adipokines like resistin (Res)[8] and myokines, including myostatin (Myo)[9–11], have been linked to MASLD. However, their precise impact on human hepatocytes remains unclear. Thus, incorporating various MASLD triggers may reveal relevant pathological features *in vitro* for drug discovery, while tailoring drug predictions to disease-relevant pathologies enhances the likelihood of successful therapeutic development.
Here, we integrate multi-modal human stem cell-based modeling with ML-based drug predictions and extensive benchmarking against clinical cohorts to identify therapeutic targets for a pressing need in hepatology – MASLD drug development. We establish a drug testing framework using iPSC-derived hepatocytes (iHeps) exposed to a compounded sequence of MASLD-related triggers. Furthermore, we utilize three independent computational drug prediction pipelines that analyze clinical data tailored to MASLD hallmark pathologies, as observed *in vitro*. By integrating these efforts and other pipeline innovations, including e.g. label-free iHep phenotyping by ML-based morphological imaging analysis, we identified the therapeutic potential of targeting flavin network biology for MASLD. Our functional validation confirmed the therapeutic potential of restoring multicellular flavin homeostasis, and we established a previously unknown functional connection to a well-known pharmaceutical, namely Aspirin.

Results:

<u>Establishing a human iPSC-derived MASLD drug testing model capturing the multi-hit disease pathology</u>

We established an integrated *in vitro* and clinically stratified *in vivo* data-based and computed drug discovery pipeline to identify anti-MASLD compounds **(Fig 1A)**. First, we derived a human iHep-based MASLD model suitable for drug testing. Conventional models rely on a single injury, like FFA exposure[12,13]. We cumulatively exposed iHeps to FFAs (oleic and palmitic acid (OA/PA)), inter-organ signaling molecules (Res/Myo), and inflammatory signaling via indirect leukocyte co-culture (peripheral blood mononuclear cells (PBMCs)) **(Fig. 1B)**. To consider genomic risk modifiers, we incorporated iHeps with and without homozygous risk mutations for the *PNPLA3*-I148M variant. To contextualize all cellular injuries, we derived an overall metric based on key MASLD pathologies **(Fig. 1B & Supp. Fig. 1A)**. Each additional injury compounded effects in both cell lines, albeit certain metabolic tipping points stood out. FFAs affected steatosis but did not induce major transcriptomic shifts or other functional hallmarks in *PNPLA3*-wildtype locus-iHeps (*PNPLA3*-I148I) **(Fig. 1B & Supp. Fig. 1B)**. Yet, *PNPLA3*-I148M iHeps displayed higher susceptibility to lipotoxic stress alone and a higher steatotic propensity **(Supp. Fig. 1A & 2A&B)**. Res/Myo addition led to the first metabolic crisis, impairing mitochondrial complex gene expression, loss of mitochondrial membrane potential ($\Delta\Psi m$) and significant down-regulation of metabolic or up-regulation of cytoskeleton pathways **(Supp. Fig. 1C, Supp. Table 1)**. Notably, without additional triggers, Res and Myo individually significantly reduced $\Delta\Psi m$ in iHeps and when combined, mitochondrial spare respiratory capacity **(Supp. Fig. 1D, Fig. 1C)**. Because gut-derived inflammation and immune cell trafficking is well established in MASLD[14,15], we incorporated endotoxin-primed leukocytes. This marked another key transition, characterized by the secretion of TNFα and IL-1β, maximum release of the hepatocyte injury marker aspartate aminotransferase (AST), and maximum metabolic dysregulation, including further downregulation of metabolic pathways and upregulation of inflammatory pathways **(Fig. 1B, Supp. Fig. 1C, Supp. Table 1)**. Mass Spec analysis of the cellular secretome confirmed elevated MASLD-related markers such as DPP4[16] or heat shock proteins, indicating cellular stress **(Supp. Fig. 1E, Supp. Table 2)**. PathfindR analysis revealed donor-specific disease trigger responses: *PNPLA3*-I148M iHeps displayed perturbations in mRNA processing-related pathways, whereas *PNPLA3*-I148I iHeps displayed mitochondrial-predominant

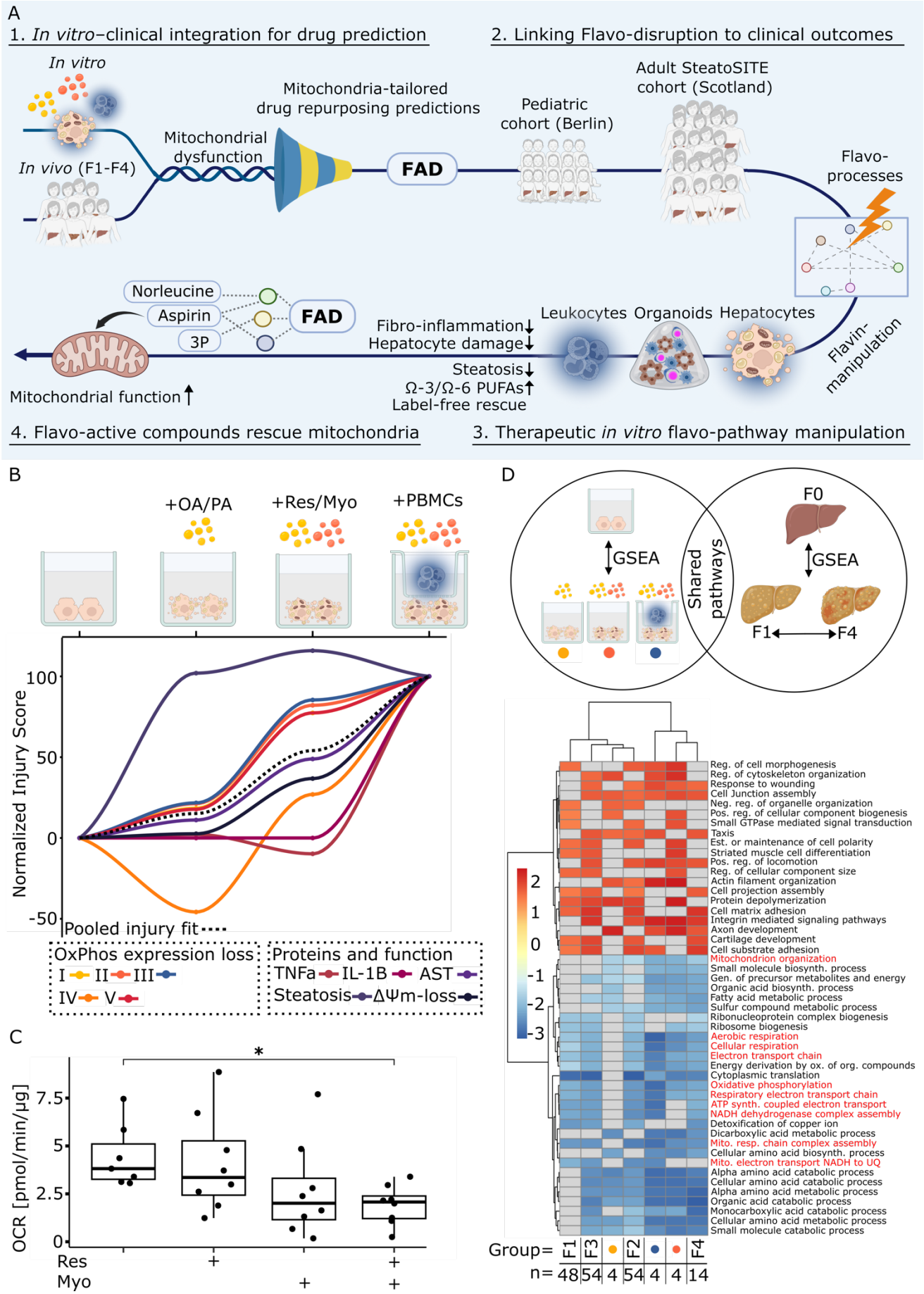
impairment **(Supp. Fig. 2C)**. Shared pathways included DNA damage and apoptosis **(Supp. Table 3)**. Importantly, benchmarking the iHep transcriptomes against clinical samples using Gene Set Enrichment Analysis revealed similar clustering of the Res/Myo and leukocyte incorporating models with the F4 disease stage **(Fig. 1D)**. Shared pathways included enhanced wound healing and diminished metabolic and mitochondrial processes. Thus, we established a multifactorial MASLD model capable of mimicking advanced pathophysiology by combining multiple factors, resulting in defined metabolic tipping points, such as mitochondrial dysfunction.

Computational discovery of FAD as a MASLD-modifying candidate via multimodal network analysis

To identify candidate compounds modifying MASLD progression, we developed a computational drug repurposing pipeline that integrates three complementary network-based approaches: (1) Protein–protein interaction (PPI)-based network proximity analysis grounded in early-stage transcriptomics, (2) fibrosis-resolved weighted gene co-expression network analysis (WGCNA) to capture disease progression signals, and (3) knowledge graph (KG) inference using PrimeKG to uncover systems-level relationships linking drugs and hepatic phenotypes **(Fig. 2A-C)**. Each method independently leverages network biology principles, but their integration, anchored in a shared biological hypothesis around mitochondrial dysfunction, adds both depth and rigor to candidate prioritization. We tailored this unified strategy to identify high-confidence therapeutic candidates for testing in our human iPSC-derived MASLD model. For both PPI and WGCNA, we used transcriptomic data from the Govaere et al. (2020)[17] cohort defining disease-relevant gene signatures and co-expression modules. For the PPI approach specifically, we refined these signatures using MitoCarta 3.0[18] focusing on mitochondrial processes. For the KG approach, we used disease phenotypes as entry points into PrimeKG, a large-scale, multimodal KG. Together, these three layers offer distinct yet convergent views on MASLD biology, enabling a systems-level prediction pipeline that links clinical data, network theory, and therapeutic inference. The following subsections describe all pipelines in detail.

PPI network-based drug proximity analysis: Differential expression analysis of liver bulk RNA-seq data identified 1,927 genes dysregulated in MASL vs. control (**Fig. 3A, Supp. Fig. 3A&B**). To refine for hepatocyte specificity, we filtered the upregulated genes using four single-cell and single-nucleus liver RNA-seq datasets from mouse and human, retaining only genes consistently elevated in hepatocytes (**Fig. 3B**). This yielded a robust hepatocyte-specific MASLD signature, with strong classification performance to effectively separate MASL from controls (**Fig. 3C, Supp. Fig. 3C**). We built a PPI network from this signature using STRING and identified topological hubs using centrality metrics. We refined them through permutation testing, yielding 21 high-confidence nodes that were both central and transcriptionally upregulated (**Fig. 3D**). These genes formed a compact subnetwork that distinguished MASL from control samples (**Fig. 3E**) and were significantly enriched for the KEGG pathway “Non-alcoholic fatty liver disease” (FDR < 0.05; **Supp. Fig. 3D**). We performed network-based drug proximity analysis to rank FDA-approved compounds based on their topological closeness to the 21-node disease module within the human interactome. This identified 34 compounds with significant proximity, listed in **Supp. Table 4**, including several that modulate mitochondrial and metabolic regulators disrupted early in MASLD (**Fig. 3F**).

WGCNA-guided mitochondrial module expansion and drug repurposing: WGCNA enables the identification of transcriptionally coordinated gene modules associated with disease phenotypes. We applied WGCNA to the Govaere et al. dataset[17] to detect modules correlated with fibrosis severity and extract key targets for therapeutic inference. WGCNA resolved 19 co-expression modules, of which six showed significant correlation with fibrosis progression **(Fig. 4A)**. These modules showed enrichment for biological processes implicated in MASLD. Notably, ME4 was enriched in extracellular matrix genes, aligning with fibrotic remodelling **(Supp. Fig. 3F)**. We prioritized ME8, enriched for mitochondrial translation and oxidative phosphorylation (OxPhos), due to its relevance to early MASLD and our MASLD drug testing model **(Fig. 4B)**.


A
1. In vitro–clinical integration for drug prediction
2. Linking Flavo-disruption to clinical outcomes
In vitro
In vivo (F1-F4)
Mitochondrial dysfunction
Mitochondria-tailored drug repurposing predictions
FAD
Pediatric cohort (Berlin)
Adult SteatoSITE cohort (Scotland)
Flavo-processes
Norleucine
Aspirin
3P
FAD
Fibro-inflammation
Hepatocyte damage
Steatosis
Ω-3/Ω-6 PUFAs
Label-free rescue
Leukocytes
Organoids
Hepatocytes
Flavin-manipulation
Mitochondrial function
4. Flavo-active compounds rescue mitochondria
3. Therapeutic in vitro flavo-pathway manipulation
B
+OA/PA
+Res/Myo
+PBMCs
Normalized Injury Score
Pooled injury fit
OxPhos expression loss
I II III
IV V
Proteins and function
TNFa IL-1B AST
Steatosis ΔΨm-loss
C
OCR [pmol/min/µg]
Res
Myo
D
GSEA
Shared pathways
F0
F1
F4
Group=
n=
Reg. of cell morphogenesis
Reg. of cytoskeleton organization
Response to wounding
Cell Junction assembly
Neg. reg. of organelle organization
Pos. reg. of cellular component biogenesis
Small GTPase mediated signal transduction
Taxis
Est. or maintenance of cell polarity
Striated muscle cell differentiation
Pos. reg. of locomotion
Reg. of cellular component size
Actin filament organization
Cell projection assembly
Protein depolymerization
Cell matrix adhesion
Integrin mediated signaling pathways
Axon development
Cartilage development
Cell substrate adhesion
Mitochondrion organization
Small molecule biosynth. process
Gen. of precursor metabolites and energy
Organic acid biosynth. process
Fatty acid metabolic process
Sulfur compound metabolic process
Ribonucleoprotein complex biogenesis
Ribosome biogenesis
Aerobic respiration
Cellular respiration
Electron transport chain
Energy derivation by ox. of org. compounds
Cytoplasmic translation
Oxidative phosphorylation
Respiratory electron transport chain
ATP synth. coupled electron transport
NADH dehydrogenase complex assembly
Detoxification of copper ion
Dicarboxylic acid metabolic process
Mito. resp. chain complex assembly
Cellular amino acid biosynth. process
Mito. electron transport NADH to UQ
Alpha amino acid catabolic process
Cellular amino acid catabolic process
Alpha amino acid metabolic process
Organic acid catabolic process
Monocarboxylic acid catabolic process
Cellular amino acid metabolic process
Small molecule catabolic process

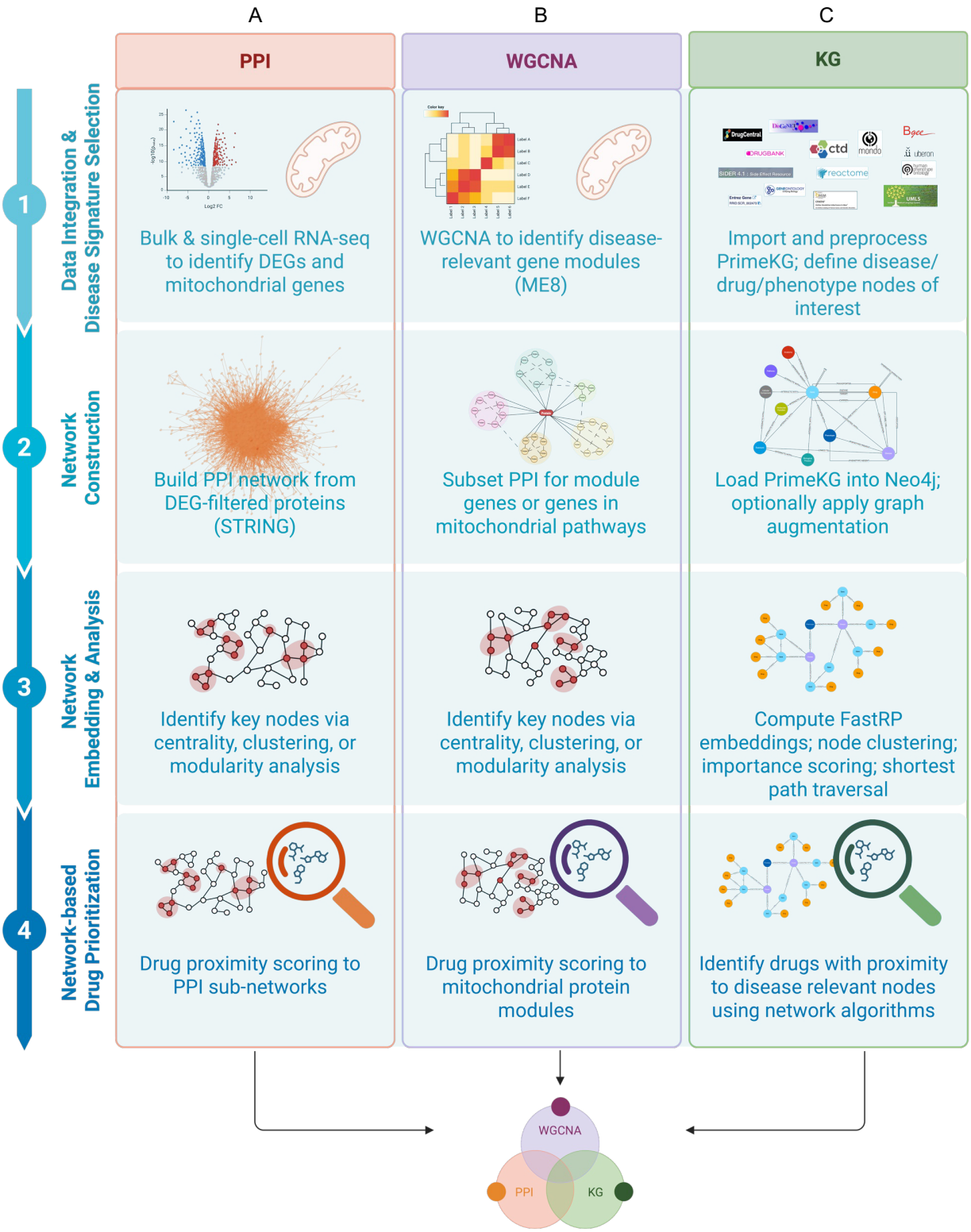
167

Fig. 1: Establishing a metabolically-associated steatotic liver disease (MASLD) drug testing model in stem cell-derived hepatocytes (iHeps) A) Schematic of the integrated drug repurposing discovery pipeline. F1-F4 = Fibrosis stage 1-4; PUFAs = polyunsaturated fatty acids **B)** Injury index in iHeps exposed to compounded metabolic injuries (Oleic and palmitic acid = OA/PA; Resistin and myostatin = Res/Myo; peripheral blood mononuclear cells = PBMCs). All parameters were standardized, ranging from Healthy (0%) to the most complex disease condition including all disease triggers (100%). Nuclear-encoded oxidative phosphorylation (OXPHOS) gene expression loss was quantified for each gene encoding individual mitochondrial complex proteins. TNFa, IL1B and AST protein data from supernatant. Steatosis and mitochondrial membrane potential (ΔΨm) of live-dye based assayed are displayed. Each parameter was assessed in at least 3 samples. **C)** Mitochondrial spare respiratory capacity of iHeps treated with Res, Myo, or both, assessed for oxygen consumption rate (OCR). Data from Seahorse XF Cell Mito Stress Test shown. Significance was assessed using multiple t-test, corrected according to the Benjamini-Hochberg method. Each data point represents an independent sample **D)** Clinical benchmarking of the *in vitro* MASLD models to clinical disease stages from Goavaere et al (n=206) stratified by fibrosis using Gene Set Enrichment Analysis (GSEA)-based clustering. GSEA was performed for all *in vitro* disease conditions compared to the untreated control. For the clinical dataset, all stages were compared to stage F0. Mitochondrial processes are marked in red.

To capture indirect regulators within the interactome, we expanded ME8 by identifying the shortest paths between ME8 genes using Dijkstra's algorithm. The resulting network comprised 7,545 proteins, with 70 key nodes identified through network centrality analysis **(Fig. 4C-D)**. These key nodes remained enriched for mitochondrial pathways **(Fig. 4E)**, confirming the importance of the pathways after expansion. We applied a network-based drug proximity model[19] to these 70 nodes, identifying 132 FDA-approved compounds with significant proximity ($Z < -1.96$) (**Supp. Table 5)**, including a set of mitochondrial-related compounds prioritized for validation.

KG-based inference using PrimeKG: PrimeKG is a multimodal KG integrating curated biological data from 20 sources. We imported the graph into Neo4j for query-based inference, visualization, and the application of graph algorithms. The graph schema and initial setup are shown in **Fig. 5A**. To identify repurposing candidates, we used two complementary queries. First, we calculated shortest paths from “non-alcoholic steatohepatitis” and “hepatic steatosis” nodes to drug nodes **(Fig. 5B–C)**, returning two unranked drug lists (500 drugs each) with 97 compounds at their intersection. These included redox cofactors, fibrates, and dual PPAR agonists. Second, we created a metabolic dysfunction-associated steatohepatitis (MASH)-specific subgraph containing all nodes and relationships connected to “non-alcoholic steatohepatitis” (**Supp. Fig. 3G, Fig. 5D**). Graph centrality algorithms applied to this 7,847-node network revealed additional candidates such as DPP-4 inhibitors. Combining the three query outputs yielded 1,153 unique drugs, with 48 shared across multiple lists **(Supp. Table 6)**. Several compounds, e.g. aspirin, have prior links to MASH, supporting the graph’s biological relevance. We excluded Troglitazone from the list due to known hepatotoxicity.

Despite differences in data sources and analytical layers, all strategies independently converged on a small subset of three compounds (**Fig. 2, Supp. Table 7**): FAD, pyridoxal phosphate (PLP), and NADH (**Supp. Fig. 3H**). These redox-cofactors emerged from entirely distinct methodologies, yet their convergence underscores a shared mechanistic axis involving mitochondrial and metabolic regulation in MASLD. From these compounds, we prioritized FAD and flavin-related processes for further validation in MASLD, due to their ill-defined role. This multi-modal strategy provides a robust framework for rational, systems-level drug repurposing in complex metabolic diseases.


A
B
C
PPI
WGCNA
KG
1
Data Integration & Disease Signature Selection
Bulk & single-cell RNA-seq to identify DEGs and mitochondrial genes
WGCNA to identify disease-relevant gene modules (ME8)
Import and preprocess PrimeKG; define disease/ drug/phenotype nodes of interest
2
Network Construction
Build PPI network from DEG-filtered proteins (STRING)
Subset PPI for module genes or genes in mitochondrial pathways
Load PrimeKG into Neo4j; optionally apply graph augmentation
3
Network Embedding & Analysis
Identify key nodes via centrality, clustering, or modularity analysis
Identify key nodes via centrality, clustering, or modularity analysis
Compute FastRP embeddings; node clustering; importance scoring; shortest path traversal
4
Network-based Drug Prioritization
Drug proximity scoring to PPI sub-networks
Drug proximity scoring to mitochondrial protein modules
Identify drugs with proximity to disease relevant nodes using network algorithms
WGCNA
PPI
KG

197
198

Fig. 2: Multimodal computational framework for drug repurposing in MASLD.
A unified pipeline integrating three network-based approaches was developed to identify candidate compounds targeting mitochondrial dysfunction in MASLD. Each method follows four key steps: data integration and signature selection, network construction, network embedding and analysis, and drug prioritization. **A) PPI-based pipeline:** Differentially expressed genes (DEGs) were derived from bulk RNA-seq (Govaere et al.) and filtered using single-cell and single-nucleus liver transcriptomes to retain hepatocyte-enriched genes. A subset overlapping with mitochondrial genes from MitoCarta 3.0 was used to construct a protein–protein interaction (PPI) network. Network hubs were identified by topological analysis, and drug proximity scores were computed to rank candidate compounds (**Supp. Table 4**). **B) WGCNA-based pipeline:** Co-expression modules associated with fibrosis stage were identified using weighted gene co-expression network analysis (WGCNA). Module ME8, enriched for mitochondrial functions, was prioritized. Network expansion identified a 'hidden layer' of intermediate nodes between module genes. Centrality analysis defined 70 key targets, and network-based proximity was used to prioritize compounds (**Supp. Table 5**). **C) Knowledge graph-based pipeline:** PrimeKG, a multimodal biomedical knowledge graph, was queried using phenotype nodes ("hepatic steatosis" and "non-alcoholic steatohepatitis"). Graph traversal, embedding-based scoring, and subgraph analysis identified drugs linked to MASLD-related nodes. Outputs from three complementary KG queries were combined (**Supp. Table 6**). **Bottom row:** A Venn diagram shows the overlap among drugs predicted by each approach. Three compounds, flavin adenine dinucleotide (FAD), pyridoxal phosphate (PLP), and nicotinamide adenine dinucleotide (NADH), were consistently identified across all pipelines (**Supp. Table 7**).

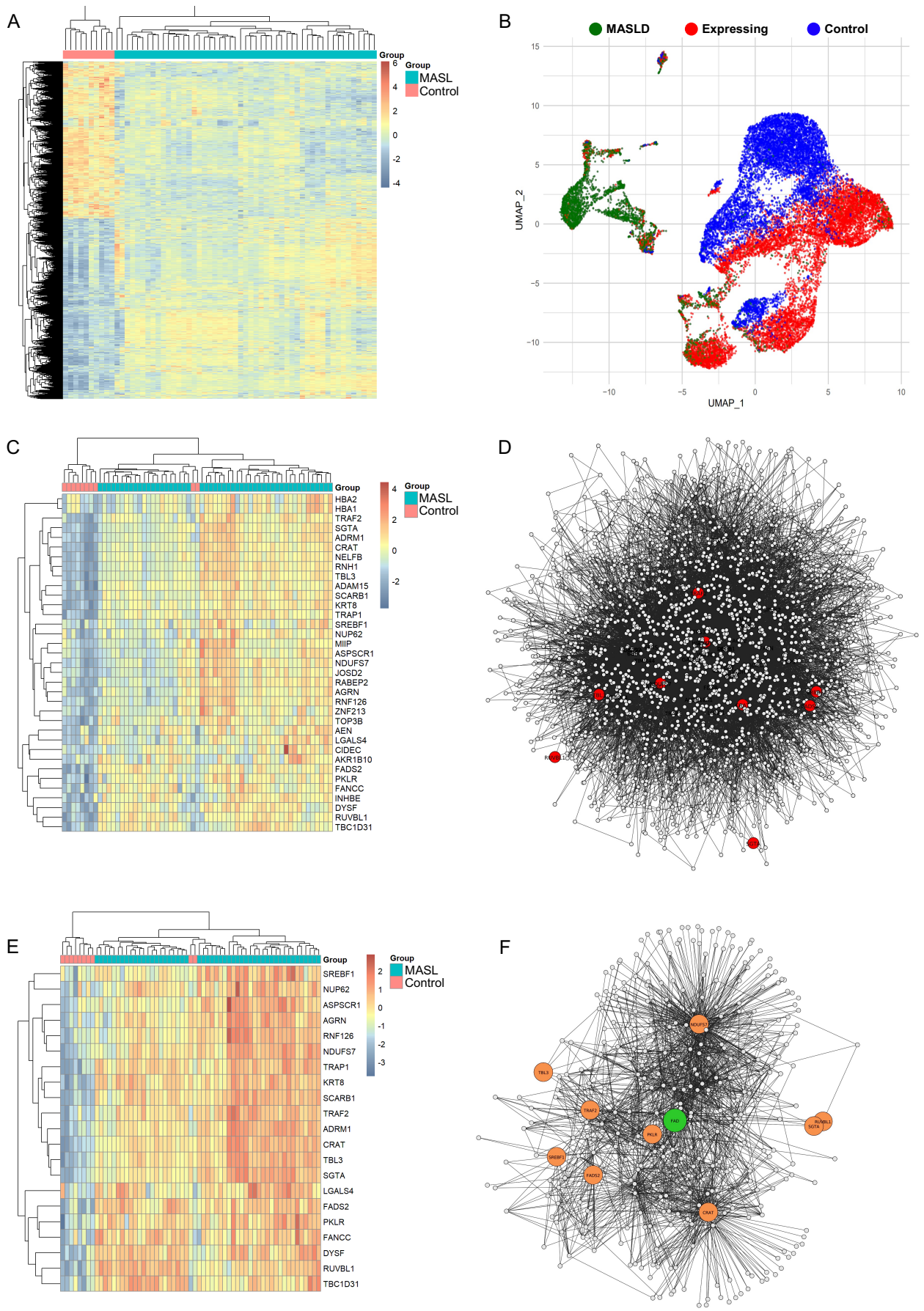

A
Group
MASL
Control
B
MASLD
Expressing
Control
UMAP_2
UMAP_1
C
Group
MASL
Control
HBA2
HBA1
TRAF2
SGTA
ADRM1
CRAT
NELFB
RNH1
TBL3
ADAM15
SCARB1
KRT8
TRAP1
SREBF1
NUP62
MIIP
ASPSCR1
NDUFS7
JOSD2
RABEP2
AGRN
RNF126
ZNF213
TOP3B
AEN
LGALS4
CIDEC
AKR1B10
FADS2
PKLR
FANCC
INHBE
DYSF
RUVBL1
TBC1D31
D
E
Group
MASL
Control
SREBF1
NUP62
ASPSCR1
AGRN
RNF126
NDUFS7
TRAP1
KRT8
SCARB1
TRAF2
ADRM1
CRAT
TBL3
SGTA
LGALS4
FADS2
PKLR
FANCC
DYSF
RUVBL1
TBC1D31
F

Fig. 3: Identification of hepatocyte-specific MASLD signatures and drug targets via integrated transcriptomic and network-based analysis. A multi-step pipeline was applied to derive a hepatocyte-specific disease module and prioritize drug candidates based on network proximity.
A) Heatmap showing the expression of 1,927 differentially expressed genes (DEGs) between MASL and control samples from human liver bulk RNA-seq data (GSE135251). Columns represent individual samples grouped by condition (MASL in red, control in turquoise), and rows represent DEGs scaled by row Z-score. Genes are hierarchically clustered. **B)** UMAP projection of a representative liver single-cell RNA-seq dataset showing major cell populations colored by group: MASLD (green), control (blue), and MASL signature-expressing cells (red overlay). MASL-specific DEGs are predominantly expressed in hepatocyte clusters. **C)** Heatmap of hepatocyte-specific upregulated genes retained after cross-validation with four single-cell and single-nucleus RNA-seq datasets (mouse and human). This refined 50-gene MASL signature displays strong separation between MASL and control samples, confirming disease relevance and cell type specificity. **D)** PPI network constructed from the refined MASL signature using STRING. Nodes represent proteins; edges indicate interactions. Red nodes highlight genes identified as network hubs by topological metrics (degree, closeness, betweenness, eigenvector centrality), filtered using permutation-based significance testing (N = 1,000). **E)** Expression heatmap of the 21 hub genes in the bulk RNA-seq dataset. Z-score-normalized rows illustrate consistent upregulation across MASL samples (red) and lower expression in controls (blue), further validating their disease association. **F)** Subnetwork showing the 21 hub genes (orange) and one of the 34 FDA-approved drugs identified via drug–target network proximity analysis. Green node indicates the example drug (FAD), and edges represent known drug–target associations within the human interactome.

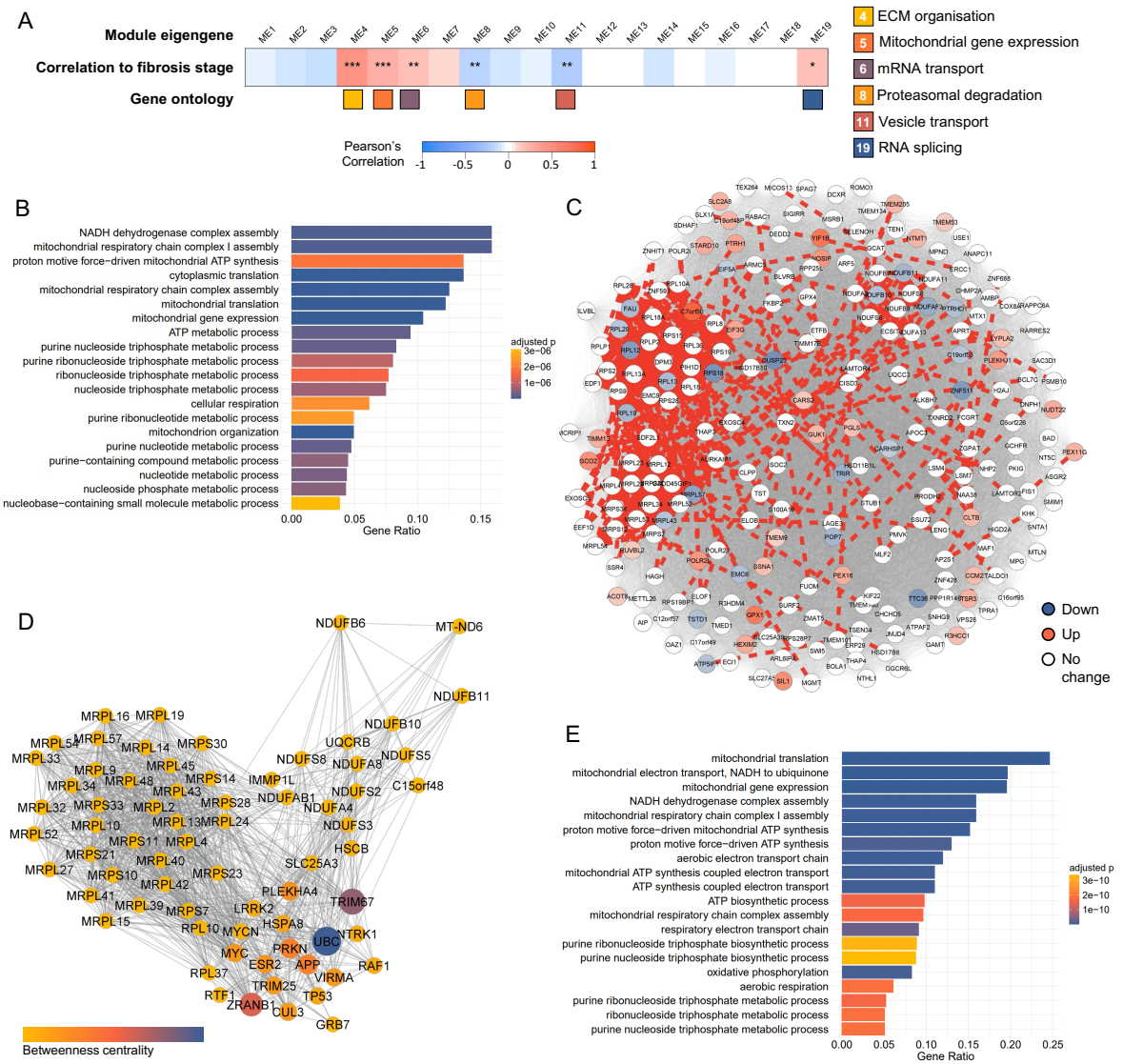


Fig. 4: Fibrosis-correlated co-expression network analysis reveals mitochondrial module ME8 as a MASLD-relevant therapeutic target.
Weighted gene co-expression network analysis (WGCNA) was applied to transcriptomic profiles from the Govaere et al. cohort to identify gene modules associated with fibrosis progression in MASLD and nominate mitochondrial drug targets. **A)** Correlation heatmap between module eigengene expression and fibrosis stage. Each row represents a gene module (ME1–ME19), with Pearson correlation coefficients shown in the heatmap. Asterisks indicate statistical significance (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). Representative enriched biological processes for significantly correlated modules are indicated with colored labels: ECM organization (blue), mitochondrial gene expression (purple), mRNA transport (orange), proteasomal degradation (green), vesicle transport (yellow), and RNA splicing (pink).
B) Top 20 enriched Gene Ontology (GO) terms for module ME8, ranked by adjusted p-value. ME8 was enriched for mitochondrial and translational processes, including NADH dehydrogenase complex assembly, oxidative phosphorylation, and ATP metabolic processes. Gene ratio represents the proportion of ME8 genes annotated to each GO term. **C)** Network visualization of ME8 showing both co-expression (grey edges) and physical protein–protein interactions (red edges). Nodes represent ME8 genes, colored by differential expression: red = upregulated, blue = downregulated, white = unchanged (relative to control). The network shows that only a subset of co-expressed genes physically interact, suggesting a layered structure of molecular coordination. **D)** PPI subnetwork constructed from ME8 hub genes after network expansion. Nodes are scaled by degree and colored by betweenness centrality (light to dark purple). This subnetwork includes direct and indirect interactors connecting ME8 genes, used for drug proximity analysis. **E)** GO term enrichment for the 70 key nodes extracted from the ME8 expanded network. Terms include mitochondrial translation, respiratory chain assembly, ATP biosynthesis, and nucleotide metabolic processes. Enrichment remains highly specific to mitochondrial function, indicating that key biological themes are preserved after network expansion.

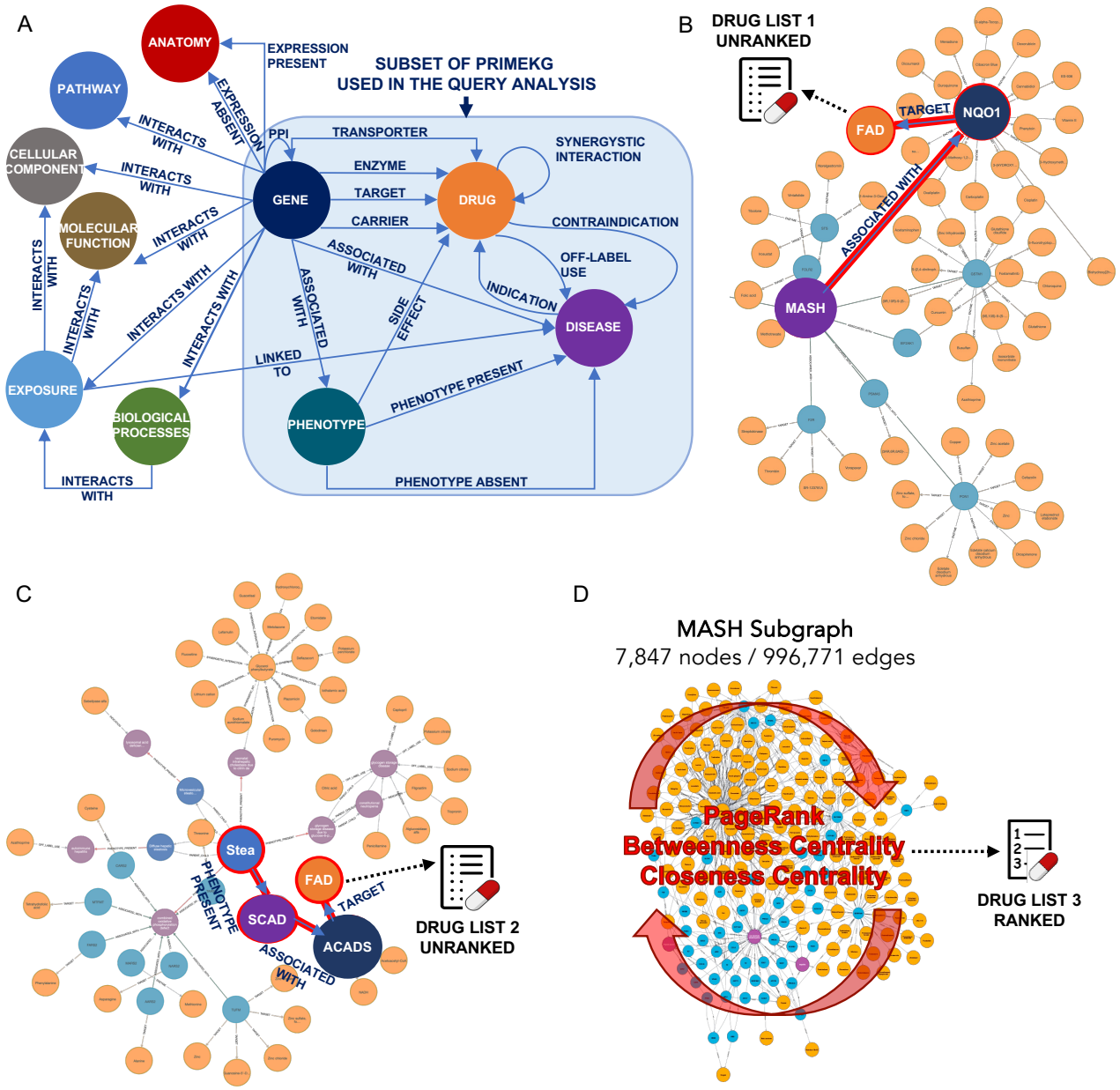
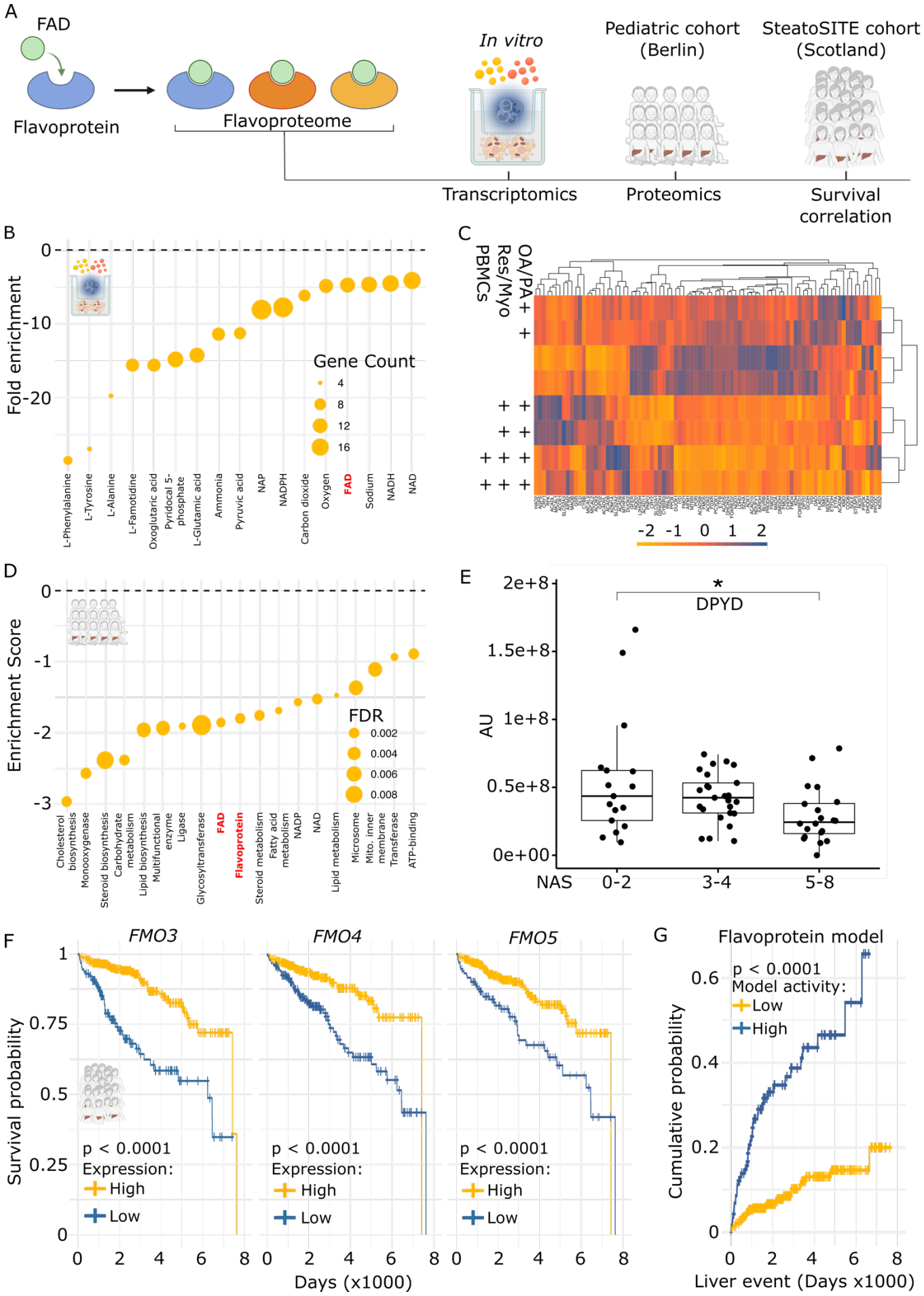


Fig. 5: Knowledge graph-based inference of MASLD drug candidates using PrimeKG. Graph-based drug repurposing was performed using PrimeKG, a multimodal knowledge graph integrating curated biomedical relationships across diseases, drugs, phenotypes, and genes. Three query strategies were used to extract compound lists relevant to hepatic steatosis and MASH. **A)** Schema of PrimeKG highlighting node types (e.g., gene, drug, disease, phenotype, pathway) and edge relationships (e.g., "associated with," "target," "interacts with," "phenotype present"). The subset of the graph used for this study is indicated in the blue box. Directed edges define relationships used to construct query paths from disease or phenotype nodes to drug targets. **B)** First traversal strategy starting from the MASH node. Gene nodes directly "associated with" MASH were retrieved, and drug nodes targeting these genes were extracted to create unranked Drug List 1. **C)** Second strategy starting from the "hepatic steatosis" phenotype node. A path was followed via "phenotype present" to reach relevant disease nodes, then further traversed through "associated with" to gene nodes, and finally to their targeting drugs. This yielded unranked Drug List 2, enriched for compounds connected through multiple phenotypic and disease layers. **D)** Third strategy involving subgraph extraction and algorithmic ranking. A MASH-specific subgraph was generated by extracting all nodes and edges connected to the "non-alcoholic steatohepatitis" node. Network centrality metrics (PageRank, betweenness centrality, and closeness centrality) were applied to score and rank drug nodes based on their topological influence within the disease module. This produced ranked Drug List 3, offering a prioritised view of candidate compounds. Stea = Hepatic steatosis; SCAD = Short-Chain Acyl-CoA Dehydrogenase; FAD = Flavin adenine dinucleotide; MASH = metabolic dysfunction-associated steatohepatitis; NQO1 = NAD(P)H:quinone oxidoreductase 1; ACADS = Acyl-CoA Dehydrogenase Short Chain

Flavoproteome perturbations in MASLD

Due to our inferred prediction of FAD´s anti-MASLD activity, we assessed flavin perturbations both *in vitro* and *in vivo* across independent populations **(Fig. 6A)**. First, metabolite enrichment analysis revealed significant enrichment *in vitro* for FAD-associated genes among those that were down-regulated **(Fig. 6B)**. To examine these changes further, we filtered the *in vitro* gene expression data for the human flavoproteome[20]. Cluster heatmap analysis displayed Res/Myo as a tipping point in flavoproteome perturbations, especially for the gene expression perturbation of *SARDH*, *PRODH*, *LDHD, DPYD*, and most of the *FMO* gene family members **(Fig. 6C, Supp. Fig 4A&B)**. *In vivo*, we validated flavoproteome perturbations using liver tissue proteomics data from our clinically stratified pediatric MASLD cohort in Berlin, Germany[21]. Using pediatric data reduces the confounding effects of cardiometabolic diseases, alcohol intake, and concomitant drug treatment present in adults. We observed significant down-regulation of flavoproteins and FAD-related processes in advanced disease stage, focusing on subcellular processes **(Fig. 6D).** All detected proteins of genes affected *in vitro* had, on average, lower abundance in advanced disease, with DPYD being significantly reduced **(Supp. Fig. 4C & Fig. 6E)**. Next, to connect flavin processes with clinical outcomes in a large adult cohort, we performed survival analysis using our MASLD SteatoSITE cohort (n = 508)[22]. As most of the *FMO* gene family was affected *in vitro*, we prioritized these genes in our analysis. While *FMO1* showed no significant association **(Supp. Fig. 4D)**, increased expression of *FMO3, FMO4*, and *FMO5* correlated significantly with a higher survival **(Fig. 6F)**. Notably, high *FMO2* expression correlated with decreased survival **(Supp. Fig. 4D)**, contradicting its recently described protective role[23]. Survival analysis of the FAD-homeostasis genes *RFT3, RFK,* and *FLAD1* reinforced flavin perturbances in MASLD pathology **(Supp. Fig. 4E)***.* While *RFK* expression correlated significantly with increased survival, elevated *RFT3* and *FLAD1* expression were associated with adverse outcomes. Thus, indicating a partial compensatory regulation of synthesizing enzymes that overcomes a co-factor shortage in advanced disease stages. To evaluate the predictive value of flavoproteome activity, we applied LASSO-penalized Cox regression, identifying a 10-gene signature that stratified patients by hepatic events (n = 446) **(Fig. 6G, Supp. Table 8).** Key contributors included the downregulation of *DPYD* and *RFK*, and the upregulation of *FMO2*, associated with an increased likelihood of adverse hepatic events. Ultimately, to enable the experimental interrogation of flavin processes in MASLD via manipulation of flavin homeostasis, we demonstrate that high-dose FAD supplementation (FAD^{H}, 50 µM) significantly raises intracellular FAD levels in our *in vitro* model by 1.3-fold **(Supp. Fig. 4F)**. Thus, we demonstrate the clinical relevance of flavin processes in MASLD pathology and establish a framework for their functional interrogation using our *in vitro* MASLD model.


A
FAD
Flavoprotein
Flavoproteome
In vitro
Pediatric cohort (Berlin)
SteatoSITE cohort (Scotland)
Transcriptomics
Proteomics
Survival correlation
B
Fold enrichment
Gene Count
L-Phenylalanine
L-Tyrosine
L-Alanine
L-Famotidine
Oxoglutaric acid
Pyridocal 5-phosphate
L-Glutamic acid
Ammonia
Pyruvic acid
NAP
NADPH
Carbon dioxide
Oxygen
FAD
Sodium
NADH
NAD
C
OA/PA
Res/Myo
PBMCs
D
Enrichment Score
FDR
Cholesterol biosynthesis
Monooxygenase
Steroid biosynthesis
Carbohydrate metabolism
Lipid biosynthesis
Multifunctional enzyme
Ligase
Glycosyltransferase
FAD
Flavoprotein
Steroid metabolism
Fatty acid metabolism
NADP
NAD
Lipid metabolism
Microsome
Mito. inner membrane
Transferase
ATP-binding
E
DPYD
AU
NAS
0-2
3-4
5-8
F
FMO3
FMO4
FMO5
Survival probability
p < 0.0001
Expression:
High
Low
Days (x1000)
G
Flavoprotein model
p < 0.0001
Model activity:
Low
High
Cumulative probability
Liver event (Days x1000)

261

Fig. 6: FAD-dependent processes are perturbed in MASLD-progression. **A)** Scheme of flavoprotein analysis *in vitro* and different *in vivo* populations **B)** Enrichment analysis for the down-regulated differentially expressed genes of the MASLD *in vitro* model incorporating OA/PA, Res/Myo and PBMCs, assessed for the associated metabolites using the human Metabolome database. **C)** Heatmap of normalized gene counts showing clustering of FAD-dependent protein gene expression across different MASLD models in iHeps **D)** Enrichment analysis of shotgun proteomics data from the Berlin pediatric MASLD patients (cohort as in Hudert et al.[21]) Enrichment analysis was done for advanced disease stage (NAS 5-8) in relation to early disease stage (NAS 1-2) **E)** Abundance of DPYD protein in the pediatric MASLD cohort described by Hudert et al.[21] Significance was assessed using multiple t-tests, corrected by the Benjamini-Hochberg method. Each dot represents an individual patient **F)** Survival analysis of FMO-gene expression and human MASLD patients from the SteatoSITE cohort[22]. Significance was assessed using a log-rank test. **G)** Predictive flavoprotein model for hepatic event probability using LASSO-penalized Cox regression. Significance was assessed using a log-rank test.

<u>Flavin modulation therapeutically intercepts metabolic crisis</u>

We assessed the therapeutic potential of flavin manipulation using iHeps and primary hepatocytes **(Fig. 7A).** Titrating FAD in iHeps revealed anti-steatotic effects at 10 µM **(Fig. 7B & Supp. Fig. 5A).** Yet, FAD^H (50 µM) restored mitochondrial function –known to be perturbed in MASH[24,25]– shown by increased coupling efficiency, decreased mitochondrial uncoupling and enhanced Complex II activity **(Fig. 7C, Supp. Fig. 5B).** We confirmed the pro-mitochondrial and anti-steatotic effects using primary human hepatocytes **(Supp. Fig. 5C-E).** FAD^H reduced hepatocyte injury, assessed by the clinical liver injury marker AST, and altered the fatty acid profile in iHeps **(Supp. Fig. 5F&G**). Specifically, FAD^H reversed injury-induced changes in saturated-to-unsaturated and omega-3-to-omega-6 fatty acid ratios, demonstrating beneficial effects on fatty acid metabolism[26] **(Fig. 7D)**. To pioneer a pipeline that can identify rescue effects without extensive cellular phenotyping, we assessed FAD^H´s effects using label-free artificial neural network-based imaging analysis of plane brightfield images **(Supp. Fig. 6).** Thereby, FAD^H mitigated injury-induced iHep morphological perturbations, indicated by tSNE clustering, with multiple mitigated features, including cytoplasm or droplet inclusion modalities **(Supp. Fig. 6C & Supp. Table 9).** This highlights a rapid and scalable approach for detecting rescue effects. To dissect therapeutic FAD-mediated effects from mitochondrial biogenesis, we tested additional overexpression of the mitochondrial biogenesis-inducer *PPARGC1A*. *PPARGC1A* is downregulated in MASH patients[27] and specifically in pro-inflammatory *in vitro* conditions in which its overexpression restores OxPhos gene expression **(Supp. Fig. 7A-C).** RNA-Seq results revealed that *PPARGC1A*-induction amplified FAD-mediated effects. However, FAD^H alone already induced significant transcriptomic changes **(Supp. Fig. 7D&E)** in iHeps, underscoring its pharmacological properties. FAD^H restored pathways suppressed in the disease model and MASH patients[28], like the complement and coagulation system, while inhibiting disease-induced pathways, including TGF-b signaling **(Fig. 7E, Supp. Fig. 1C).** Combinational *PPARGC1A* overexpression further increased metabolic pathway activity, e.g. TCA cycle **(Supp. Fig. 7F).** Functionally, combining FAD^H with *PPARGC1A* significantly reduced steatosis compared to FAD^H alone **(Supp. Fig. 7G)**, supporting synergistic effects. To validate FAD´s anti-fibrogenic and anti-TGF-modulating effects, we harnessed fibro-competent multi-cellular liver organoids from iPSCs containing hepatocytes, macrophages, stellate and endothelial cells[29] **(Supp. Fig. 8A-C)**. FAD^H reduced fibrogenic gene signature in organoids injured by the secretome of the inflammatory iHep MASLD model **(Supp. Fig. 8D)**. Next, we evaluated FAD´s immunomodulatory effects in primary immune cells exposed to the mentioned secretome **(Supp. Fig. 9)**. FAD^H intercepted the activation of innate and adaptive immune cell compartments, indicated by suppressed surface expression of MASLD-associated inflammatory activation markers[30] in monocytes and both $CD4^+$ and $CD8^+$ T-cells. Notably, this anti-inflammatory effect persisted in LPS-induced inflammation (**Supp. Fig 9D**). To expand the therapeutic scope beyond direct FAD supplementation, we predicted pharmacological compounds targeting similar proteins as FAD (considered flavo-active) of which norleucine, 3P, and aspirin fully restored mitochondrial respiration in primary hepatocytes **(Fig. 7F)**. These results support the therapeutic potential of modulating flavin pathways and point to additional clinically tractable compounds.

Thus, we demonstrate functional evidence for therapeutic manipulation of flavin-processes for alleviating hepatocyte-centric pathologies in MASLD.

302

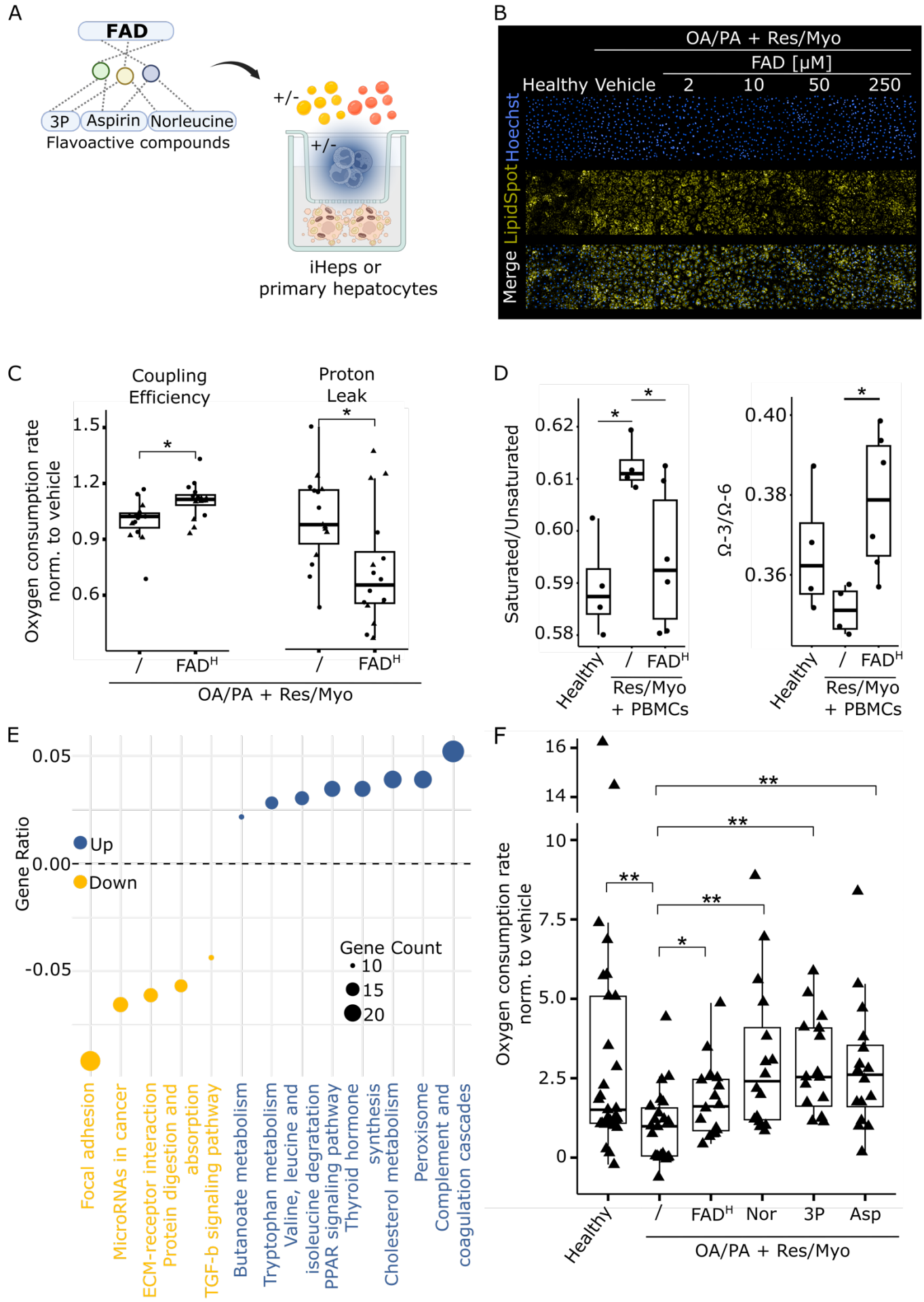

A
FAD
3P
Aspirin
Norleucine
Flavoactive compounds
+/-
+/-
iHeps or
primary hepatocytes
B
OA/PA + Res/Myo
FAD [μM]
Healthy
Vehicle
2
10
50
250
Hoechst
LipidSpot
Merge
C
Coupling Efficiency
Proton Leak
Oxygen consumption rate norm. to vehicle
1.5
1.2
0.9
0.6
*
*
/
FAD^H
/
FAD^H
OA/PA + Res/Myo
D
Saturated/Unsaturated
0.62
0.61
0.60
0.59
0.58
*
*
Ω-3/Ω-6
0.40
0.38
0.36
*
Healthy
/
FAD^H
Res/Myo
+ PBMCs
Healthy
/
FAD^H
Res/Myo
+ PBMCs
E
Gene Ratio
0.05
0.00
-0.05
Up
Down
Gene Count
10
15
20
Focal adhesion
MicroRNAs in cancer
ECM-receptor interaction
Protein digestion and absorption
TGF-b signaling pathway
Butanoate metabolism
Tryptophan metabolism
Valine, leucine and isoleucine degratation
PPAR signaling pathway
Thyroid hormone synthesis
Cholesterol metabolism
Peroxisome
Complement and coagulation cascades
F
Oxygen consumption rate norm. to vehicle
16
14
10
7.5
5.0
2.5
0
**
**
**
**
*
Healthy
/
FAD^H
Nor
3P
Asp
OA/PA + Res/Myo

Fig. 7: Manipulating FAD-homeostasis alleviates MASLD hallmarks *in vitro*. FAD^H = high-dose FAD (50 µM) **A)** Schematic of flavin-augmentation experiments in iPSC-derived (iHeps) and primary hepatocytes. **B)** Representative images of fluorescent staining of lipid droplets and nuclei in iHeps that are untreated or exposed to oleic acid and palmitic acid (OA/PA) + resistin and myostatin (Res/Myo) exposed to different FAD concentrations. **C)** Mitochondrial coupling efficiency and proton leak of iHeps (●) and primary mouse hepatocytes (▲) treated exposed to OA/PA + Res/Myo with or without FAD^H (50 µM) assessed using the Seahorse XF Cell Mito Stress Test. Significance was assessed using multiple t-tests, corrected by the Benjamini-Hochberg method. Each dot represents an independent sample **D)** Quantification of the saturated over non-saturated and omega-3 over omega-6 fatty acid ratios normalized to protein concentration in healthy iHeps or exposed to Res/Myo and peripheral blood mono-nuclear cells (PBMCs) with or without FAD^H (50µM). Significance was assessed using multiple t-tests, corrected by the Benjamini-Hochberg method. Each dot represent an independent sample. **E)** Top KEGG pathways enriched in iHeps exposed to the 3-Hit disease model treated with FAD^H (50 µM) relative to the vehicle control (both transduced with an EGFP control). **F)** Testing of mitochondrial spare respiratory capacity for additional flavo-active compounds. Mitochondrial spare respiratory capacity was assessed using the Seahorse XF Cell Mito Stress Test. Norleucine (Nor) was used in a 1 mM, Aspirin (Asp) in 5 µM and 3-Pyridin-4-yl-1H-indazol-5-ylamine dihydrochloride (3P) in 1.5 µM concentration. Significance was assessed using multiple t-tests, corrected by the Benjamini-Hochberg method. Each dot represents an independent sample. / = Vehicle

Discussion:

MASLD is the most common chronic liver disease, yet treatment options are limited[31]. Although several Phase 3 clinical trials were conducted, only Resmetirom received FDA approval. Still, Resmetirom improved fibrosis only in a small number of patients, emphasizing the need for additional treatment strategies. Herein, we integrate stem cell-derived systems with computational drug predictions, uncovering the therapeutic potential of targeting the flavoproteome in MASLD. We established a framework that couples human-relevant *in vitro* disease modeling with network-based and ML-driven hypothesis generation, based on stratified molecular patient data. Specifically, we developed a donor-adjustable drug-testing pipeline capturing hallmark pathologies of MASLD, from inter-organ metabolic triggers to a fibro-inflammatory response. We developed a computational toolbox that predicts biological pathway networks and drug candidates with high accuracy to intercept pathophysiological processes. We increased the chances of successful experimental validation by using a closed-loop iterative feedback system, reducing the need for high-throughput screening. This strategy identified FAD and other compounds active in the flavoprotein network, including Aspirin, which was recently described to improve steatosis in MASLD through an unknown mechanism[32].

Previous MASLD drug repurposing was limited in both models and computational thoroughness, utilizing microarray data from a small patient cohort and evaluating anti-steatotic drugs in HepG2 cells[33]. Contrarily, we harnessed recent advances in ML, not only in drug prioritization but also in detecting rescue effects by label-free imaging based solely on brightfield images. This approach enables straightforward rescue detection and highlights potentially affected morphological structures, as demonstrated for endothelial cells[34]. Further, harmonizing pathology- and cell-type-tailored multi-modal computational predictions, clinical validation and authentic pre-clinical modeling, advanced drug repurposing frameworks from earlier reports[33,35].

This work pioneers the integration of three complementary computational strategies: PPI-based drug proximity analysis, fibrosis-stage-resolved WGCNA, and multimodal KG inference into a unified pipeline. Translating these predictions into relevant *in vitro* models establishes a link between computational findings and experimental validation. Focusing on mitochondrial dysfunction as a central driver of MASLD enabled the biologically meaningful filtering of gene signatures and improved the specificity of compound prioritization. The multimodal KG introduced a systems-level perspective, identifying compounds associated with hepatic phenotypes through complex network relationships. Notably, all three strategies converged on FAD, highlighting the methodology's robustness and supporting a mechanistic role. Yet, limitations may persist in the KG-based component. KGs may offer biased inferences, relying on curated databases that are not routinely updated and are constrained by existing knowledge. Furthermore, integrating non-topological features, like drug chemical structure, could enhance the predictive power. Nonetheless, the KG remains a valuable hypothesis-generation tool when combined with orthogonal computational and experimental approaches, as demonstrated here.

Regarding *in vitro* modeling, we established a roadmap of compounded hepatocellular injury in iHeps. Dissecting the injury response revealed a crucial role for Res and Myo in metabolic and mitochondrial dysfunction in human inter-organ biology, supporting their debated role in humans as reported in mice and carcinoma cell lines[36–40]. Furthermore, leukocyte co-culture led to a second metabolic tipping point,

emphasizing the importance of liver-immune cell crosstalk in MASLD, similarly demonstrated in insulin resistance modeling[41]. Our modular disease model, combined with multi-cellular organoids and primary immune cells, and rigorous computational drug prioritization, enabled stage-specific interventions in complex systems, capturing inter-organ signaling and fibro-inflammatory pathology. Although our Res/Myo focus as extrahepatic signals further uncovered their role in MASLD, incorporating gut-derived signals or more defined insulin resistance signaling may further advance our model[42,43]. Furthermore, large-scale multi-donor analysis[44] can leverage the model´s full potential by enabling donor-specific drug responses for personalized therapy assessment.

By revealing the importance of flavin processes in MASLD, we demonstrate a promising axis for therapeutic intervention. Clinically, riboflavin, the precursor of FAD, serves as a treatment for genetic disorders of flavin metabolism[45] or migraine prevention[46]. Riboflavin deficiency increases MASLD activity in rodents, zebrafish, and HepG2 cells[47–49], and induces a MASLD phenotype via decreased *PPARA* activity in mice[50]. Stringently, we show increased PPAR signaling upon FAD supplementation **(Fig. 7E),** indicating a reciprocal interplay. Despite early reports of flavin reduction in cirrhotic liver disease[51], underlined by our survival analysis, this is, to our knowledge, the first study to show flavo-protective effects for human MASLD. Previous studies lacked human relevance, showing decreased hepatic lipid accumulation under riboflavin and selenium co-supplementation in rats[52], not studying fibro-inflammatory responses. For better clinical translation, we tested more bioavailable flavo-proteome-modulating drugs, like Aspirin or norleucine, an experimental non-proteinogenic amino acid commonly used for protein engineering, for which exact molecular targets need to be identified in future studies. Future translational studies will determine whether manipulating flavin pathways alone suffices for treating MASLD in a clinical setting or if additional strategies, like mitochondrial biogenesis stimulus via *PPARGC1A,* are required.

Thus, we showcase how integrating complex stem-cell-derived models with computational drug predictions enables rapid preclinical discovery and human model-centric functional validation of therapeutic candidates. Expanding this framework by incorporating patient-specific cells or combinatorial drug approaches could further advance the dissection of complex diseases and enable more precise therapeutic interventions.

Online Methods:

<u>iPSCs lines</u>

We used the two human iPSC donor lines BIHi001-B major homozygous for PNPLA3 rs738409 variant (PNPLA3- I148I) and BIHi005-A minor homozygous for the PNPLA3 rs738409 variant (PNPLA3-I148M), provided by the BIH Core Unit Pluripotent Stem Cells and Organoids. Both cell lines are registered in the Human Pluripotent Stem Cell Registry (RRID:CVCL_YN70 and RRID:CVCL_IT56). We cultured cells on vitronectin-coated (Thermo) plates in mTeSR medium (Stem Cell Tech). All experiments were performed at 37°C in a 5% O_2 and 5% CO_2 hypoxic atmosphere.

<u>Differentiation into iHeps</u>

We adapted a 28-day protocol[53], optimized by Seifert & Ludwik et al[54] and modified by us for the maturation of iHeps, as following:

Generation and thawing of ROCKi-Adapted iPSCs:

iPSCs were expanded in mTeSR medium (StemCell Technologies) containing 10 µM ROCKi (Tocris), detached with 0.5 mM EDTA in PBS for 5 min at room temperature, resuspended in Bambanker freezing medium (NIPPON), frozen at -80°C using a Mr. Frosty container, and transferred to vapor-phase liquid nitrogen (-135°C). For thawing, vials were warmed in a 37°C water bath until a small ice crystal remained, diluted dropwise with 10× volume of mTeSR + ROCKi, centrifuged (3 min, 300 g), and resuspended in fresh ROCKi-containing mTeSR.

Hepatic Differentiation Protocol:

Upon reaching ~80% confluency, iPSCs were single-cell seeded onto Geltrex-coated 6-, 12-, or 24-well plates (Day 0) and differentiated as follows:

Days 0–5: DE-induction medium (RPMI1640, 2% B27, 1% PenStrep, 100 ng/ml Activin A, 50 ng/ml Wnt3, 0.5 mM sodium butyrate), followed by 1 day without sodium butyrate.

Days 6–12: KO-medium (KO-DMEM, 20% KO Serum, 1% PenStrep, 1% Glutamax, 1% NEAA, 1% DMSO, 0.1 mM β-mercaptoethanol) for hepatic commitment.

Days 13–19: HBM medium (Lonza) supplemented with 20 ng/ml HGF, 20 ng/ml OSM, and 1% PenStrep.

Days 20–28: Maturation in HCM Bulletkit medium (Lonza, excluding EGF), with supplementation of 5 nM T3 (Tocris), 100 nM sodium selenite (Sigma), and 100 µM CDCA (Sigma)[55,56].

<u>Primary hepatocytes</u>

Primary hepatocytes were isolated from C57BL/6J mice (12–20 weeks old) using a well-established two-step perfusion technique. Initially, the liver was perfused with a calcium-free buffer to remove blood and non-parenchymal cells, followed by perfusion with collagenase (LiberaseTM, Roche, Switzerland) to dissociate the tissue enzymatically. The resulting cell suspension was purified using a Percoll gradient, yielding hepatocytes with >90% viability as previously described[57]. Purified cells were seeded onto collagen type I-coated culture plates (5 µg/cm²) at densities suitable for experimental needs. Cells were maintained in William's E medium supplemented with 2 mM L-glutamine, 5% PenStrep (penicillin–streptomycin), and hepatocyte growth supplements (CM4000; Thermo Fisher Scientific). This complete medium supported optimal growth and maintenance of hepatocytes *in vitro*.
All animal experiments complied with ethical regulations and were approved by the Committee on Research Animal Care of the Charité Universitätsmedizin Berlin, Berlin, Germany (License number: G0174/20).
The primary human hepatocytes were isolated from a 63-year-old male patient undergoing liver surgery due to colorectal liver metastasis. Cell isolation was done under informed consent and the ethics statement EA1/257/14 according to the protocol described in Horner et al[58]. PHHs were frozen in FCS with 10% DMSO and stored in vapor phase liquid nitrogen below -135°C until further use. For the experiments, we utilized a recently described protocol [59]. Briefly, cells were plated on collagen-coated plates (collagen type 1, rat tail, Gibco), with a Matrigel overlay, and maintained in HCM medium (Lonza).

Primary human peripheral blood mononuclear cells

Blood isolation was done in accordance with ethical standards of the Charité Universitätsmedizin Berlin under the ethical approval EA2_065_21. We obtained healthy human PBMCs using Ficoll-Paque (Sigma). To minimize PBMC donor variability, we derived them from a standard "master" donor who a physician clinically monitored to ensure the absence of clinical illness or medication intake. Specifically, we mixed whole blood with PBS (w/o Ca^{2+} Mg^{2+}) in a 1:1 ratio and overlaid this dilution on the Ficoll in a 3 to 2 ratio, respectively. We then centrifuged the mixture for 30 minutes at 300 g at room temperature, using the lowest levels of brake and acceleration. The leukocyte layer was transferred to a new tube and washed once with PBS. After lysing red blood cells by incubating the cells for 10 min in ACK lysis buffer, we washed twice, counted and froze cells down in a 10% DMSO solution in FCS using Mr. Frosty. The cells were stored in vapor phase liquid nitrogen below -135°C until further use.

Compounded metabolic injury model

We tested different combinations of cumulative metabolic injury hits in iHeps. Generally, we exposed iHeps for 3 days to either FFAs alone (1 Hit), combined with the inter-organ signaling molecules Res and Myo (2 Hit) or in addition with co-cultured PBMCs (3 Hits). For rescue experiments, we also included combinations as indicated.
For FFA injury, we exposed cells to oleic and palmitic acid (OA and PA) in a 5 to 2 ratio on day 28 for 3 days. Oleic and palmitic acid (both Merck) were coupled to the HCM bullet kit-derived BSA by incubation at 37°C for at least 2 h. We used 125 µM as total fatty acid concentration.
To account for inter-organ signaling in our model, we exposed iHeps to recombinant resistin and myostatin (100 ng/ml each, Peprotech) for three days.
For modeling inflammation, we included PBMCs in an indirect trans-well co-culture using Millicell trans-well inserts (Millipore) with a 0.4 µm pore size, 1 day after the addition of the other injuries (Day 29). We thawed the PBMCs 18 h before co-culture and cultured them in RPMI medium containing 10% heat-inactivated fetal calf serum (FCS) and Pen-strep. Within this 18 h pre-incubation, PBMCs incubated with 5 µg/ml lipopolysaccharide (LPS, eBioscience or Stem Cell Technologies) or without. Before adding them to the inner compartments of the trans-wells, we washed of the RPMI medium and re-suspended them in the respective HCM culture medium. PBMC viability after coculture was consistently over 90%, as assessed via Trypan Blue staining. To derive the overall injury metric, we normalized each disease marker for all conditions to untreated iHeps (0%) and iHeps exposed to 3 Hits (100%) and then applyed the geom_smooth function (method='loess', formula= y~x) in RStudio.
Due to our focus on multiple acquired metabolic injuries, compounding their mitochondrial dysfunction, we selected PNPLA3-I148I iHeps for in-depth characterization and mechanistic rescue experiments.

Live dyes

We incubated the cells for 30 min with LipidSpot 488/610 and MitoTracker™ Red CMXRos staining to assess steatosis and mitochondrial membrane potential. We used all dyes according to manufacturer´s instructions. Specifically, Mitotracker was used with a 200 nM end concentration. Eventually, we either performed live-cell imaging or fixed cells in the culture plate using 4% paraformaldehyde, followed by subsequent PBS washes before imaging.

Fluorescence imaging and image analysis

464 We imaged cells in their respective well plates using the Phenix High-Content Imaging system (Revity). We
465 captured at least 10 individual images per well, which we analyzed using the Signals Image Artist software
466 (Version 1.3.29). For measuring steatosis or mitochondrial membrane potential, we manually set upper and
467 lower thresholds for LipidSpot and Mitotracker staining in each experiment to eliminate background and
468 artefacts. For Mitotracker staining, we then measured intensity for the selected image region. To assess
469 steatosis, we measured LipidSpot area and divided it by the respective Mitotracker area or number of
470 nuclei for each image. Respective signals for steatosis or Mitotracker we then further normalized as
471 indicated.
472
473 RNA Isolation
474 We isolated RNA using a hybrid protocol of trizol RNA isolation combined with column-based purification
475 (Qiagen or Thermo). We lysed cells in 1 ml of trizol (Thermo) to which we subsequently added 0.2 ml
476 chloroform. All chloroform-trizol mixtures were then centrifuged for 10 min, using 12,000 g at 4°C. Then, we
477 transferred the aqueous phase to a new tube and added 70% ethanol in a 1:1 ratio. This mixture we loaded
478 on RNeasy Micro Kit columns (Qiagen). From this point, we performed every step according to the
479 manufacturer´s recommendations (including the on-column DNAse treatment). We assessed the
480 concentration of the isolated RNA using Nanodrop.
481
482 qPCR
483 For quantifying gene expression using qPCR, we reverse-transcribed RNA into cDNA using a cDNA
484 Synthesis kit (High-Capacity cDNA Reverse Transcription Kit (Thermo, Cat. No. 4368813), biotechrabbitTM
485 or Applied Biosystems™ High-Capacity RNA-to-cDNA™ Kit) according to manufacturer´s instructions. We
486 then performed RT-qPCR using the PowerUp SYBR Green Master Mix (Applied Biosystems, Cat. No.
487 A25776) on a QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific Inc.) according to the
488 manufacturer's instructions. Primers were synthesized by Eurofins Genomics, with sequences listed in
489 Table 1. Gene expression changes were calculated using the ddCT method.
490
491 **Table 1: Oligos used in this study**

hNDUFA8_F	ACTGGTCAACAAGTGTGCTTTGG
hNDUFA8_R	CACACTCGTCAAACTTTGCCTGC
hNDUFB1_F	TGAGTTCCCTGTTGCCCTTGGT
hNDUFB1_R	GGGACAAGAACATGAACCCAGTG
hSDHA_F	GAGATGTGGTGTCTCGGTCCAT
hSDHA_R	GCTGTCTCTGAAATGCCAGGCA
hSDHB_F	GCAGTCCATAGAAGAGCGTGAG
hSDHB_R	TGTCTCCGTTCCACCAGTAGCT
hUQCRB_F	TAAGAGGGCACTGGACCTGAAC
hUQCRB_R	CATGATTACTTCTTTGCCCATTCTT
hCox7A2_F	CGGAGAAGCAAAAACTGTTCCAG
hCox7A2_R	CCAACTGTAAGAATCATGGTGGC
hCox5B_F	GGAGATCATGCTGGCTGCAAAG
hCox5B_R	GCAGCCTACTATTCTCTTGTTGG
hATP5A1_F	GCTCCTTACTCTGGCTGTTCCA
hATP5A1_R	GCGGAGCAACAGAGACATCTGA
hATP5G3_F	GCAACAGTAGGAGTGGCTGGTT
hATP5G3_R	GCTTCAGACAAGGCAAATCCCAG
PPARGC1A_F	CCAAAGGATGCGCTCTCGTTCA
PPARGC1A_R	CGGTGTCTGTAGTGGCTTGACT
COL1A1_F	GATTCCCTGGACCTAAAGGTGC
COL1A1_R	AGCCTCTCCATCTTTGCCAGCA
COL4A1_F	TGTTGACGGCTTACCTGGAGAC
COL4A1_R	GGTAGACCAACTCCAGGCTCTC
COL4A2_F	GGATAACAGGCGTGACTGGAGT
COL4A2_R	CTTTGCCACCAGGCAGTCCAAT
TGFB1_F	TACCTGAACCCGTGTTGCTCTC
TGFB1_R	GTTGCTGAGGTATCGCCAGGAA
RPS18_F	GCAGAATCCACGCCAGTACAAG
RPS18_R	GCTTGTTGTCCAGACCATTGGC

492

493
494 Supernatant analysis
495 We collected supernatant from cells exposed to the indicated disease models with or without rescue
496 conditions after 18-24h, albeit consistent timing for ensured for individual experiments. We stored the
497 samples at -80°C until further use. The molecules TNFa, IL1B and aspartate aminotransferase (AST) were
498 measured by Labor Berlin–Charité Vivantes GmbH, Berlin, Germany, using standard clinical biochemical
499 procedures that can be detailed upon request.
500
501 Secretome analysis using Liquide chromatography – tandem mass spectrometry (LC-MS/MS)
502 For secretome analysis, we analyzed the conditioned supernatant of cells using mass spectrometry. For
503 more sensitive detection of secreted proteins, we added HCM medium without additional growth factors or
504 BSA for 24 h after exposure to the respective disease model. We centrifuged the samples and stored the
505 supernatant at -80°C until the analysis.
506 For the analysis, 18 µl of each culture supernatant were separated by one-dimensional SDS-polyacrylamide
507 gel electrophoresis according to Beier et al.[60]. Each sample lane was divided into three sub-samples (all
508 protein bands above the bovine serum albumin (BSA) band, the BSA band and all protein bands below the
509 BSA band). In-gel digestion of proteins, extraction of peptides, desalting and preparation of the samples for
510 LC-MS/MS analysis were performed as described by Beier et al [60].
511 The LC-MS/MS analysis was performed using a nanoAQUITY UPLC system (Waters corporation, Milford,
512 MA, USA) coupled to an LTQ Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific Inc.,
513 Waltham, Massachusetts; USA). The LC gradient used is described in Lerch et al[61]. A m/z of 350-1750 was
514 used for primary MS scans. Except for the exclusion time, the other MS parameters were taken from Beier
515 et al. (2021). An exclusion time of 12 s was chosen.
516 MS/MS raw files were analysed using MaxQuant (Max Planck Institute of Biochemistry, Martinsried,
517 Germany, www.maxquant.org, version 2.4.3.0 according to the parameters described previously (Beier et
518 al.[60]) using N-terminal acetylation as an additional variable modification against a database of human
519 proteins (taxonomy id 3A9606_29) and isoforms downloaded from Uniprot on 2023-06-27 and a database
520 of Bos taurus proteins downloaded from Uniprot on 2023-12-20.
521 Proteins are considered to be reliably identified if they are identified by at least two unique peptides, each
522 of them detected by at least one MS/MS scan in two different samples/replicates. In addition, their total
523 intensity must be greater than 100.000. Protein quantification was performed using MaxQuant (version
524 2.4.3.0) intensity-based absolute quantification (iBAQ) (Schwanhäusser et al.)[62]. The means of two biological
525 replicates (for iHeps without any trigger only one replicate was measured) were normalised and plotted as a
526 heatmap using the R package tidyHeatmap2 after subtracting the means of the blank samples (culture
527 medium). The data input is provided in **Supplementary Table 2.**
528 The mass spectrometry proteomic data have been deposited in the ProteomeXchange Consortium via the
529 PRIDE partner repository (Perez-Riverol et al[63]) with the dataset identifier PXD063495. Reviewers will have
530 access to the MS data at project accession: PXD063495 using the token lSRc2SuWKYkF. Alternatively,
531 reviewer can access the dataset by logging in to the PRIDE website using the following account details:
532 Username: reviewer pxd063495@ebi.ac.uk, Password: 7jwiwNNKtUYa
533
534 Bulk RNA Sequencing
535 mRNA sequencing for characterizing the MASLD *in vitro* model was performed at the BIH Genomics Core
536 facility using Illumina TruSeq stranded mRNA library preparation and for sequencing using 50 mln of
537 PE100nt reads per sample on a 25B flowcell of the Illumina NovaSeq X Plus. For exploring transcriptomic
538 changes by FAD treatment (FAD rescue) we used the services from Novogene. The sequencing of the FAD
539 rescue samples was performed using the PE150 sequencing strategy on NovaSeq Xplus sequencer 25B
540 sequencing flow cell (Illumina, San Diego) by Novogene Germany (Munich, Germany).
541
542 Features of the MASLD *in vitro* model dataset:
543 The MASLD *in vitro* model bulk rna-seq dataset consisted of 32 samples from two donor lines *PNPLA3*-
544 I148I and -I148M (16 for each donor line, four of each sample), see Table 2.
545
546 **Table 2: Sample conditions used for bulk RNA-sequencing and downstream analysis for**
547 **characterizing the *in vitro* MASLD models.**

Name	**Condition**
HCM	healthy control (untreated cells)
OA PA	cells that were solely treated with an excess of free fatty acids

OA PA ResMyo	cells that in addition to free fatty acid excess also were exposed to 2 MASLD associated signaling molecules from other tissues (adipose and muscle tissue)
OA PA ResMyo PBMCs	cells that experienced both of the above-mentioned conditions + a co-culture of stimulated primary immune cells (immune cells were not included in the RNA sequencing)

Processing and analysis of MASLD *in vitro* model RNA-seq dataset:
We started by carrying out a whole genome decoy-aware indexing and transcript quantification using Salmon v1.10.1 [64]. The Salmon pipeline estimates transcript-level abundance from RNA-seq read data. We then used DESeq2 [65], to test for differentially expressed genes between the following conditions:

I. OA PA vs HCM
II. OA PA ResMyo vs HCM
III. OA PA ResMyo PBMCs vs HCM

For derivation of the OXPHOS expression loss in the overall injury index, the counts of all genes of each oxidative phosphorylation complex were normalized to the total counts and normalized to the untreated control, to derive the expression changes. A similar strategy was used for deriving the inflammatory gene expression changes.

Signaling pathway enrichment in MASLD *in vitro* model dataset:
For each set of differentially expressed genes, we tested for enriched KEGG pathways[66].
To identify the common and distinct signalling pathways enriched in the two cell lines, we used the pathfindR pipeline[67]. pathfindR leverages interaction information from a protein-protein interaction network (PIN) to help us identify enriched biological pathways from gene expression data.
We mapped the differentially expressed genes with p values less than 0.05 onto BioGrid[68] as the PIN, then performed an active subnetwork search using a Greedy algorithm[69]. The resulting active subnetworks were then filtered based on the following scores: (a) quantile threshold greater than 0.8 and (b) number of significant genes they contain greater than 10. Using the genes in each of the active subnetworks, the significantly enriched terms (pathways/gene sets) were identified. Gene Ontology "biological processes"[70] were used as the gene sets for enrichment analysis. Enriched terms with Bonferroni adjusted p values larger than 0.05 were discarded and the lowest adjusted p value (over all active subnetworks) for each term was kept. This process of active subnetwork search and enrichment analyses was repeated for 10 iterations, performed in parallel.

Clinical benchmarking using gene set enrichment analysis:
To functionally characterize the *in vitro* models, we used gene set enrichment analysis (GSEA). GSEA is a computational method used to identify whether a set of predefined genes shows statistically significant differences in expression between two biological states (e.g., disease vs. control). Unlike traditional approaches that focus on individual genes, GSEA evaluates the expression of entire gene sets (groups of functionally related genes) to see if they are more likely to be upregulated or downregulated in a particular condition. In the context of this study GSEA helps in functional characterization by providing insights into the biological processes that are associated with the observed changes in gene expression.
Differentially expressed genes (DEGs) from each of the MASLD *in vitro* model datasets were subjected to gene set enrichment analysis (GSEA) using the "GSEA" function in R. The "msigdb" package in R was used to extract predefined gene sets (GO:BP) from the Molecular Signatures Database. Gene Ontology (GO) biological processes were enriched with a significance threshold of $p < 0.05$. This analysis was also performed for the DEGs from the Govaere et al. dataset[17], stratified by the fibrosis stages described in their study. The analyzed gene sets are summarized in Table 3.

Table 3: Datasets used for clinical benchmarking

MASLD *in vitro* model	**Govaere et al Fibrosis Stage**
OA PA vs HCM	Control
OA PA ResMyo vs HCM	F1

OA PA ResMyo PBMCs vs HCM	F2
	F3
	F4

593
594 Features of the FAD rescue dataset:
595 The FAD rescue bulk rna-seq dataset initially consisted of 12 samples (4x samples per condition) derived
596 from the *PNPLA3*-I148I line, see Table 4. Due to poor sample quality of one sample from the OA/PA +
597 Res/Myo + PBMCs + EGFP+ FAD condition which did not pass the sample QC, we had to remove it from
598 the analysis.
599
600 **Table 4: Sample conditions used for bulk RNA-sequencing and downstream analysis of the FAD**
601 **rescue.**

Name	**Condition**
OA PA ResMyo PBMCs + EGFP	iHeps exposed to the OA/PA, Res/Myo and PBMC disease model which were lentivirally transduced with a vehicle control
OA PA ResMyo PBMCs + EGFP + FAD	iHeps exposed to the OA/PA, Res/Myo and PBMC disease model which were lentivirally transduced with a vehicle control and treated with 50 µM of FAD
OA PA ResMyo PBMCs + *PPARGC1A* + FAD	iHeps exposed to the OA/PA, Res/Myo and PBMC disease model which were lentivirally transduced with *PPARGC1A*-expressing construct and treated with 50 µM of FAD

602
603 The processing and data preparation was carried out by Novogene´s bioinformatic services. The generation
604 of DEGs which were also used as an input for KEGG pathway analysis we carried out using the Novogene
605 magic web application for the following comparisons:
606 **I.** OA PA ResMyo PBMCs + EGFP vs OA PA ResMyo PBMCs + EGFP + FAD
607 **II.** OA PA ResMyo PBMCs + EGFP + FAD vs OA PA ResMyo PBMCs + *PPARGC1A* + FAD
608
609 Shiny Go analysis
610 For identifying the metabolites that are associated with the down-regulated genes in the MASLD *in vitro*
611 model we used the Shiny Go Web application[71]. We used the complete list of down-regulated DEGs from
612 the comparison of PNPLA3-I48I iHeps exposed to OA/PA+Res/Myo+PBMCs to HCM condition as an input.
613 We analyzed the enrichment for the human metabolome database using an FDR cut-off of 0.05 and
614 minimum pathway size of 10.
615
616 Drug repurposing approaches in a clinical MASLD dataset
617 Our computational drug repurposing pipeline has three approaches (Table 5): (1) Analysis of single cell and
618 single nucleus MASLD datasets to identify genes enriched in diseased hepatocytes and early stage MASLD
619 patients, (2) Weighted gene co-expression network analysis of fibrosis stages in MASH and (3) interrogation
620 of a biomedical knowledge graph to identify drug compounds associated with hepatic steatosis. We
621 implemented approaches (1 and 2) using a published clinical MASLD dataset (Govaere et al., 2020). For the
622 following sections and the results, we utilize the new nomenclature of MASLD. Thus, MASL, refers to NAFL,
623 MASLD to NAFLD and MASH to NASH in the respective data sets that use the old nomenclature as
624 identifier.
625
626 **Table 5: Computational drug repurposing pipelines**

Approach	**Sequence of analysis**
Single cell approach	1. MASLD Bulk RNA sequencing (early-stage vs control) 2. Single cell and single nucleus RNA-sequencing datasets 3. Protein-protein interaction networks 4. Drug-disease proximity analysis

WCGNA approach	1. MASLD Bulk RNA (fibrosis stages) 2. Protein-protein interaction networks 3. Drug-disease proximity analysis
Knowledge Graph approach	1. PrimeKG – Breadth-First Search (all shortest paths) 2. PrimeKG – Drug proximity search

Re-analysis of a bulk RNA-sequencing MASLD dataset:
Publicly available bulk RNA sequencing data from frozen liver biopsies were downloaded from the NCBI GEO repository (Home - GEO - NCBI (nih.gov), accession: GSE135251[17]). This dataset comprises 10 healthy controls, 51 MASL samples, and 155 MASH patients with varying stages of fibrosis.

To analyze this dataset, we utilised the Nextflow workflow system. We first used nf-core/fetchngs v 1.10.0 to fetch the metadata and raw FastQ files from the Gene Expression Omnibus. We then used nf-core/rnaseq (version 3.12.0) to analyse the RNA sequencing data and produce a gene expression matrix. Here we utilized STAR to map the raw FastQ reads to the reference genome (GRCh38.p14.) and project the alignments onto the transcriptome followed by Salmon for quantification. Differential expression analysis (DESeq2) was performed comparing MASL and healthy samples with an absolute log fold change greater than 1 and a multiple testing corrected p-value lower than 0.05 was deemed significant. Visualisations included volcano plots, MA plots, and heatmaps of the top DEGs.

Single-cell and single-nucleus RNA-seq MASLD datasets
To refine and validate the hepatocyte specificity of the bulk RNA-seq signature, four single-cell or single-nucleus RNA-seq datasets were analyzed: GSE210501, GSE166504, GSE212837, and GSE189600. These datasets include mouse models of MASLD as well as human liver samples at various stages of disease. Each dataset was processed using the Seurat package (v4.3.0) following standard quality control procedures, including filtering of cells with high mitochondrial content or low feature counts. Normalization, identification of highly variable features, PCA, UMAP dimensionality reduction, and clustering were performed. Hepatocytes were identified based on canonical markers such as human and mouse genes for Alb, Cyp2e1, and Ttr. For each dataset, the refined MASLD gene signature was tested for differential expression within hepatocyte populations comparing disease versus control conditions. AUC analyses and UMAP-based visualizations were used to assess the discriminatory power of the signature, and expression scores were computed based on average expression of the upregulated targets.

Protein-protein interaction (PPI) network construction in MASLD samples:
To identify an MASL signature, we restricted the analysis of the GSE135251 dataset to MASL patients and healthy control samples. MASH patients were not considered as we wanted to identify the early changes underpinning steatosis, prior to the onset of fibrosis and inflammation.
To prioritize functionally central regulators within the MASLD transcriptional landscape, we constructed a protein–protein interaction (PPI) network using STRING v11.0 (combined score > 400). The input consisted of upregulated hepatocyte-specific genes derived from the bulk analysis. Shortest paths were computed between all gene pairs, yielding a high-confidence undirected network. Ensembl protein identifiers were converted to HGNC gene symbols using a curated alias dictionary. Subcellular localization data from the COMPARTMENTS database were used to annotate membrane and nuclear-associated genes.

Weighted gene co-expression network analysis segmented by MASH fibrosis stage:
Gene co-expression network analysis was performed using the WGCNA R package[72,73] (version 1.73). Since the aim of this work was to construct disease specific co-expression network, the non-MASLD control samples were not included in this analysis.
First, the co-expression similarity matrix, calculated using a Pearson correlation coefficient for all possible pairs of genes, was transformed into an adjacency matrix using a soft-thresholding power β such that the resulting network structure satisfied the scale-free topology criterion. For this study, $\beta = 28$ was chosen. The adjacency matrix was then converted to a topological overlap matrix (TOM). Co-expression module identification was accomplished with the function cutreeDynamic from the dynamicTreeCut R package (version 1.63) using TOM dissimilarity (1-TOM) as the distance measure and a minimum module size cutoff of 30. Genes that were not assigned to any co-expression module were located in the 'grey' module. The resulting co-expression modules contained genes that were highly interconnected (co-expressed with each other).

To evaluate the correlation of co-expression module to fibrosis stage, the 216 MASLD samples were divided into the following subgroups using the clinical information provided by Govaere et al.[17]: MASLD (51 samples), MASH F0-F1 (34 samples), MASH F2 (53 samples), MASH F3 (54 samples) and MASH F4 (14 samples), where 'F' indicates the degree of fibrosis in each group, while MASLD patients displayed no fibrosis. We then converted the fibrosis stage metadata into a continuous numerical vector. Specifically, samples with no fibrosis were assigned a value of 0, samples with stage 1 fibrosis were assigned a value of 1, and so forth up to stage 4. The moduleEigengenes function (WGCNA R package) was used to calculate module eigengenes.
We then calculated the Pearson correlation coefficients between module eigengene expression and fibrosis stage, as well as its statistical significance, using functions from the WGCNA package. This approach allowed us to quantify the strength and direction of the association between gene module expression and fibrosis severity. Modules were considered significantly associated to fibrosis if p value < 0.05.

Pathway enrichment analysis of WGCNA modules

We performed overrepresentation analysis (ORA) to identify the main biological pathways representative of each module. First, the ensembl IDs for the genes in each co-expression module were converted to entrez IDs using the biomaRt R package (version 2.28.0). Enriched Gene Ontology terms, KEGG and Reactome pathways were then identified using the gProfiler2 R package (version 0.2.2). Enriched pathways were defined as those with FDR-adjusted p value < 0.05.

WGCNA-guided disease network construction

After having identified disease-relevant co-expression modules via WGCNA, we proceeded to expand such networks using the Dijkstra algorithm to identify all shortest paths between any two genes in a selected co-expression module. Each selected co-expression module was analyzed separately. This approach ensured that the resulting drug predictions could be directly linked to a biological process affected in MASH.
First, the module genes were mapped to the human interactome. The human interactome was assembled from the following databases: STRING, IntAct, BioGRID, HI-Union, Interactome3D, HINT, APID, BioPlex3.0, SignaLink3.0. Were necessary, the databases were filtered to: include only interactions between human proteins, include only interactions validated experimentally and remove self-loops (protein interaction with itself. For data obtained from the string database, we also removed all interactions with confidence score < 0.4. This resulted in a network of 18748 proteins and 773143 edges. All shortest paths between all pairs of module genes in the human interactome were found using the all_shortest_paths function from the NetworkX python package (version 3.0).

Topological analysis of the PPI and WCGNA disease networks

Eigenvector centrality, degree, betweenness centrality and random walk with restart (RWR) were utilized to identify key proteins in the MASL PPI network and in each WCGNA disease module. To identify significant key proteins, permutation tests were performed 1000 times for each of the network centrality algorithms. Each test generated a random network with a preserved degree distribution of the original disease network. These cumulative results were used to calculate the empirical p-value of the network algorithm. Proteins with empirical p value < 0.001 (from WGCNA module analysis) and empirical p value < 0.01 (from the single cell approach analysis) for either centrality measure were considered key nodes. The R and Python igraph packages were used for topological analysis respectively.

Drug-disease proximity analysis

FDA approved compounds and their interactions were sourced from ChEMBL[74], DrugBank[75], STITCH[76] and the Barabási group[77]. The network-based distance between drug targets and key disease proteins in the human interactome was calculated as described by Guney et al[19]. Given K, a set of disease proteins, T a set of drug targets, and d(K, T), the shortest path length between nodes k ∈ K and t ∈ T in the human interactome, we define:

$$d_c(K,T) = \frac{1}{||T||}\sum_{t\in T} min_{k\in K} d(K,T) \quad (1)$$

The closest distance measure d(K, T) (average shortest path length between any drug target and its closest disease target, eq. 1) was used here because it showed the best performance in drug-disease pair prediction in the study by Guney et al[19].
To assess the statistical significance of the distance between a key protein and a drug dc(K, T), we performed 1000 permutation tests, each time with two randomly selected gene sets. Selection of random gene sets was performed so that the degree distribution of the random sets was similar to the real sets. However, due to the scale-free nature of biological networks, there are only a few high-degree nodes in the human interactome. To avoid repetitive selection of the same, high degree nodes during random selection,

739 we used a binning approach in which nodes within a certain degree interval were grouped together such
740 that there were at least 100 nodes in each bin. When selecting the random gene sets, we sampled nodes
741 from the same bins as the real drug/disease targets. The mean and standard deviation of the reference
742 distribution was then used to calculate the Z-score of the observed distance. Drugs with Z-score <= -1.96
743 (for WGCNA-driven drug discovery) and Z-score <= -2 (for PPI-driven drug discovery) were considered
744 significantly proximal.
745
746 Network visualization and annotation
747 Final network visualisations were generated using Cytoscape (v3.9.1), iGraph and NetworkX (v2.6.3) with
748 Force Directed, Fruchterman–Reingold and circular layouts. In the WGCNA network modules **(Fig. 4C-D)**,
749 nodes were colored by differential expression and betweenness centrality respectively. Differential
750 expression was conducted using edgeR (v4.4.2) comparing all MAFLD to control cases. Betweenness
751 centrality computation was performed in Cytoscape. To illustrate the interaction landscape surrounding
752 FAD, the top convergent candidate from our pipeline, a subnetwork was constructed including all key
753 regulators and their 1-hop neighbours. FAD was manually introduced into the network and connected to its
754 known targets. Nodes were coloured by relevance (the drug in green, direct targets in orange, background
755 in grey) and annotated using custom scripts. Figures were rendered at high resolution using CairoSVG and
756 Matplotlib.
757
758 Knowledge graph analysis
759 To enable graph traversal and algorithm calculation, we imported PrimeKG into Neo4j, a graph database
760 management system that stores data as nodes, edges, and their associated attributes (Neo4j Technology
761 Inc., 2015).
762 PrimeKG - Simple Graph Queries of complete graph network
763 To identify potential drug repurposing candidates, we performed Breadth-First Search (BFS) algorithm to
764 compute the shortest paths starting from our nodes of interest. In the first approach, we initiated the query
765 from the disease node "non-alcoholic steatohepatitis" (NASH, referring to the new nomenclature MASH),
766 identified genes associated with the disease, and then queried for drugs known to interact with these
767 genes. In a second approach, we began with the phenotype "hepatic steatosis" node and expanded our
768 search to nearby disease nodes connected to hepatic steatosis, extracting genes associated with these
769 diseases and identifying drugs that interact with them. In both queries, we excluded relationships that were
770 labelled "side effect" and "contraindication".
771 PrimeKG – Drug proximity scoring to disease-relevant nodes in disease subgraph
772 Creation of the Disease Subgraph for Non-Alcoholic Steatohepatitis (NASH):
773 Starting with the NASH node, we performed a three-hop traversal to expand the subgraph, which ensured a
774 more comprehensive representation of the network by incorporating interactions from nodes directly or
775 indirectly connected to NASH. To ensure the quality of the subgraph, we filtered the relationships by
776 excluding those related to "side effects" and "contraindications", which are not directly relevant to the
777 therapeutic context of NASH. We will henceforth refer to this subgraph as "nash_subgraph".
778 Graph Algorithms for Proximity Analysis:
779 To identify possible repurposing candidate drugs for NASH, we applied three graph centrality algorithms to
780 "nash_subgraph": PageRank (using a damping factor of 0.85 to ensure convergence and a maximum of
781 1,000 iterations), Betweenness Centrality, and Closeness Centrality using the Neo4j Graph Data Science
782 (GDS) library. For each measure, we filtered for Drug nodes, ranked them in descending order of PageRank
783 score to obtain one list, and retained the top 500 drugs for further analysis.
784 To obtain a prioritized drug list for downstream integration into the WGCNA and PPI-network based
785 pipelines, we carried out a union of the drugs from the lists returned from the simple graph query and graph
786 algorithm proximity calculations.
787
788 Pediatric MASLD cohort Mass Spec analysis
789 We validated the perturbations of FAD-dependent processes in pediatric MASLD patients using the cohort
790 described in Hudert et al[21]. The normalized liver proteomic profiles of the pediatric cohort (see clinical
791 characterization and methods in Hudert et al[21]) were grouped into two categories according to their NAS
792 scores 0-2 versus 5-8 for pathway analyses. For plotting individual proteins, the following three categories
793 were established: NAS 0-2, 3-4, and 5-8.
794 Pathway analyses: The log2 fold change between the two NAS groups was calculated for each protein
795 (70% valid values in at least one group, imputation of missing values from normal distribution) (Perseus
796 v1.6.0.2), which served as input to the String database https://string-db.org/. The annotated keywords
797 (Uniprot) pathway was chosen to identify significantly regulated pathways between the two NAS groups.
798
799 Analysis of survival and liver-related events in adult MASLD cohort

800 We analyzed the SteatoSITE cohort to correlate survival and hepatic events (decompensation) with
801 flavoproteome gene expression. Kaplan-Meier Plots were generated as described in Schwabe et al and
802 tested for significance using a log-rank test. Specifically, the survival plot for individual genes uses only
803 SteatoSITE biopsy cases with successful RNAseq (n = 508). For the predicted flavo-model, we performed
804 lasso-penalised cox regression including all genes from the flavoproteome for all cases with successful
805 RNAseq where there was no decompensation coding event prior to the biopsy (n = 446). For the
806 coefficients from the resulting 10-gene signature, we derived a single metric value for each case by
807 multiplying the coefficient by the cpm for the named gene and summing them. We then derived a single
808 cutpoint for that new metric that best divided patients into high and low risk of decompensation using the
809 surv_cutpoint function in R. Based on that, we then classified them into high and low risk groups to
810 generate the Kaplan-Meier plot.
811
812 Quantification of intracellular FAD
813 We quantified intracellular FAD levels using the fluorescence-based detection of the FAD assay kit (Abcam,
814 ab204710) according to the manufacturer's instructions. Before harvesting cells for the assay, we washed
815 3x times with PBS to ensure no significant amounts of exogenous FAD remained. We performed the assay
816 read-out using the Spectramax i3X Plate reader.
817
818 Identification of compounds with similar network topology to FAD
819 Given the impact of mitochondrial dysfunction in MASLD[78], we refined MASL signatures to a subset of the
820 MASLD DEG samples associated with mitochondrial structure or function. The list of mitochondrial-
821 associated genes was sourced from the MitoCarta 3.0 database[18]. We then selected compounds with
822 shared nodes to FAD. We further expanded the drug search to include 24,000 investigational and
823 experimental compounds in phase 1-3 clinical trials **(Supp. Table 10)**. From these, we prioritized norleucine
824 (Nor), aspirin (Asp), 3-Pyridin-4-yl-1H-indazol-5-ylamine dihydrochloride (3P), A-443654 and BCTC for
825 rescue experiments.
826
827 Mechanistic rescue experiments
828 For the rescue experiments, we co-exposed the cells to the indicated concentrations of FAD (Roth), Nor
829 (Thermo), Asp (Sigma), 3P (Sigma), A-443654 (biomol), BCTC (Sigma), plus the indicated injury condition
830 using the duration and concentrations described for the iHeps. Thus, the total exposure to each compound
831 was 72 h. After the co-exposure, we harvested the respective material for the analysis as indicated. BCTC
832 and A-443654 did not show significant differences, which is why they are omitted in the plot.
833
834 Measurement of the oxygen consumption rate (OCR)
835 For measuring the OCR in iHeps, we detached them from one full 6-well plate and seeded them on one full
836 Geltrex (Thermo) coated Seahorse Plate on day 28 of differentiation. For detachment, we washed iHeps
837 with PBS and then incubated the cells for at least 10 min with 10x TrypLE (Gibco) until they rounded up. We
838 added 5-fold volume of PBS with 20% KO-Serum, centrifuged the cells for 3 min at 200 g, and
839 resuspended them in HCM complete media containing all supplements as described. We first added 50 µl
840 of the cells for attachment, followed by 50 µl of more medium 5 h later. For testing Res and Myo, we
841 exposed cells to each factor individually or both combined for 3 days, compared to the vehicle only. For
842 testing the rescue effect by FAD or other compounds, we co-exposed iHeps and primary mouse
843 hepatocytes to the compound plus OA/PA and Res/Myo for three days. We measured OCR using the
844 Seahorse Extracellular Flux (XF)-96 Bioanalyzer from Agilent Technologies, as described previously[79].
845 Briefly, the XF basal medium from Agilent Technologies was supplemented with 10 mM glucose, 1 mM
846 pyruvate, and 1 mM glutamine. The electron transport chain complexes inhibitors (oligomycin (2 µM), FCCP
847 (1 µM), rotenone (1 µM), and antimycin (1 µM)) were titrated and delivered through injection at designated
848 ports A, B, and C of the Bioanalyzer. The resulting data were normalized to the protein content in each well
849 using the Pierce™ BCA assay (Thermo Fisher Scientific).
850
851 Measurement of electron transport chain complex activities via respirometry-based assay
852 The electron transport chain complex I, II, and IV activities were measured in frozen hepatocytes using
853 Seahorse XF-96 Extracellular Flux Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) as described[80].
854 Primary mouse hepatocytes were treated with the respective treatments for 72 hours. After treatment, the
855 medium was removed, cells were washed with PBS three times. Then, the cells were harvested and frozen
856 at -70°C. On the day of assay, cells were homogenized in MAS buffer (70 mM sucrose, 220 mM mannitol, 5
857 mM KH2PO4, 5 mM MgCl2, 1 mM EGTA, 2 mM HEPES, pH 7.4) using 10 strokes in a glass Dounce
858 homogenizer. The homogenate was centrifuged at 200 g for 10 minutes at 4°C, and the supernatant was
859 collected. Protein concentration was measured using the Pierce™ BCA assay (Thermo Fisher Scientific).

860 For the Seahorse XF-96 assay, 15 µg of homogenate protein was loaded into each well in 20 µl of MAS
861 buffer. The plate was centrifuged at 2,000 g for 5 minutes at 4°C, after which 130 µl of MAS buffer
862 containing cytochrome c (10 µg/ml) was added to each well. Substrates were delivered through port A
863 (NADH, 1 mM; or succinate + rotenone, 5 mM + 2 µM). Inhibitors rotenone and antimycin A (2 µM and 4 µM,
864 respectively) were delivered through port B, TMPD + ascorbic acid (0.5 mM + 1 mM) through port C, and
865 azide (50 mM) through port D. The data was normalized to the proteins content. OCR signal after azide
866 injection was subtracted from NADH (complex I activity), succinate + rotenone (complex II activity), and
867 TMPD+ascorbic acid (complex IV activity).
868
869 Lipidomics
870 We harvested iHeps that we left untreated or exposed to Res/Myo and PBMCs with or without 50 µM of
871 FAD. We omitted exogenous FFAs to focus on cell-intrinsic changes. After exposure, we harvested the cells
872 by washing them twice with PBS and subsequently incubating them in 10X TryPLE. After the cells
873 dissociated from the dish, we inactivated TryPLE using PBS with 20% KO-Serum and centrifuged them for
874 3 min using 200 g. The Pellets were then stored in -80°C for 24 h until the analysis.
875 For the analysis cells were hydrolyzed with 100 µL 10 mol/L NaOH within 120 min at 80 °C. The samples
876 were neutralized with 100 µL acetic acid and filled up to 5 mL with methanol containing internal standards
877 (Cayman Chemical, Ann Arbor MI).
878 HPLC measurement was performed using an Agilent 1290 HPLC system with binary pump, autosampler
879 and column thermostat equipped with a Phenomenex Kinetex-C18 column 2.6 µm, 2.1 x 150 mm column
880 (Phenomenex, Aschaffenburg, DE) using a solvent system of acetic acid (0.05%) and acetonitrile. All
881 solvents and buffers in LC-MS-grade were purchased from VWR Germany
882 The solvent gradient started at 70 % acetonitrile and was increased to 98 % within 10 min and hold until 14
883 min with a flow rate of 0.4 mL/min and 5 µL injection volume.
884 The HPLC was coupled with an Agilent 6470 triplequad mass spectrometer with electrospray ionisation
885 source operated in negative selected ion mode.
886
887 Label-free brightfield imaging analysis
888 For the cell morphology analysis by label-free brightfield imaging using the VAIDR technology platform (TRI
889 Thinking Research Instruments GmbH, Hamburg, Germany), we took a minimum of 5 images per condition
890 as input.
891 Image segmentation:
892 The VAIDR image analysis software was used to develop a segmentation method to detect the relevant
893 sub-cellular structures of the hepatocytes on the label-free images. To this end, [six] image segments [of
894 512x512 pixels] were manually annotated using the software's label editor. For the annotation, the following
895 classes were used: Nucleus, membrane, cytoplasm, lipid, and mitochondria. The software then
896 automatically trained a segmentation algorithm based on the annotated images. This algorithm was applied
897 to all [96 full-size] images in the dataset, yielding one segmentation mask per image, which assigns each
898 pixel to one of the classes mentioned above.
899 Feature extraction:
900 Based on these masks and the original images, 185 numerical features were calculated as follows: For each
901 contiguous region, called "object", belonging to a particular class, the area, roundness, as well as the
902 image intensity and texture intensity, were calculated. [For example, groups of pixels of the "nucleus" class
903 represent the detected nuclei. For each such group, the area, roundness, intensity and texturation yield a
904 rich phenotypic profile of the respective nucleus.] The object-level results were aggregated per well by
905 calculating the mean and standard deviation of each parameter and each object class over all objects in a
906 well. The object count for each class was used as an additional feature.
907 Feature evaluation:
908 The resulting well-level phenotypic features were evaluated in two ways: (i) visually, by calculating the t-SNE
909 (1) plot. (ii) For each feature, a T-test was performed to assess whether untreated diseased cells could be
910 distinguished from FAD-treated diseased cells based on that feature alone. The T-test was perfomed using
911 the function ttest_ind from the Python package scipy.stats.
912
913 Lentivirus production and transduction experiments
914 We generated lentiviral particles by co-transfecting HEK293T cells with the plasmids VSV-G, pCMV8.9
915 (both provided by the Viral Core Facility of the Charité Universitätsmedizin Berlin) and either EGFP (cloned
916 into the pLX304 backbone) or *PPARGC1A* (#Addgene #143709). Two days after transfection, supernatant
917 was harvested, filtered and concentrated 20-fold using the Takara Lenti-X-Concentrator. Virus was
918 aliquoted and frozen in -80 until further use.
919 For transduction, iHeps were exposed to a low-level viral titer in their native medium for at least 4 hours,
920 with the addition of 5 µg/ml polybrene (approximately 0.1 functional MOI) 2 days before exposure to the

disease model. After transduction, the cells were treated with fresh medium and used for further experiments as indicated.

FAD rescue in fetal liver-like organoids (FLOs)

For generating FLOs, we used the protocol from Quach et al[29] using the *PNPLA3*-I148M line. At day 18, we added supernatants from healthy iHeps or supernatant from iHeps exposed to OA/PA, Res/Myo and PBMCs in a to 2 / 5 ratio to the FLO medium, to which either a vehicle control or 50 µM FAD was freshly added before the culture. Medium was changed every other day. After 120 h, cells were harvested for analysis.

FAD rescue in primary immune cells

We freshly isolated healthy PBMCs as described above, used for disease modeling, and cultured 50000 cells in 100 µl of RPMI with 10% heat-inactivated FCS and Pen-Strep in a 96-well plate. To these, we either added 50 µl of supernatant from healthy iHeps or from iHeps exposed to OA/PA, Res/Myo and PBMCs with or without freshly added FAD (50 µM). We included 5 µg/ml of LPS (Stem Cell Technologies) as a positive control, with or without 50 µM FAD. After 18 h, we harvested the cells for full-spectrum flow cytometry, stained for the Panel in Table 6. Briefly, we transferred cells into FACS tubes and washed off the medium using FACS buffer (PBS with 2 mM EDTA). After centrifugation, we discarded the supernatant, and the viability dye was added to the residual volume. After 10 minutes, we added the antibody mix that we prepared in blocking buffer. After 20 min, we washed the samples using FACS buffer. Subsequently, we fixed the cells in 2% formalin for 10 min, which we washed off using FACS buffer. Then, samples were acquired on a Cytek® Aurora Cytometer with 3 lasers (Violet, Blue, Red) and 38 detectors with the Spectroflo® v3.0.3 software. Daily QC was performed according to the manufacturer's instructions. After unmixing, cells were analyzed with the gating indicated.

Table 6: Flow cytometry-antibodies and dyes used in this study

Channel	Marker	Target	Clone	Manufacturer	Ordering No
BV421	CD80	Myeloid activation	2D10	BioLegend	305222
PacB	CD66b	Neutrophils	G10F5	BioLegend	305112
BV480	CD4	CD4TC	RPA-T4	BD	746541
V500	CD16	Neutrophils/monocytes	3G8	BD	561394
SV538	CD3	TC	UCHT1	BioLegend	300484
BV570	HLA-DR	MHC-II	L243	BioLegend	307638
BV711	CD40L	TC activation	24-31	BioLegend	310838
BV750	CD56	NK cells	5.1H11	BioLegend	362556
BB515	CD8	CD8TC	RPA-T8	BD	564526
Alexa532	CD45	Pan-immune	HI30	eBio	58-0459-42
PE-Cy5	CD19	B cells	HIB19	BioLegend	302210
PerCP-Cy5.5	CD14	Monocytes	M5E2	BioLegend	301824
PE-Fire810	CD86	Myeloid activation	IT2.2	BioLegend	305455
APC	CD40	Myeloid activation	5C3	BioLegend	334310
Al647	PD1	TC activation	NAT105	BioLegend	367420
Al700	CD11c	APC/DC	Bu15	BioLegend	337220
Zombie NIR	Viability			BioLegend	423106
APC Fire 810	CD25	TC activation	M-A251	BioLegend	356150

Multiplexed immunofluorescence (mIF) of organoids

mIF was performed on 5-µm thick formalin-fixed paraffin-embedded organoid sections as previously described[81]. Briefly, antigen retrieval was performed in heated citrate buffer (pH=6.0). Antibodies were eluted after each cycle of staining using a 2-mercaptoethanol/SDS (2-ME/SDS) buffer. Primary and secondary antibodies used in this study are listed in Table 7. Images were acquired on a Zeiss AxioObserver 7 and processed in FIJI, as previously reported[82].

Table 7: Antibodies used for mIF

Marker	Source	Identifier
Rabbit anti-CD31	abcam	Cat#ab28364
Mouse anti-HepPar1	Dako	Cat#OCHIE5
Goat anti-IBA-1	abcam	Cat#ab5076
Goat anti-PDGFRA	R&D Systems	Cat#AF-307-SP
Donkey anti-Goat IgG Alexa Fluor 488	Sigma-Aldrich	Cat#SAB4600032
Donkey anti-Mouse IgG Alexa Fluor 647	Invitrogen	A32787
Goat anti-Rabbit IgG Alexa Fluor 750	Cell Signaling	Cat#4418S

Statistical testing

We performed data visualization and statistical significance assessment using RStudio or GraphPad Prism (Version 10). We used the statistical tests as indicated, two-sided if not stated otherwise.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The main data supporting the results in this study are available within the Article and its Supplementary Information. Raw data of the *in vitro* transcriptomics will be provided to the reviewers if requested and uploaded to a public repository upon publication. The SteatoSITE transcriptomic data are deposited in the European Nucleotide Archive (https://www.ebi.ac.uk/ena; study accession number: PRJEB58625). Gene expression data are also freely available for user-friendly interactive browsing online at https://shiny.igc.ed.ac.uk/SteatoSITE_gene_explorer/. Full data access can be requested at https://regeneration-repair.ed.ac.uk/research/research-resources/steatosite. Mass Spec raw data can be found as indicated. Proteomic data from the pediatric cohort was utilized from Hudert et al[21].

References

1. Loewa, A., Feng, J. J. & Hedtrich, S. Human disease models in drug development. *Nat. Rev. Bioeng.* **1**, 545–559 (2023).
2. Cerneckis, J., Cai, H. & Shi, Y. Induced pluripotent stem cells (iPSCs): molecular mechanisms of induction and applications. *Signal Transduct. Target. Ther.* **9**, 112 (2024).
3. Younossi, Z. M. *et al.* The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review. *Hepatology* **77**, 1335–1347 (2023).
4. Le, P. *et al.* Estimated Burden of Metabolic Dysfunction–Associated Steatotic Liver Disease in US Adults, 2020 to 2050. *JAMA Netw. Open* **8**, e2454707 (2025).
5. Koyama, Y. & Brenner, D. A. Liver inflammation and fibrosis. *J. Clin. Invest.* **127**, 55–64 (2017).
6. Buzzetti, E., Pinzani, M. & Tsochatzis, E. A. The multiple-hit pathogenesis of non-alcoholic fatty liver disease (NAFLD). *Metabolism* **65**, 1038–1048 (2016).

1027 7. Hendriks, D. *et al.* Engineered human hepatocyte organoids enable CRISPR-based target discovery and
1028 drug screening for steatosis. *Nat. Biotechnol.* **41**, 1567–1581 (2023).
1029 8. Ajmera, V. *et al.* Novel plasma biomarkers associated with liver disease severity in adults with
1030 nonalcoholic fatty liver disease. *Hepatology* **65**, 65–77 (2017).
1031 9. García, P. S., Cabbabe, A., Kambadur, R., Nicholas, G. & Csete, M. Elevated Myostatin Levels in Patients
1032 with Liver Disease: A Potential Contributor to Skeletal Muscle Wasting. *Anesth. Analg.* **111**, 707–709
1033 (2010).
1034 10. Delogu, W. *et al.* Myostatin regulates the fibrogenic phenotype of hepatic stellate cells via c-jun N-
1035 terminal kinase activation. *Dig. Liver Dis.* **51**, 1400–1408 (2019).
1036 11. Wilkes, J. J., Lloyd, D. J. & Gekakis, N. Loss-of-Function Mutation in Myostatin Reduces Tumor Necrosis
1037 Factor α Production and Protects Liver Against Obesity-Induced Insulin Resistance. *Diabetes* **58**, 1133–
1038 1143 (2009).
1039 12. Jeon, J., Lee, S. H., Kim, D. & Sung, J. H. In vitro hepatic steatosis model based on gut–liver-on-a-chip.
1040 *Biotechnol. Prog.* **37**, e3121 (2021).
1041 13. Ricchi, M. *et al.* Differential effect of oleic and palmitic acid on lipid accumulation and apoptosis in
1042 cultured hepatocytes. *J. Gastroenterol. Hepatol.* **24**, 830–840 (2009).
1043 14. Carpino, G. *et al.* Increased Liver Localization of Lipopolysaccharides in Human and Experimental
1044 NAFLD. *Hepatology* **72**, 470–485 (2020).
1045 15. An, L. *et al.* The Role of Gut-Derived Lipopolysaccharides and the Intestinal Barrier in Fatty Liver
1046 Diseases. *J. Gastrointest. Surg.* **26**, 671–683 (2022).
1047 16. Baumeier, C. *et al.* Elevated hepatic DPP4 activity promotes insulin resistance and non-alcoholic fatty
1048 liver disease. *Mol. Metab.* **6**, 1254–1263 (2017).
1049 17. Govaere, O. *et al.* Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals
1050 gene signatures for steatohepatitis and fibrosis. *Sci. Transl. Med.* **12**, eaba4448 (2020).
1051 18. Rath, S. *et al.* MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization
1052 and pathway annotations. *Nucleic Acids Res.* **49**, D1541–D1547 (2021).

1053 19. Guney, E., Menche, J., Vidal, M. & Barábasi, A.-L. Network-based in silico drug efficacy screening. *Nat.*
1054 *Commun.* **7**, 10331 (2016).
1055 20. Lienhart, W.-D., Gudipati, V. & Macheroux, P. The human flavoproteome. *Arch. Biochem. Biophys.* **535**,
1056 150–162 (2013).
1057 21. Hudert, C. A. *et al.* Genetic determinants of steatosis and fibrosis progression in paediatric non-
1058 alcoholic fatty liver disease. *Liver Int.* **39**, 540–556 (2019).
1059 22. Kendall, T. J. *et al.* An integrated gene-to-outcome multimodal database for metabolic dysfunction-
1060 associated steatotic liver disease. *Nat. Med.* **29**, 2939–2953 (2023).
1061 23. Ke, C. *et al.* FMO2 ameliorates nonalcoholic fatty liver disease by suppressing ER-to-Golgi transport of
1062 SREBP1. *Hepatology* **81**, 181–197 (2025).
1063 24. Koliaki, C. *et al.* Adaptation of Hepatic Mitochondrial Function in Humans with Non-Alcoholic Fatty Liver
1064 Is Lost in Steatohepatitis. *Cell Metab.* **21**, 739–746 (2015).
1065 25. Pérez-Carreras, M. *et al.* Defective Hepatic Mitochondrial Respiratory Chain in Patients With
1066 Nonalcoholic Steatohepatitis. *Hepatology* **38**, 999–1007 (2003).
1067 26. Simopoulos, A. P. The importance of the ratio of omega-6/omega-3 essential fatty acids. *Biomed.*
1068 *Pharmacother.* **56**, 365–379 (2002).
1069 27. Arconzo, M. *et al.* Hepatic-specific Pgc-1α ablation drives fibrosis in a MASH model. *Liver Int.* **44**, 2738–
1070 2752 (2024).
1071 28. Subudhi, S. *et al.* Distinct Hepatic Gene-Expression Patterns of NAFLD in Patients With Obesity.
1072 *Hepatol. Commun.* **6**, 77–89 (2022).
1073 29. Rezvani, M. *et al.* Fetal Liver-like Organoids Recapitulate Blood-Liver Niche Development and
1074 Multipotent Hematopoiesis from Human Pluripotent Stem Cells. Preprint at
1075 https://doi.org/10.1101/2024.10.11.617794 (2024).
1076 30. Zhang, C. & Yang, M. Targeting T Cell Subtypes for NAFLD and NAFLD-Related HCC Treatment: An
1077 Opinion. *Front. Med.* **8**, 789859 (2021).
1078 31. Tacke, F. *et al.* EASL–EASD–EASO Clinical Practice Guidelines on the management of metabolic
1079 dysfunction-associated steatotic liver disease (MASLD). *J. Hepatol.* **81**, 492–542 (2024).

1080 32. Simon, T. G. *et al.* Aspirin for Metabolic Dysfunction–Associated Steatotic Liver Disease Without
1081 Cirrhosis: A Randomized Clinical Trial. *JAMA* **331**, 920 (2024).
1082 33. Zareifi, D. S. *et al.* A network-based computational and experimental framework for repurposing
1083 compounds toward the treatment of non-alcoholic fatty liver disease. *iScience* **25**, 103890 (2022).
1084 34. Kusumoto, D. *et al.* Anti-senescent drug screening by deep learning-based morphology senescence
1085 scoring. *Nat. Commun.* **12**, 257 (2021).
1086 35. Seagle, H. M. *et al.* Genomics-Informed Drug Repurposing Strategy Identifies Novel Therapeutic Targets
1087 for Metabolic Dysfunction-Associated Steatotic Liver Disease. Preprint at
1088 https://doi.org/10.1101/2025.02.18.25321035 (2025).
1089 36. Wen, F. *et al.* Acute Elevated Resistin Exacerbates Mitochondrial Damage and Aggravates Liver
1090 Steatosis Through AMPK/PGC-1α Signaling Pathway in Male NAFLD Mice. *Horm. Metab. Res.* **53**, 132–
1091 144 (2021).
1092 37. Yang, H.-M. *et al.* Resistin impairs mitochondrial homeostasis via cyclase-associated protein 1-mediated
1093 fission, leading to obesity-induced metabolic diseases. *Metabolism* **138**, 155343 (2023).
1094 38. Zhou, L. *et al.* Resistin reduces mitochondria and induces hepatic steatosis in mice by the protein kinase
1095 C/protein kinase G/p65/PPAR gamma coactivator 1 alpha pathway. *Hepatology* **57**, 1384–1393 (2013).
1096 39. Liu, X., Pan, J. P., Bauman, W. A. & Cardozo, C. P. AdipoRon prevents myostatin-induced upregulation of
1097 fatty acid synthesis and downregulation of insulin activity in a mouse hepatocyte line. *Physiol. Rep.* **7**,
1098 (2019).
1099 40. Acquarone, E., Monacelli, F., Borghi, R., Nencioni, A. & Odetti, P. Resistin: A reappraisal. *Mech. Ageing*
1100 *Dev.* **178**, 46–63 (2019).
1101 41. Groeger, M. *et al.* Modeling and therapeutic targeting of inflammation-induced hepatic insulin
1102 resistance using human iPSC-derived hepatocytes and macrophages. *Nat. Commun.* **14**, 3902 (2023).
1103 42. Yang, J. *et al.* Integrated-gut-liver-on-a-chip platform as an in vitro human model of non-alcoholic fatty
1104 liver disease. *Commun. Biol.* **6**, 310 (2023).
1105 43. Kozyra, M. *et al.* Human hepatic 3D spheroids as a model for steatosis and insulin resistance. *Sci. Rep.*
1106 **8**, 14297 (2018).

1107 44. Kimura, M. *et al.* En masse organoid phenotyping informs metabolic-associated genetic susceptibility to
1108 NASH. *Cell* **185**, 4216-4232.e16 (2022).
1109 45. Balasubramaniam, S., Christodoulou, J. & Rahman, S. Disorders of riboflavin metabolism. *J. Inherit.*
1110 *Metab. Dis.* **42**, 608–619 (2019).
1111 46. Chen, Y.-S. *et al.* Effect of Vitamin B2 supplementation on migraine prophylaxis: a systematic review
1112 and meta-analysis. *Nutr. Neurosci.* **25**, 1801–1812 (2022).
1113 47. Park, K. *et al.* Flavin Adenine Dinucleotide Depletion Caused by electron transfer flavoprotein subunit
1114 alpha Haploinsufficiency Leads to Hepatic Steatosis and Injury in Zebrafish. *Hepatol. Commun.* **5**, 976–
1115 991 (2021).
1116 48. Tang, J. *et al.* Severe riboflavin deficiency induces alterations in the hepatic proteome of starter Pekin
1117 ducks. *Br. J. Nutr.* **118**, 641–650 (2017).
1118 49. Wang, Y. *et al.* Effects of riboflavin deficiency and high dietary fat on hepatic lipid accumulation: a
1119 synergetic action in the development of non-alcoholic fatty liver disease. *Nutr. Metab.* **21**, 1 (2024).
1120 50. Masschelin, P. M. *et al.* Vitamin B2 enables regulation of fasting glucose availability. *eLife* **12**, e84077
1121 (2023).
1122 51. Jusko, W. J. & Lewis, G. P. Effect of hepatic disease on liver flavin and protein levels. *Am. J. Clin. Nutr.*
1123 **25**, 265–268 (1972).
1124 52. Zhang, D. *et al.* Selenium and vitamin B_2 cosupplementation alleviates high-fat-diet-induced
1125 nonalcoholic fatty liver disease in rats by regulating hepatic lipid metabolism. *Eur. J. Lipid Sci. Technol.*
1126 **126**, 2300020 (2024).
1127 53. Kajiwara, M. *et al.* Donor-dependent variations in hepatic differentiation from human-induced
1128 pluripotent stem cells. *Proc. Natl. Acad. Sci.* **109**, 12538–12543 (2012).
1129 54. Seifert*, J. *et al.* Low nanomolar Triiodothyronine Concentrations Drive Further Maturation of hiPSC-
1130 Derived Hepatocyte-Like Cells. Preprint at https://doi.org/10.21203/rs.3.rs-6595009/v1 (2025).
1131 55. Ma, H. *et al.* The nuclear receptor THRB facilitates differentiation of human PSCs into more mature
1132 hepatocytes. *Cell Stem Cell* **29**, 795-809.e11 (2022).

1133 56. Nell, P. *et al.* Identification of an FXR-modulated liver-intestine hybrid state in iPSC-derived hepatocyte-
1134 like cells. *J. Hepatol.* **77**, 1386–1398 (2022).
1135 57. Charni-Natan, M. & Goldstein, I. Protocol for Primary Mouse Hepatocyte Isolation. *STAR Protoc.* **1**,
1136 100086 (2020).
1137 58. Horner, R. *et al.* Impact of Percoll purification on isolation of primary human hepatocytes. *Sci. Rep.* **9**,
1138 6542 (2019).
1139 59. Kwon, Y. *et al.* Induction of steatosis in primary human hepatocytes recapitulates key
1140 pathophysiological aspects of metabolic dysfunction-associated steatotic liver disease. *J. Hepatol.* **82**,
1141 18–27 (2025).
1142 60. Beier, N. *et al.* Adaptation of Dinoroseobacter shibae to oxidative stress and the specific role of RirA.
1143 *PloS One* **16**, e0248865 (2021).
1144 61. Lerch, M. F. *et al.* A non-coding RNA from the intercellular adhesion (ica) locus of Staphylococcus
1145 epidermidis controls polysaccharide intercellular adhesion (PIA)-mediated biofilm formation. *Mol.*
1146 *Microbiol.* **111**, 1571–1591 (2019).
1147 62. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**,
1148 337–342 (2011).
1149 63. Perez-Riverol, Y. *et al.* The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res.* **53**, D543–D553
1150 (2025).
1151 64. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware
1152 quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
1153 65. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq
1154 data with DESeq2. *Genome Biol.* **15**, 550 (2014).
1155 66. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30
1156 (2000).
1157 67. Ulgen, E., Ozisik, O. & Sezerman, O. U. pathfindR: An R Package for Comprehensive Identification of
1158 Enriched Pathways in Omics Data Through Active Subnetworks. *Front. Genet.* **10**, 858 (2019).

1159 68. Oughtred, R. *et al.* The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* **47**, D529–D541
1160 (2019).
1161 69. Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A. F. Discovering regulatory and signalling circuits in
1162 molecular interaction networks. *Bioinformatics* **18**, S233–S240 (2002).
1163 70. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids*
1164 *Res.* **32**, 258D – 261 (2004).
1165 71. Ge, S. X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants.
1166 *Bioinformatics* **36**, 2628–2629 (2020).
1167 72. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC*
1168 *Bioinformatics* **9**, 559 (2008).
1169 73. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat.*
1170 *Appl. Genet. Mol. Biol.* **4**, Article17 (2005).
1171 74. Mendez, D. *et al.* ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **47**, D930–
1172 D940 (2019).
1173 75. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids*
1174 *Res.* **46**, D1074–D1082 (2018).
1175 76. Szklarczyk, D. *et al.* STITCH 5: augmenting protein-chemical interaction networks with tissue and
1176 affinity data. *Nucleic Acids Res.* **44**, D380-384 (2016).
1177 77. Cheng, F., Kovács, I. A. & Barabási, A.-L. Network-based prediction of drug combinations. *Nat. Commun.*
1178 **10**, 1197 (2019).
1179 78. Myint, M. *et al.* Inflammatory signaling in NASH driven by hepatocyte mitochondrial dysfunctions. *J.*
1180 *Transl. Med.* **21**, 757 (2023).
1181 79. Kumar, P., Hassan, M., Tacke, F. & Engelmann, C. Delineating the heterogeneity of senescence-induced-
1182 functional alterations in hepatocytes. *Cell. Mol. Life Sci.* **81**, 200 (2024).
1183 80. Juanola, O. *et al.* Intestinal microbiota drives cholestasis-induced specific hepatic gene expression
1184 patterns. *Gut Microbes* **13**, 1911534 (2021).

81. Guillot, A., Kohlhepp, M., Bruneau, A., Heymann, F. & Tacke, F. Deciphering the Immune Microenvironment on A Single Archival Formalin-Fixed Paraffin-Embedded Tissue Section by An Immediately Implementable Multiplex Fluorescence Immunostaining Protocol. *Cancers* **12**, 2449 (2020).

82. Guillot, A. *et al.* Mapping the hepatic immune landscape identifies monocytic macrophages as key drivers of steatohepatitis and cholangiopathy progression. *Hepatology* **78**, 150–166 (2023).

Acknowledgements
We thank the members of the BIH Core Unit pluripotent Stem cells and Organoids, specifically Valeria Fernandez Vallone, Janine Cernoch and Claudia Schaar for their technical support. We also thank Janine Altmüller and Tatiana Borodina of the MDC/BIH Genomics Platform, Berlin for technical support. Additionally, we would like to thank the HPC for Research/Clinic cluster at the Berlin Institute of Health for providing data storage. The authors would like to acknowledge Novogene UK (Munich, Germany) for providing bulk RNA sequencing and bioinformatics analysis services.
We are also grateful for the support provided by all the current and former members of the Rezvani Lab, specifically Natalia Martagón Calderón and Reinhild Dünnebacke. We would also like to thank the following people for sharing their expertise: Marième Gueye, Amr Al Shebel, Tilman Breiderhoff, Kerstin Sommer, Sandra Gerstenberg, Verena Klämbt, Marah Alhabahbeh, Adrien Guillot, Hanyang Liu, Tian Lan and Guo Yin. Further, we want to thank Aras Mattis and lab members from the University of California San Francisco for discussion. Furthermore, we would like to extend our deepest gratitude to the individuals who have donated materials to this biomedical research.

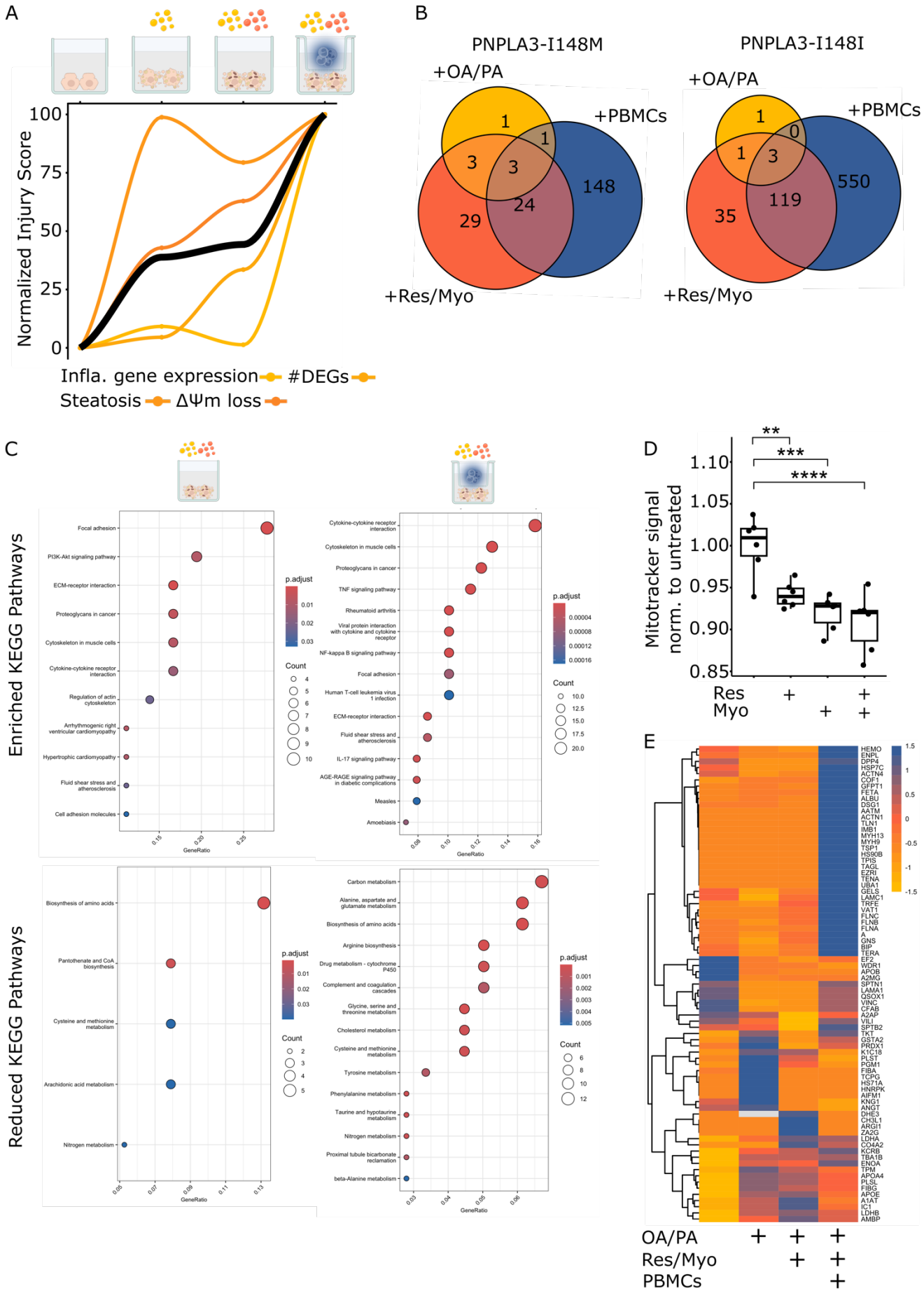
Author contributions
M.R. and Na.H. conceived the study. M.R., J.W., Na.H., F.B., G.Y. and A.C. designed and supervised the study. K.L. and H.S. contributed to protocol establishment. Ni.H., I.S. and P.T. provided and isolated clinical specimens. P.K. and C.E. contributed SeaHorse analysis. S.Q. and M.M. contributed to *in vitro* FLO and iHep assays. M.K. and S.E. contributed the Mass spec analysis. B.C. established the label-free AI brightfield imaging analysis. M.Ro. contributed the Lipidomics analysis. D.M. and C.H. contributed the characterization and proteomic analysis of the pediatric MASLD co-hort. I.L. and L.H. contributed with the full spectrum flow cytometry analysis. T.K. and J.F. analyzed the SteatoSITE co-hort. F.T., P.R. and P.B. participated in discussions and provided intellectual input. J.W., M.R., Na.H., F.B., G.Y. and A.C. wrote the paper.

Funding
The authors of this scientific article would like to acknowledge the funding provided by the Deutsche Forschungsgemeinschaft (DFG) Emmy Noether Programme to M. R. (RE 3749/2-1) and the Young Investigator Award awarded by Charite 3R to J.W. The Berlin-based authors (especially M.R., J.W. and F.T.) would also like to acknowledge the “CRC/TR 412”/“TRR 412” funded by the DFG. Na.H. is funded by LifeArc. SteatoSITE was funded by Innovate UK (Precision medicine: impacting through innovative technology - Ref: TS/R017581/1). P.R. is funded by an MRC Senior Clinical Fellowship (MR/W015919/1).

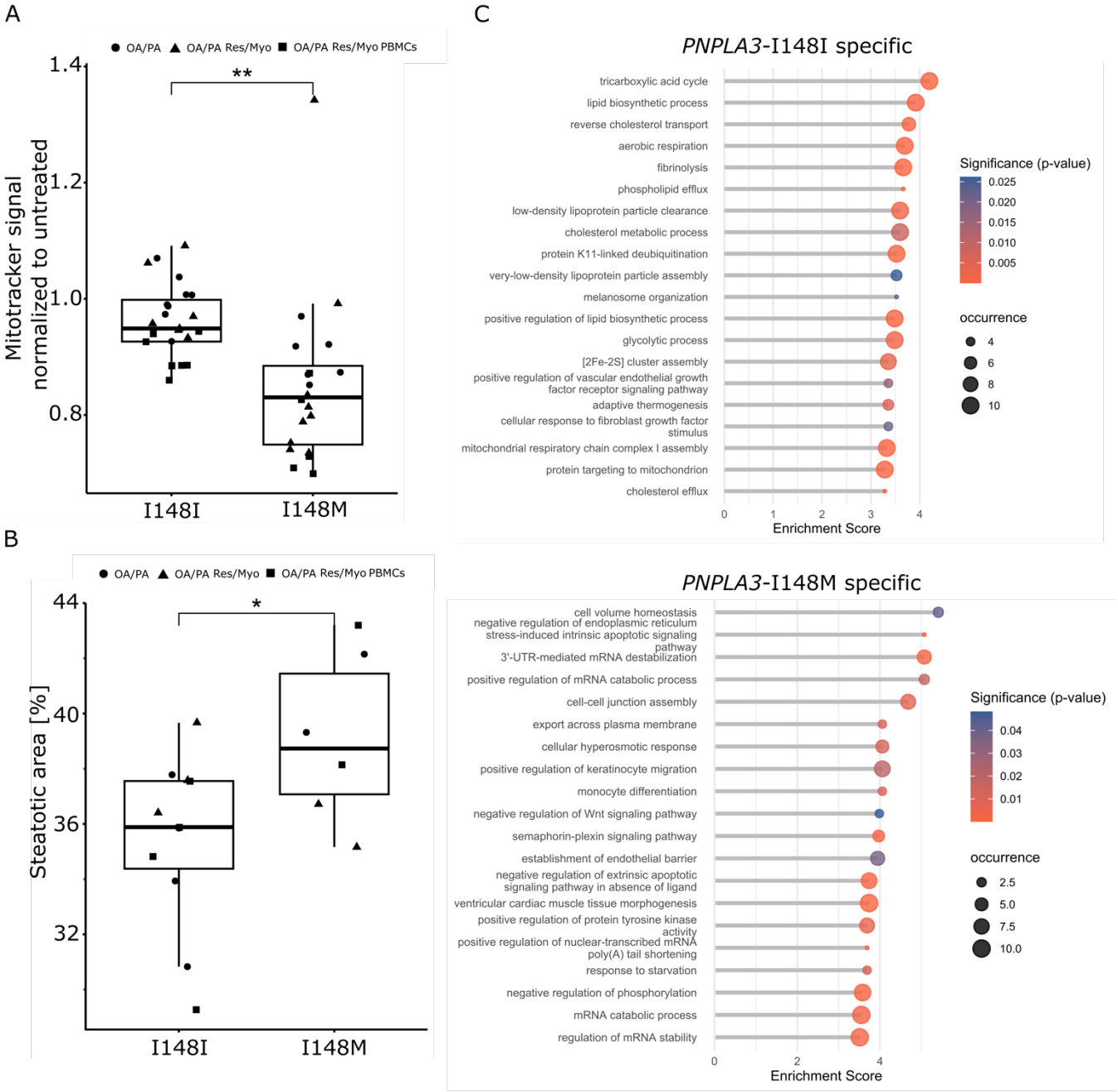
Competing interests
Na.H., M.R. and J.W. are inventors of the patent application EP24206430. Na.H. is a cofounder of KURE.ai and CardiaTec Biosciences. J.A.F serves as a consultant or advisory board member for Stellaris, Resolution Therapeutics, Kynos Therapeutics, Ipsen, River 2 Renal Corp., Stimuliver, and Global Clinical Trial Partners, has received speaker fees from HistoIndex and Resolution Therapeutics, and research grant funding from GlaxoSmithKline and Genentech. P.R. has served as a consultant for MSD and Macomics and has received research support from Genentech, Intercept and Neogenomics.

Code availability
Publicly available packages were used in all analyses, all of which are referenced in the Methods section along with the specific parameters applied.

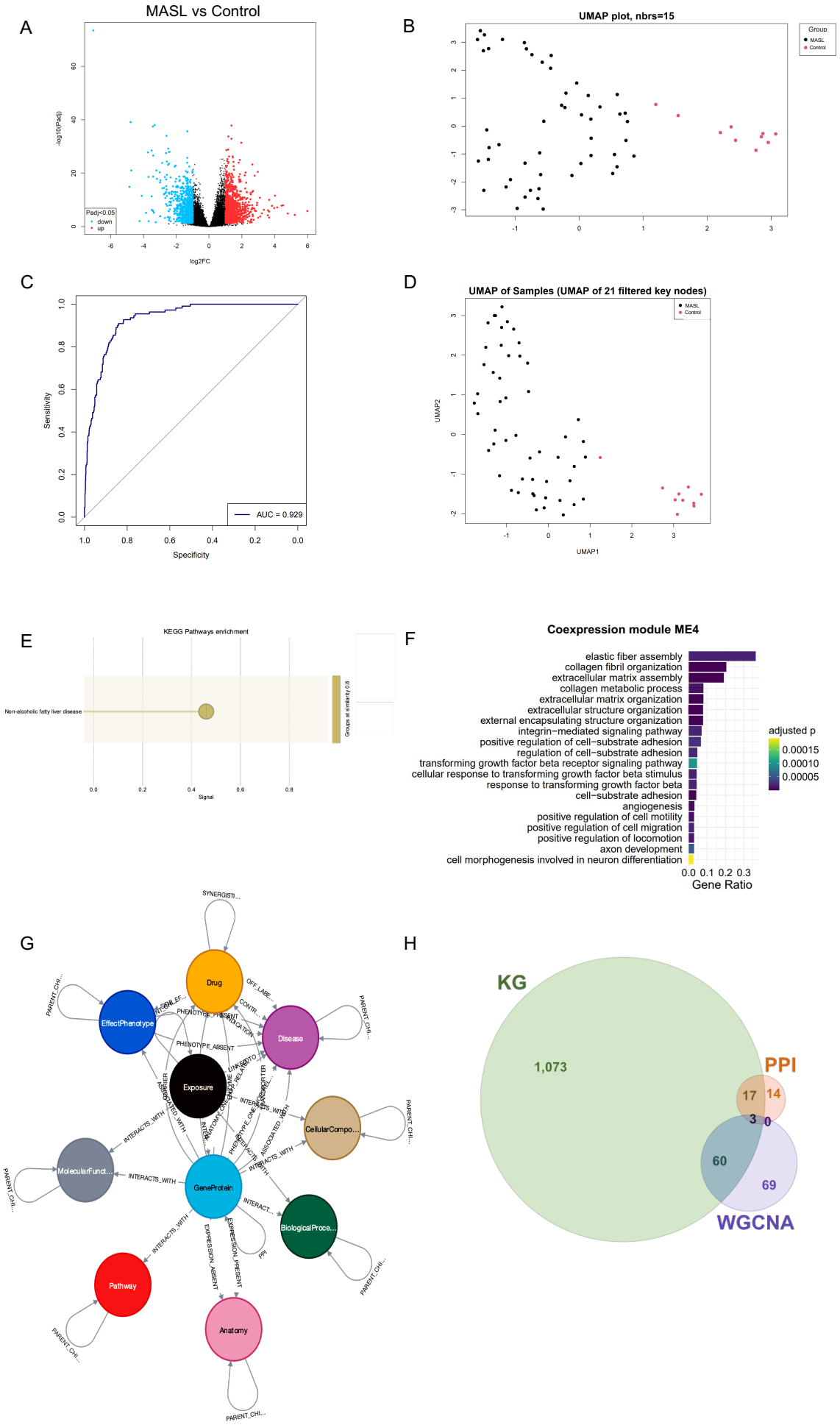
1237
1238 **Supplementary Information**
1239 **Supplementary Tables**
1240 **Supplementary Table 1:** KEGG pathways perturbed in PNPLA3-I148I iHeps exposed to different MASLD
1241 models
1242 **Supplementary Table 2:** Detected human proteins of Mass Spec analysis of conditioned medium from
1243 PNPLA3-I148I iHeps exposed to different MASLD models.
1244 **Supplementary Table 3:** PathfindR analysis for identifying cell line specific or shared pathways in iHeps
1245 exposed to the 3-Hit MASLD model
1246 **Supplementary Table 4:** Compound prediction derived from PPIs
1247 **Supplementary Table 5:** Compound prediction derived from WGCNA
1248 **Supplementary Table 6:** Compound prediction derived from KG
1249 **Supplementary Table 7:** Compounds predicted in multiple computational approaches
1250 **Supplementary Table 8:** Predictive 10-Gene flavo-signature
1251 **Supplementary Table 9:** Features derived by label-free brightfield imaging different between diseased cells
1252 with or without FAD^H assessed for significance
1253 **Supplementary Table 10:** Flavo-active compounds
1254
1255 **Supplementary Figures:**


A
Normalized Injury Score
100
75
50
25
0
Infla. gene expression
#DEGs
Steatosis
ΔΨm loss
B
PNPLA3-I148M
+OA/PA
+PBMCs
+Res/Myo
1
1
3
3
148
24
29
PNPLA3-I148I
+OA/PA
+PBMCs
+Res/Myo
1
0
1
3
550
119
35
C
Enriched KEGG Pathways
Focal adhesion
PI3K-Akt signaling pathway
ECM-receptor interaction
Proteoglycans in cancer
Cytoskeleton in muscle cells
Cytokine-cytokine receptor interaction
Regulation of actin cytoskeleton
Arrhythmogenic right ventricular cardiomyopathy
Hypertrophic cardiomyopathy
Fluid shear stress and atherosclerosis
Cell adhesion molecules
GeneRatio
p.adjust
Count
Cytokine-cytokine receptor interaction
Cytoskeleton in muscle cells
Proteoglycans in cancer
TNF signaling pathway
Rheumatoid arthritis
Viral protein interaction with cytokine and cytokine receptor
NF-kappa B signaling pathway
Focal adhesion
Human T-cell leukemia virus 1 infection
ECM-receptor interaction
Fluid shear stress and atherosclerosis
IL-17 signaling pathway
AGE-RAGE signaling pathway in diabetic complications
Measles
Amoebiasis
GeneRatio
p.adjust
Count
Reduced KEGG Pathways
Biosynthesis of amino acids
Pantothenate and CoA biosynthesis
Cysteine and methionine metabolism
Arachidonic acid metabolism
Nitrogen metabolism
GeneRatio
p.adjust
Count
Carbon metabolism
Alanine, aspartate and glutamate metabolism
Biosynthesis of amino acids
Arginine biosynthesis
Drug metabolism - cytochrome P450
Complement and coagulation cascades
Glycine, serine and threonine metabolism
Cholesterol metabolism
Cysteine and methionine metabolism
Tyrosine metabolism
Phenylalanine metabolism
Taurine and hypotaurine metabolism
Nitrogen metabolism
Proximal tubule bicarbonate reclamation
beta-Alanine metabolism
GeneRatio
p.adjust
Count
D
Mitotracker signal norm. to untreated
1.10
1.05
1.00
0.95
0.90
0.85
**
***
****
Res
Myo
E
OA/PA
Res/Myo
PBMCs

Supplementary Fig. 1: Characterizing the MASLD drug testing model in stem cell-derived hepatocytes (iHeps). A) MASLD Injury index in PNPLA3-I148M iHeps exposed to different MASLD models. All parameters were standardized, ranging from Healthy (0%) to the most complex disease condition including all disease triggers (100%). Inflammatory (Infla.) gene expression includes the genes *TNFA*, *CXCL8*, *CXCL1* and *CCL2*. ΔΨm = mitochondrial membrane potential **B)** Venn diagram displaying the number of differentially expressed genes (DEGs) from bulk RNA sequencing across disease conditions relative to the untreated control. Circle sizes are based on the log2-transformed DEG number. **C)** KEGG Pathway enrichment analysis for the up- and down-regulated genes in *PNPLA3*-I148I iHeps, for the 2-Hit and 3-Hit MASLD model. The 1-Hit model did not have enough DEGs for enrichment analysis and is therefore not shown. **D)** Mitotracker signal of *PNPLA3*-I148M iHeps treated with resistin (Res), myostatin (Myo), or both. Significance was assessed using multiple t-test, corrected according to the Benjamini-Hochberg method. Each dot represents an independent sample **F)** Heatmap of the mean IBAQ values of the identified human proteins. Proteins were identified in the supernatants of untreated iHeps and iHeps exposed to different stimuli (OA/PA = oleic and palmitic acid, Res/Myo, PBMCs = peripheral blood mononuclear cells) using LC-MS/MS. Heatmap shows the mean IBAQ values of n = 2 independent wells (with the exception of untreated iHeps (n=1)) (see **Supp. Table 2**).

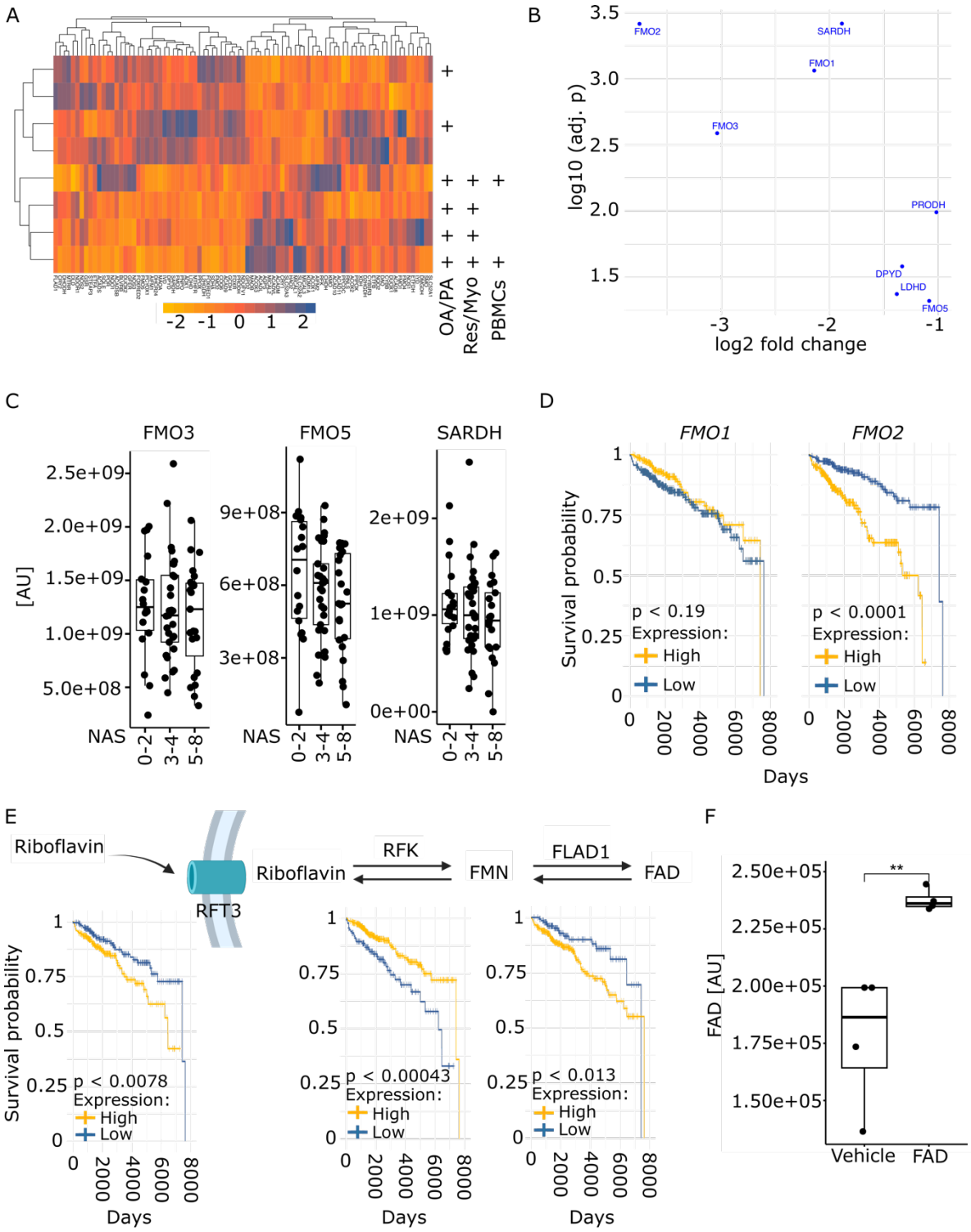


Supplementary Fig. 2: Identifying donor specific effects in stem cell-derived hepatocytes (iHeps) with or without genetic MASLD predisposition A) Mitotracker signal in *PNPLA3*-I148I and -I148M iHeps, normalized to the respective untreated control. Significance was assessed using a t-test. Each dot represent an independent sample. **B)** Steatotic area in *PNPLA3*-I148I and -I148M iHeps, normalized to the mitochondrial area. Significance was assessed using a t-test. Each dot represent an independent sample. **C)** PathfindR analysis for enriched Pathways in the 3-Hit disease model specific the *PNPLA3*-I148I or -148M iHeps.


A
MASL vs Control
-log10(Padj)
log2FC
Padj<0.05
down
up
B
UMAP plot, nbrs=15
Group
MASL
Control
C
Sensitivity
Specificity
AUC = 0.929
D
UMAP of Samples (UMAP of 21 filtered key nodes)
MASL
Control
UMAP2
UMAP1
E
KEGG Pathways enrichment
Non-alcoholic fatty liver disease
Groups at similarity 0.8
Signal
F
Coexpression module ME4
elastic fiber assembly
collagen fibril organization
extracellular matrix assembly
collagen metabolic process
extracellular matrix organization
extracellular structure organization
external encapsulating structure organization
integrin−mediated signaling pathway
positive regulation of cell−substrate adhesion
regulation of cell−substrate adhesion
transforming growth factor beta receptor signaling pathway
cellular response to transforming growth factor beta stimulus
response to transforming growth factor beta
cell−substrate adhesion
angiogenesis
positive regulation of cell motility
positive regulation of cell migration
positive regulation of locomotion
axon development
cell morphogenesis involved in neuron differentiation
adjusted p
0.00015
0.00010
0.00005
Gene Ratio
G
Drug
EffectPhenotype
Disease
Exposure
CellularCompo...
MolecularFunct...
GeneProtein
BiologicalProce...
Pathway
Anatomy
H
KG
1,073
PPI
17
14
3
0
60
69
WGCNA

1310

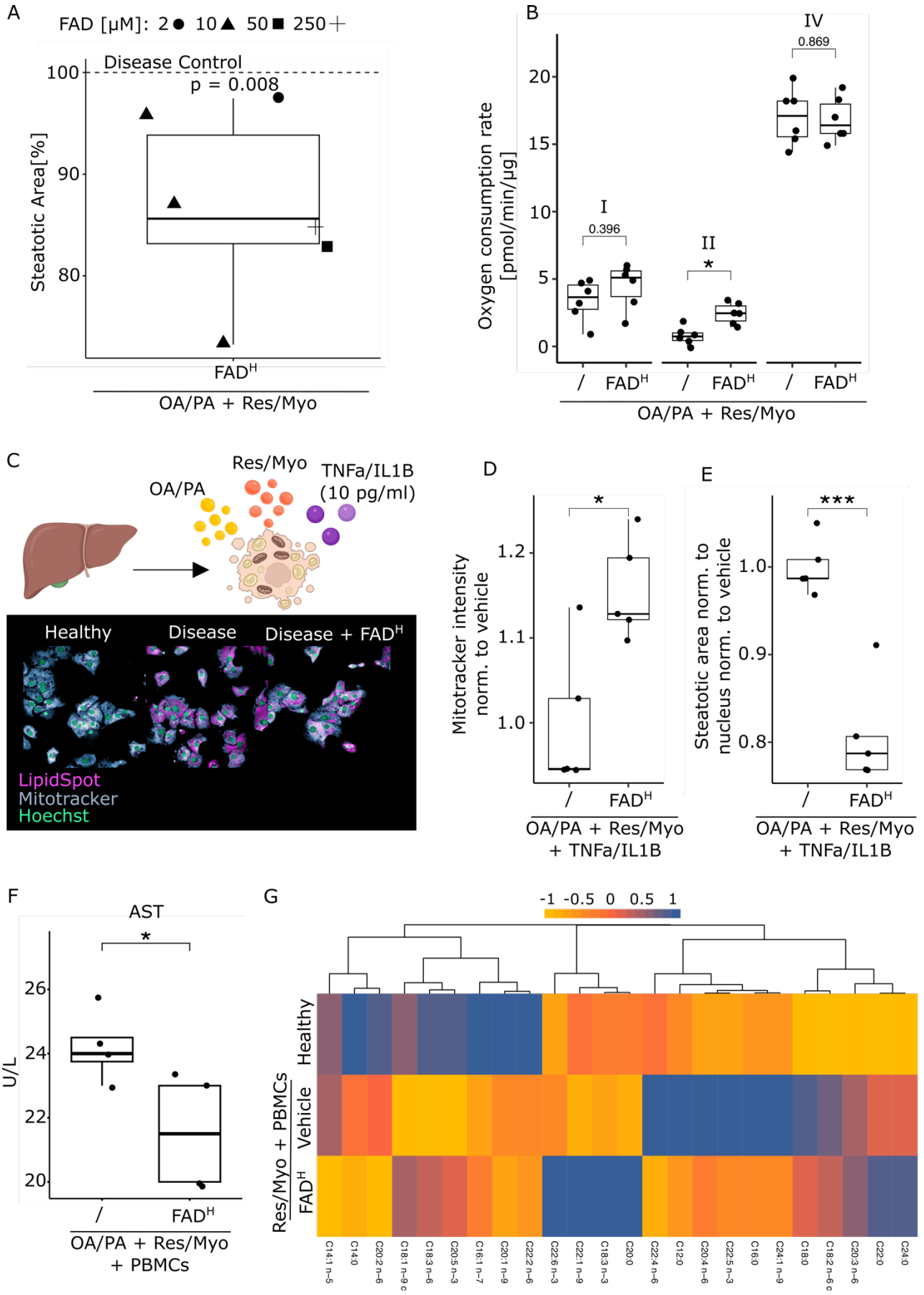
Supplementary Fig. 3: Validation of disease signature, module enrichment, and compound convergence across computational pipelines. Supporting analyses for transcriptomic signature refinement, co-expression module evaluation, and compound overlap across the PPI, WGCNA, and KG pipelines. **A)** UMAP projection of bulk RNA-seq samples (GSE135251) based on global gene expression profiles. Samples cluster by condition, with MASL (red) clearly separated from control (black), indicating robust transcriptomic differences. **B)** Volcano plot showing differential expression analysis between MASL and control samples. Significantly upregulated (red) and downregulated (blue) genes are highlighted (adjusted $p < 0.05$), with log2 fold change on the x-axis and –log10 p-value on the y-axis. **C)** Receiver operating characteristic (ROC) curve demonstrating the classification performance of the refined hepatocyte-specific MASL signature. The area under the curve (AUC) is 0.929, indicating strong discriminative power. **D)** UMAP embedding of bulk RNA-seq samples based on the expression of 21 key network nodes identified from the PPI analysis. The filtered signature maintains separation between MASL (red) and control (black) groups. **E)** KEGG pathway enrichment analysis of the 21 key PPI nodes. The only significantly enriched term was "Non-alcoholic fatty liver disease," supporting the specificity of the derived subnetwork. **F)** Top 20 enriched Gene Ontology (GO) terms for WGCNA module ME4, which was positively correlated with fibrosis severity. Terms include extracellular matrix organization, integrin signaling, and cellular migration. Bars are ranked by gene ratio and colored by adjusted p-value. **G)** Schematic overview of the PrimeKG graph schema showing major node types (e.g., Drug, Disease, Gene/Protein, Exposure, Phenotype, etc.) and example edge types (e.g., "interacts with," "associated with," "phenotype present"). This reflects the ontology structure used in KG-based query design. **H)** Venn diagram summarizing the overlap among predicted compounds from the three computational pipelines. Out of 1,199 unique compounds, three—flavin adenine dinucleotide (FAD), pyridoxal phosphate (PLP), and NADH—were identified by all three methods (KG, WGCNA, and PPI).



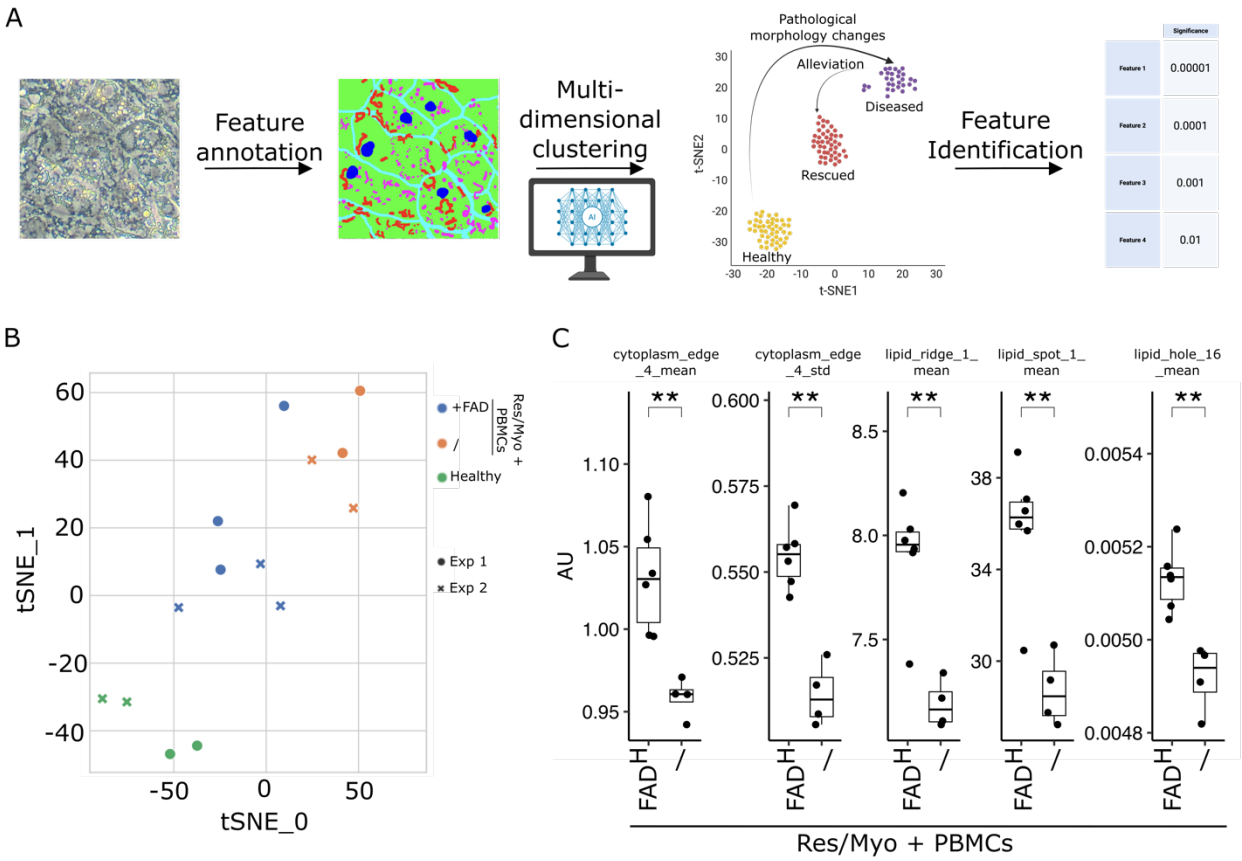
1347

Supplementary Fig. 4: Perturbance and manipulation of FAD homeostasis in MASLD A) Heatmap of normalized gene counts showing clustering of FAD-dependent protein gene expression across different MASLD models in an independent differentiation of PNPLA3-I148I iHeps **B)** DEGs the 3-Hit *in vitro* disease model in relation to healthy control filtered for the human flavoproteome. **C)** Abundance of *FMO3*, *FMO5* and *SARDH* proteins in the pediatric MASLD cohort described by Hudert et al. Significance was assessed using multiple t-tests, corrected by the Benjamini-Hochberg method. Each dot represents and independent patient **D)** Survival analysis of FMO-gene expression and human MASLD patients from the SteatoSITE co-hort. **E)** Survival analysis of FAD-homeostasis gene expression and human MASLD patients from the SteatoSITE co-hort. **F)** Concentration of FAD in iHeps exposed to the 3-Hit MASLD model after 3-day treatment with or without 50 µM of FAD. Each dot represents an independent sample.

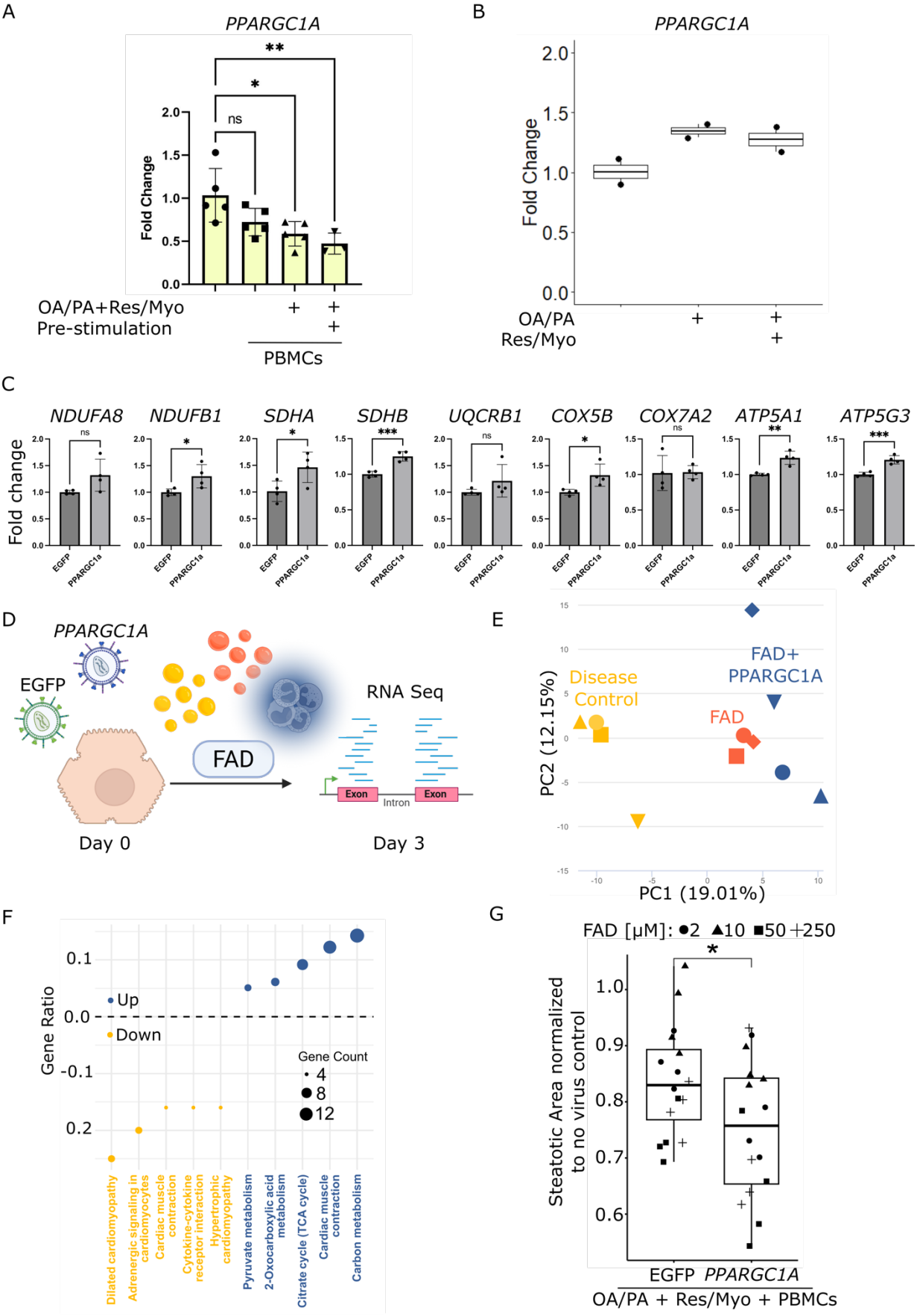
1348


A
FAD [μM]: 2 10 50 250
Disease Control
p = 0.008
Steatotic Area[%]
100
90
80
FAD^H
OA/PA + Res/Myo
B
Oxygen consumption rate [pmol/min/μg]
I
0.396
II
*
IV
0.869
/
FAD^H
OA/PA + Res/Myo
C
OA/PA
Res/Myo
TNFa/IL1B (10 pg/ml)
Healthy
Disease
Disease + FAD^H
LipidSpot
Mitotracker
Hoechst
D
Mitotracker intensity norm. to vehicle
*
OA/PA + Res/Myo + TNFa/IL1B
E
Steatotic area norm. to nucleus norm. to vehicle
***
OA/PA + Res/Myo + TNFa/IL1B
F
AST
U/L
*
OA/PA + Res/Myo + PBMCs
G
-1 -0.5 0 0.5 1
Healthy
Vehicle
FAD^H
Res/Myo + PBMCs
C14:1 n-5
C14:0
C20:2 n-6
C18:1 n-9 c
C18:3 n-6
C20:5 n-3
C16:1 n-7
C20:1 n-9
C22:2 n-6
C22:6 n-3
C22:1 n-9
C18:3 n-3
C20:0
C22:4 n-6
C12:0
C20:4 n-6
C22:5 n-3
C16:0
C24:1 n-9
C18:0
C18:2 n-6 c
C20:3 n-6
C22:0
C24:0

Supplementary Fig. 5: Characterizing the MASLD alleviating effects of FAD. A) Quantification of the steatotic area normalized to the nucleus count in iHeps exposed to the 2-Hit MASLD model with different concentrations of FAD. Significance was used using a one-sided one sample t-test using the 100% vehicle control marked by the dotted line as a reference. Each dot represents an independent sample **B)** Mitochondrial complex activity (I, II and IV) in mouse hepatocytes exposed to the 2 Hit MASLD model with or without 50 μM FAD^H treatment. Significance was assessed using multiple t-test, corrected according to the Benjamini-Hochberg method. Each dot represents an independent sample **C)** Scheme, representative images and quantification of mitochondrial membrane potential **(D)** and steatotic area **(E)** for validating FAD effects in primary human hepatocytes exposed to OA/PA, Res/Myo and the pro-inflammatory cytokines IL1B and TNFA. Mitochondrial membrane potential and steatotic area were normalized to the Vehicle control (plus steatotic area was normalized to the nucleus count). Each dot represents an independent sample **F)** Concentration of AST in culture supernatant from iHeps exposed to the 3-Hit disease model, with or without 50 μM of FAD^H. Cells were transduced with EGFP before disease modeling. Each dot represents an independent sample **G)** Lipidomic heatmap analysis of iHeps either untreated (healthy control) or exposed to the disease trigger Res/Myo and trans-well PBMC co-culture with or without FAD^H (50 μM). Fatty acid species abundance was normalized by protein concentration and to total amount of measured fatty acid content. Heatmap shows the mean of n = 4-6 independent samples.

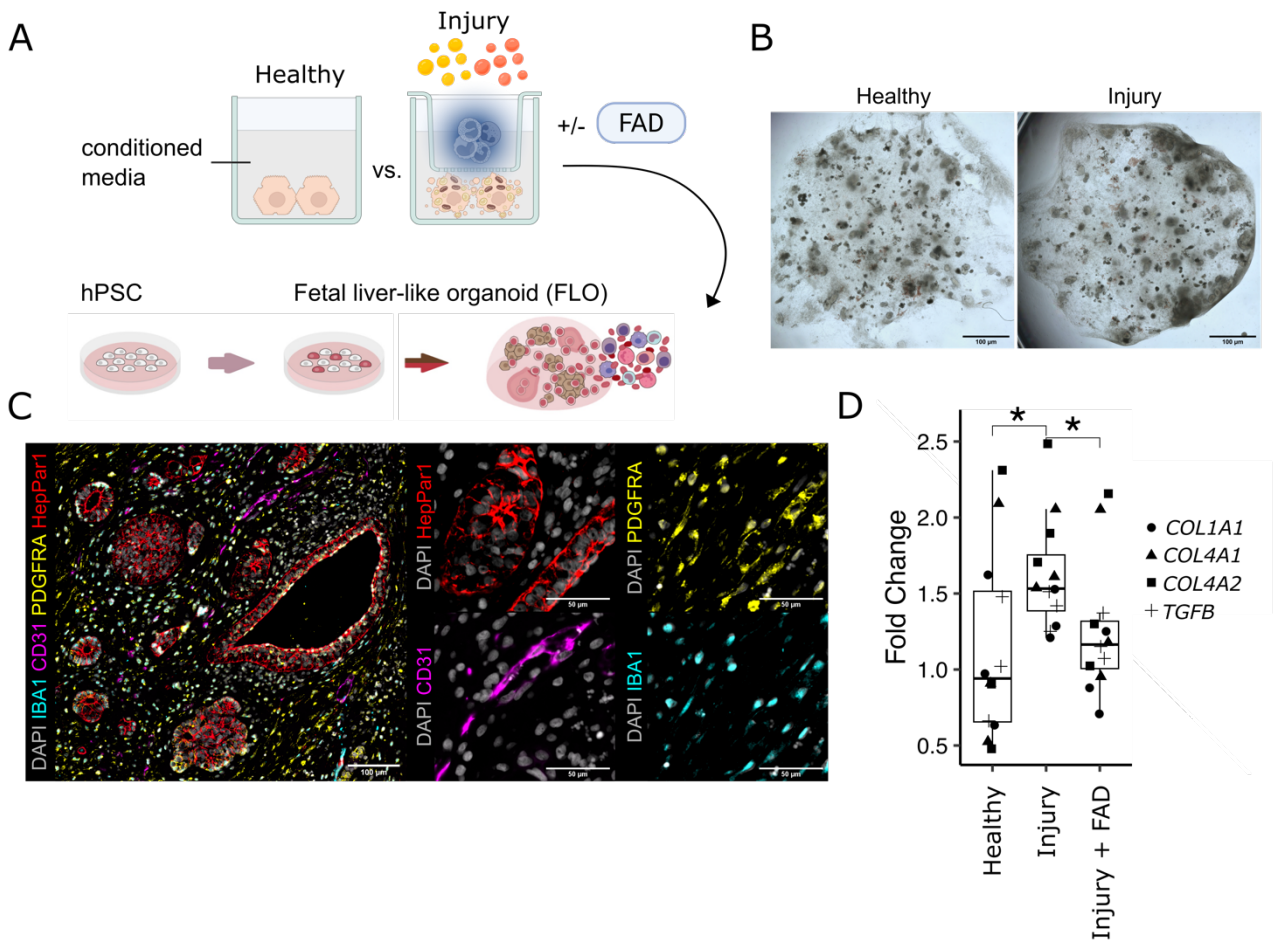


Supplementary Fig. 6: Label-free brightfield imaging analysis for detecting rescue effects of FAD. A) Scheme of label-free brightfield imaging analysis. **B)** tSNE plot of iHeps exposed to Res/Myo and PBMCs with or without FAD^H (50 µM) or being left untreated. **C)** Top 5 morphological parameter different between iHeps exposed to Res/Myo + PBMCs with or without FAD treatment, assessed for significance using t-tests. Each dot represents an independent sample.

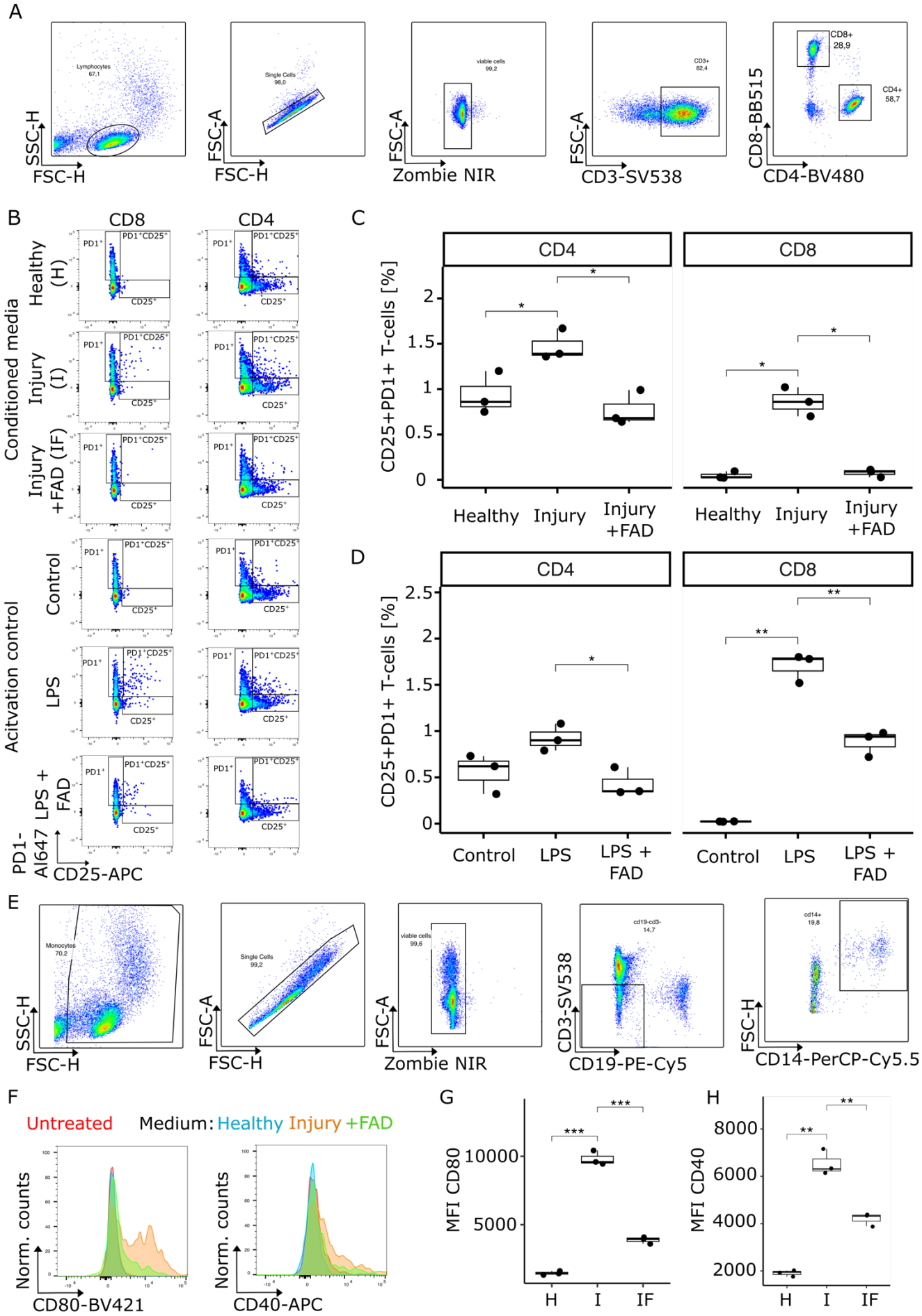

A
PPARGC1A
Fold Change
ns
*
**
OA/PA+Res/Myo
Pre-stimulation
PBMCs
B
PPARGC1A
Fold Change
OA/PA
Res/Myo
C
NDUFA8
NDUFB1
SDHA
SDHB
UQCRB1
COX5B
COX7A2
ATP5A1
ATP5G3
Fold change
EGFP
PPARGC1a
D
PPARGC1A
EGFP
FAD
RNA Seq
Exon
Intron
Day 0
Day 3
E
Disease
Control
FAD
FAD+
PPARGC1A
PC2 (12.15%)
PC1 (19.01%)
F
Gene Ratio
Up
Down
Gene Count
Dilated cardiomyopathy
Adrenergic signaling in cardiomyocytes
Cardiac muscle contraction
Cytokine-cytokine receptor interaction
Hypertrophic cardiomyopathy
Pyruvate metabolism
2-Oxocarboxylic acid metabolism
Citrate cycle (TCA cycle)
Cardiac muscle contraction
Carbon metabolism
G
FAD [μM]: ●2 ▲10 ■50 +250
Steatotic Area normalized to no virus control
EGFP
PPARGC1A
OA/PA + Res/Myo + PBMCs

1418

Supplementary Fig. 7: Characterizing changes and additional rescuing effects of *PPARGC1A* and its overexpression in MASLD iHeps. A) Gene expression levels of *PPARGC1A* in iHeps exposed to different PBMC co-culture conditions, normalized to untreated cells. Dots represent independent samples. Significance was assessed using one-way anova with Tukey-Post Hoc **B)** Gene expression levels of *PPARGC1A* in iHeps exposed to OA/PA or OA/PA with Res/Myo. Gene expression levels were normalized to the untreated control. Dots represent independent samples. **C)** Expression levels of mitochondrial complex genes in iHeps transduced with *PPARGC1A* or EGFP control assessed via qPCR. Significance was assessed using t-tests. Dots represent independent samples **D)** Schematic of the RNA seq experiment analyzing iHeps transduced with *PPARGC1A* or EGFP control, exposed to the 3-Hit disease model, with or without 50 μM FAD^H. **E)** PCA plot, displaying all tested conditions. **F)** Top KEGG pathways enriched in iHeps exposed to FAD^H plus *PPARGC1A overexpression* relative to FAD^H-treated EGFP control. **G)** Steatotic area in iHeps exposed to the 3-Hit disease model, transduced with EGFP or *PPARGC1A* with different concentrations of FAD. Statistical significance was assessed using a t-test. Dots represent independent samples. Steatotic area was normalized to nuclei count and is relative to the non-transduced vehicle control of each condition.



Supplementary Fig. 8: Characterizing FAD effects in multi-cellular fetal liver organoids (FLOs). A) Scheme of evaluating anti-TGFB effects of FAD^H in FLOs exposed to supernatant from injured iHeps. **B)** Brightfield images of FLOs assessing encapsulation, treated with conditioned medium from healthy iHeps or iHeps undergoing the 3-hit MASLD injury. **C)** Immunofluorescent images of FLOs, showing the abundance of hepatocytes ($HepPar1^+$), stellate cells ($PDGFRA^+$), endothelial cells ($CD31^+$) and macrophages ($IBA1^+$). **D)** Gene expression analysis of fibrosis signature genes in FLOs exposed to conditioned medium from healthy of injured iHeps (3-Hit MASLD model) with or without FAD^H (50 µM). For each gene, each dot represents an independent organoid. Significance was assessed using multiple t-test corrected according to the Benjamini-Hochberg method.


A
Lymphocytes 67,1
SSC-H
FSC-H
Single Cells 98,0
FSC-A
FSC-H
viable cells 99,2
FSC-A
Zombie NIR
CD3+ 82,4
FSC-A
CD3-SV538
CD8+ 28,9
CD4+ 58,7
CD8-BB515
CD4-BV480
B
CD8
CD4
Conditioned media
Healthy (H)
Injury (I)
Injury +FAD (IF)
Acitvation control
Control
LPS
LPS + FAD
PD1+
PD1+CD25+
CD25+
PD1-Al647
CD25-APC
C
CD4
CD8
CD25+PD1+ T-cells [%]
Healthy
Injury
Injury +FAD
D
CD4
CD8
CD25+PD1+ T-cells [%]
Control
LPS
LPS + FAD
E
Monocytes 70,2
SSC-H
FSC-H
Single Cells 99,2
FSC-A
FSC-H
viable cells 99,6
FSC-A
Zombie NIR
cd19-cd3- 14,7
CD3-SV538
CD19-PE-Cy5
cd14+ 19,8
FSC-H
CD14-PerCP-Cy5.5
F
Untreated
Medium: Healthy Injury +FAD
Norm. counts
CD80-BV421
Norm. counts
CD40-APC
G
MFI CD80
H
I
IF
H
MFI CD40
H
I
IF

1482

Supplementary Fig. 9: Characterizing anti-inflammatory effects by FAD^H in primary human immune cells A) Full spectrum flow cytometry gating strategy for CD4 and CD8 T-cell characterization in PBMCs isolated from a healthy individual **B)** Representative dot plots of T-cell activation marker in $CD4^+$ and $CD8^+$ T-cells stimulated with iHep supernatant or LPS exposure with or without FAD^H. **C)** Quantification of T-cells double positive for the activation marker PD1 and CD25 after exposure to iHep supernatants with or without FAD^H. Multiple t-test to assess significance, corrected according to the Benjamini-Hochberg method. **D)** Quantification of T-cells double positive for the activation marker PD1 and CD25 after exposure to LPS with or without FAD^H. Multiple t-test to assess significance, corrected according to the Benjamini-Hochberg method. **E)** Full spectrum flow cytometry gating strategy for $CD14^+$ monocyte characterization in PBMCs isolated from a healthy individual. **F)** Representative histogram plots for the activation marker CD40 and CD80 in $CD14^+$ monocytes after exposure to iHep supernatants with or without FAD^H. **G)** Quantification of the parameters shown in F). Multiple t-test to assess significance, corrected according to the Benjamini-Hochberg method.