

1 **A prognostic molecular signature of hepatic steatosis is spatially heterogeneous and**
2 **dynamic in human liver**

3 Andrew S. Perry^{1*}, Niran Hadad^{2*}, Emeli Chatterjee^{3*}, Maria Jimenez Ramos^{4**}, Eric Farber-
4 Eger^{1**}, Rashedeh Roshani^{5**}, Lindsay K. Stolze¹, Shilin Zhao¹, Liesbet Martens^{6,7}, Timothy J.
5 Kendall^{4,8}, Tinne Thone^{6,7}, Kaushik Amancherla¹, Samuel Bailin⁹, Curtis L. Gabriel⁹, John
6 Koethe⁹, J. Jeffrey Carr¹, James Greg Terry¹, Jane Freedman¹, Kahraman Tanriverdi¹, Eric
7 Alsop², Kendall Van Keuren-Jensen², John F.K. Sauld¹⁰, Gautam Mahajan¹⁰, Sadiya Khan¹¹,
8 Laura Colangelo¹², Matthew Naylor¹³, Susan Fisher-Hoch¹⁴, Joseph McCormick¹⁴, Kari E.
9 North¹⁵, Jennifer Below⁵, Quinn Wells¹, Dale Abel¹⁶, Ravi Kalhan¹¹, Charlotte Scott^{6,7}, Martin
10 Guilliams^{6,7}, Jonathan A. Fallowfield^{4*}, Nicholas E. Banovich^{2*}, Saumya Das^{3*}, Ravi Shah^{1*}

11
12 Affiliations:

13 ¹Vanderbilt Translational and Clinical Cardiovascular Research Center, Vanderbilt University
14 School of Medicine, Nashville, TN, USA
15 ²Translational Genomics Research Institute, Phoenix, AZ, USA
16 ³Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts,
17 USA
18 ⁴Institute for Regeneration and Repair, University of Edinburgh, Edinburgh, UK
19 ⁵Vanderbilt Genetics Institute, Division of Genetic Medicine, Department of Medicine, Vanderbilt
20 University Medical Center, Nashville, TN, USA
21 ⁶Laboratory of Myeloid Cell Biology in Tissue Homeostasis and Regeneration, VIB-UGent
22 Center for Inflammation Research, Ghent, Belgium
23 ⁷Department of Biomedical Molecular Biology, Ghent University, Belgium.
24 ⁸Edinburgh Pathology, University of Edinburgh, Edinburgh, UK
25 ⁹Division of Infectious Diseases, Department of Medicine, Vanderbilt University Medical Center,
26 Nashville, TN, USA
27 ¹⁰Emulate Inc., Boston, MA, USA
28 ¹¹Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
29 ¹²Department of Preventive Medicine, Northwestern University Feinberg School of Medicine,
30 Chicago, IL, USA
31 ¹³Sections of Cardiovascular Medicine and Preventive Medicine and Epidemiology, Department
32 of Medicine, Boston University School of Medicine, Boston, MA, USA
33 ¹⁴School of Public Health, The University of Texas Health Science Center at Houston,
34 Brownsville, TX, USA
35 ¹⁵CVD Genetic Epidemiology Computational Laboratory, Gillings School of Global Public Health,
36 University of North Carolina, Chapel Hill, NC, USA
37 ¹⁶Department of Medicine, David Geffen School of Medicine and UCLA Health, University of
38 California-Los Angeles, Los Angeles, CA, USA
39

1 A.S.P, N.H, and E.C. are co-first authors.
2 M.J.R., E.F.E, and R.R. contributed equally to this work.
3 J.A.F., N.E.B., S.D., and R.S. are co-senior authors.

4
5 **Corresponding authors:**

6 Ravi Shah
7 Vanderbilt Translational and Clinical Research Center
8 Vanderbilt University Medical Center
9 2525 West End Ave, Suite 300
10 Nashville, TN 37203, USA
11 ravi.shah@vumc.org

12
13 Jonathan A. Fallowfield
14 Institute for Regeneration & Repair, University of Edinburgh,
15 Edinburgh BioQuarter, 4-5 Little France Drive, Edinburgh, EH16 4UU
16 Jonathan.Fallowfield@ed.ac.uk

17
18 Nicholas E. Banovich
19 The Translational Genomics Research Institute
20 445 N. Fifth Street
21 Phoenix, AZ 85004, USA
22 nbanovich@tgen.org

23
24 Saumya Das
25 Simches Cardiovascular Research Center
26 Massachusetts General Hospital
27 185 Cambridge Street
28 Boston, MA 02114, USA
29 sdas@mgh.harvard.edu

1 **SUMMARY**

2 Metabolic dysfunction-associated steatotic liver disease (MASLD) prevalence is increasing in
3 parallel with an obesity pandemic, calling for novel strategies for prevention and treatment. We
4 defined a circulating proteome of human MASLD across ≈ 7000 proteins in ≈ 5000 individuals
5 from diverse, at-risk populations across the metabolic health spectrum, demonstrating
6 reproducible diagnostic performance and specifying both known and novel metabolic pathways
7 relevant to MASLD (central carbon and amino acid metabolism, hepatocyte regeneration,
8 inflammation, fibrosis, insulin sensitivity). A parsimonious proteomic signature of MASLD was
9 associated with a protection from MASLD and its related multi-system metabolic consequences
10 in >26000 free-living individuals, with an additive effect to polygenic risk. The MASLD proteome
11 was encoded by genes that demonstrated transcriptional enrichment in liver, with spatial
12 transcriptional activity in areas of steatosis in human liver biopsy and dynamicity for select
13 targets in human liver across stages of steatosis. We replicated several top relations from
14 proteomics and spatial tissue transcriptomics in a humanized “liver-on-a-chip” model of MASLD,
15 highlighting the power of a full translational approach to discovery in MASLD. Collectively, these
16 results underscore utility of blood-based proteomics as a dynamic “liquid biopsy” of human liver
17 relevant to clinical biomarker and mechanistic applications.

18

19 **Keywords:** metabolic dysfunction-associated steatotic liver disease, proteomics, diabetes

1 INTRODUCTION

2 Metabolic dysfunction-associated steatotic liver disease (MASLD) is present in >30% of
3 individuals worldwide¹ and has emerged as a predominant driver of end-stage liver disease and
4 an important contributor to a range of other chronic, life-limiting illnesses (e.g., cancer,
5 cardiovascular disease, renal dysfunction)². The heterogeneous progression from MASLD to
6 chronic steatohepatitis—and limited effectiveness of therapies that interrupt this process—have
7 fueled efforts to identify at-risk populations early for prevention. While large studies of genetic
8 propensity for MASLD have yielded promising targets (e.g., *PNPLA3*³), genetic variation alone
9 may not capture dynamic behavioral and environmental contributions (diet, obesity, diabetes,
10 etc.) that may be reflected in molecular states that precede MASLD. Circulating proteomics⁴⁻⁶
11 and hepatic tissue profiles⁷ have been at the forefront of resolving these limitations in MASLD,
12 though limited by lack of integration of circulating “omics” and liver characterization or reliance
13 on invasive biopsy-guided approaches. In response, a few recent studies have begun to
14 integrate liver phenotypes and circulating proteomics in humans with some promising results,
15 albeit in small cohorts with available liver tissue and in targeted populations with specific
16 exposures (e.g., alcohol)^{8,9}. Integrative studies that link the circulating proteome to dynamic
17 hepatic tissue states across the MASLD spectrum (at single cell and spatial resolution) and
18 large clinical longitudinal studies to assess clinical susceptibility and consequences of MASLD
19 will provide the broad translational landscape necessary to simultaneously specify diagnostic
20 and prognostic biomarkers and prioritize dynamic targets for mechanistic inquiry.

21 We leveraged three large prospective observational cohorts (N=4996) across a broad
22 spectrum of metabolic risk with complementary non-invasive measures of hepatic steatosis to
23 identify a circulating proteome of MASLD. Single-protein and multi-protein signatures of MASLD
24 were associated with imaging-defined MASLD across a spectrum of metabolic risk, with clinical
25 diagnostic performance beyond known MASLD risk factors. Identified proteins specified
26 pathways of central carbon and amino acid metabolism, hepatocyte regeneration, inflammation,

1 fibrosis, and insulin sensitivity. Moreover, this multivariable signature of MASLD was associated
2 with incident non-alcoholic chronic liver disease, diabetes, and a host of metabolic conditions in
3 >26000 participants over nearly a decade of follow-up in the UK Biobank, with additivity over
4 polygenic risk estimates and prognostic reliability down to a clinically translatable 21-protein
5 panel. We observed very strong enrichment of the MASLD-associated proteome in human liver
6 at a transcriptional level, as well as spatial transcriptional localization to areas of histologically
7 defined steatosis in human liver in early MASLD. Expression of key targets appeared
8 predominantly in hepatocytes, macrophages, and fibroblasts and several targets' (prioritized via
9 spatial transcription and proteomics) expression correlated with progression of steatosis in >600
10 human livers and in a humanized "liver-on-a-chip" model of early MASLD. Collectively, these
11 results inform fundamental connections between a blood-based proteomic signature of MASLD
12 as a "liquid biopsy" of potentially dynamic states in human liver, connecting clinical biomarker
13 discovery and hepatic biology in MASLD.

14

15 **RESULTS**

16 **General flow of study and characteristics of study samples**

17 Our study consisted of five integrated steps (**Figure 1**; details in **Methods**), specifically
18 (1) identification and validation of a "MASLD proteome" (across 4996 participants across three
19 prospective observational studies: Coronary Artery Risk Development in Young Adults
20 [CARDIA], UK Biobank, Cameron County Hispanic Cohort [CCHC]); (2) characterization of
21 tissue origin and implicated molecular pathways; (3) relation of this proteome to MASLD-
22 relevant clinical outcomes, including complementarity with human genetics (across 26421
23 participants in UK Biobank over a median 13.7 years of follow-up); (4) examining expression of
24 genes encoding the MASLD proteome in human liver across stages of MASLD (bulk RNA-seq
25 in human liver; SteatoSITE, N=523 biopsies); (5) specifying cell and spatially resolved

1 expression of these genes in human liver with and without MASLD (scRNA-seq, N=19; spatial
2 transcriptomics, N=4).

3 The characteristics of the study samples are shown in **Supplementary Tables 1-3**.

4 Overall, our study included 2679 CARDIA study participants after excluding participants with
5 other potential causes for hepatic steatosis (>14 alcoholic drinks/week, hepatitis C, cirrhosis,
6 HIV, and use of amiodarone, valproic acid, methotrexate, tamoxifen, or diltiazem)¹⁰. CARDIA
7 participants were randomly split into derivation (N=1876) and validation (N=803) subsamples,
8 with an overall median age of 51 years, 57% female, and 47% Black individuals (**Supplemental**
9 **Table 1**). Participants were predominantly overweight and obese (median body mass index
10 [BMI] 29 kg/m²), with low alcohol use (median 1 drink/week) and 15% had diabetes. The study
11 population from UK Biobank included 26421 participants that represented a broader age range
12 (25th to 75th percentile: 50-64 years), 54% were women, and participants were predominantly
13 White (94%) and overweight (median 27kg/m²) with a lower prevalence of diabetes (5.8%) and
14 greater alcohol intake (43% reporting ≥ 3 times a week). A total of 2111 UK Biobank participants
15 had MRI measures of hepatic steatosis (**Supplemental Table 2**). Participants from the
16 Cameron County Hispanic cohort (CCHC; N=206) had similar age (25th to 75th percentile: 46-66
17 years) and sex (66% female) distributions, composed entirely of White Hispanics, with greater
18 obesity (median BMI 31 kg/m²) and prevalence of diabetes (32%), with 49% reporting no alcohol
19 use (**Supplemental Table 3**).

20

21 **Circulating proteomics of MASLD identify broad canonical pathways of human** 22 **metabolism with predominant gene expression in the liver**

23 Across 2679 participants in CARDIA with SomaScan 7k proteomics, we identified 237
24 unique proteins (259 SomaScan aptamers) associated with liver attenuation on computed
25 tomography (lower liver attenuation ~ more hepatic steatosis) across both derivation and
26 validation subsamples (adjusted for age, sex, race, BMI; **Figure 2A-B**; full regression estimates

1 in **Supplementary Data File SD02**). Regression estimates were robust to multivariable
2 adjustment, including metabolic risk factors, renal function, physical activity, and alcoholic drinks
3 per week (relation of regression estimates across adjustments: Spearman $\rho=0.95$; $P < 2.2 \times 10^{-16}$;
4 **Supplemental Figure 1**). We observed significant enrichment of genes encoding the MASLD
5 proteome in the liver (**Figure 2C**), specifying broad pathways implicated in central metabolic
6 processes (e.g., carbon, pyruvate, amino acid, carbohydrate metabolism) and fibrosis (**Figure**
7 **2D**), including known and emerging mechanisms of MASLD, namely amino acid metabolism
8 (*ACY1*^{6,11}, *FAH*¹²), alcohol processing (e.g., *ADH1A*^{13,14}), fructose catabolism (*ALDOB*,
9 *SORD*¹⁵), bile acid and steroid metabolism (*AKR1D1*¹⁶, *AKR1C4*^{17,18}), gluconeogenesis
10 (*FBP1*¹⁹), and multi-substrate detoxification, intermediary metabolism, and fibrosis (*GSTA1*²⁰,
11 *ASL*²¹, *UGDH*²²), among others. To identify potential central mediators of MASLD, we next
12 conducted an interaction (hub gene) analysis including 235 genes (of the 237 unique genes),
13 with nodes identifying genes with central relevance to MASLD (**Figure 2E**). Identified nodes
14 included pathogenic mediators of hepatocyte regeneration and fibrosis regulation (*EGFR*²³, *IGF-*
15 *1*²⁴), apoptosis regulation (*MET*²⁵), inflammatory mediators (*CXCL2*²⁶, *CRP*, *SERPINE1*),
16 extracellular matrix responses to hepatic injury (*VTN*²⁷, *ACAN*²⁸), glycogen metabolism
17 (*PYGL*²⁹), and mitochondrial pyruvate metabolism (*PKLR*³⁰, *PC*³¹), among several other
18 canonical markers of insulin sensitivity and adiposity (*ADIPOQ*, *INS*). These results suggested a
19 predominant hepatic origin for the circulating MASLD proteome, implicating canonical
20 metabolic-inflammatory-fibrotic pathways in liver degeneration.

21 22 **Integrated proteomics identifies a diagnostic biomarker of MASLD with wide replication** 23 **across metabolic risk states**

24 Penalized regression (LASSO) generated a 336-apramer model for liver attenuation
25 adjusted for age, sex, race, and BMI (hereafter referred to as “MASLD score”). The MASLD
26 score correlated with liver attenuation in both derivation and validation subsamples within

1 CARDIA (Spearman $\rho=0.69$ and 0.56 , respectively; **Figure 3A**; model coefficients in
2 **Supplementary Data File SD04**), differed across clinical thresholds for MASLD (**Figures 3B**,
3 **3D**), with retention of correlations using the top 21 aptamers from the MASLD score
4 (**Supplemental Figure 3A**). Addition of the MASLD score to standard clinical risk (age, sex,
5 race, BMI, alcoholic drinks per week, aspartate aminotransferase [AST; aptamer based], alanine
6 aminotransferase [ALT; aptamer based], hemoglobin A1c) markedly improved MASLD
7 discrimination (C-index 0.84 [95% CI $0.80-0.88$] to 0.94 [95% CI $0.92-0.96$], $P=4.3 \times 10^{-7}$; **Figure**
8 **3F**). This discriminative performance was maintained using the top 21 aptamers from the
9 MASLD score (21 chosen given the limit of current absolute proteomic detection [Olink] for
10 clinical translation). Of note, despite the well described association between obesity and
11 MASLD, we only observed a modest correlation between BMI and the protein score (Spearman
12 $\rho=-0.27$, $P<2.2 \times 10^{-16}$), with smaller effects by age, sex, race, and alcohol use (**Supplemental**
13 **Figure 2**).

14 For external rigor of our approach, we next replicated the MASLD score with hepatic
15 steatosis in >2000 participants from 2 different studies with distinct approaches to MASLD
16 quantification (CCHC, ultrasound; MRI, UK Biobank). As CCHC and UK Biobank used Olink
17 proteomics platforms, recalibration of the MASLD score was performed (see **Methods**). Of note,
18 CT liver attenuation (the measure in CARDIA) is an opposite directionality relative to ultrasound
19 or MRI (higher attenuation ~ lower steatosis ~ lower MRI or ultrasound measure). In 2111 UK
20 Biobank participants, our results were largely similar, including (1) a similar relation between the
21 recalibrated MASLD score and MRI-determined proton-derived fat fraction (Spearman $\rho=-0.5$,
22 $P<5.9 \times 10^{-135}$); **Figure 3C, Supplemental Figure 3B**); (2) improvement in MASLD discrimination
23 above clinical models (defined as PDFFF>5.5%, a clinically accepted threshold³²; C-statistic 0.79
24 [$0.77-0.82$] to 0.83 [$0.81-0.85$], $P=5.2 \times 10^{-8}$, **Figure 3G**); (3) similar associations with age, sex,
25 race, BMI, and alcohol use (**Supplemental Figure 2**). Similarly, in CCHC—despite at far higher

1 metabolic risk with far more prevalent MASLD (**Supplemental Table 3**)—we found a magnitude
2 of correlation consistent with other cohorts for controlled attenuation parameter (an ultrasound-
3 based measure of steatosis; Spearman $\rho=-0.54$, $P<2.2\times 10^{-16}$; **Figure 3E**).

4 Recognizing the importance of adiposity on the risk of MASLD, we interrogated whether
5 the relationship between the MASLD score and hepatic steatosis was modified by
6 overweight/obesity status. In the CARDIA validation subsample (N=803), we observed a
7 stronger correlation among participants with obesity (BMI ≥ 30 kg/m²; Spearman $\rho=0.65$) than
8 participants with normal/overweight BMI (<30 kg/m²; Spearman $\rho=0.41$; **Supplemental Figure**
9 **4**). In a linear model for the outcome of hepatic steatosis (which was CT assessed liver
10 attenuation in CARDIA), we observed a statistically significant interaction between the MASLD
11 score and BMI wherein the magnitude of the relationship between the MASLD score and CT
12 liver attenuation increases with greater BMI (interaction $\beta=7.5$, $P=1.4\times 10^{-8}$). In conjunction with
13 the observed relationship in CCHC (a cohort with elevated metabolic risk), these findings
14 support the relevance of the MASLD score in at-risk populations.

15

16 **A MASLD proteomic score identifies individuals at long-term chronic non-alcoholic liver** 17 **disease risk, MASLD-related diseases, and survival in over 26,000 people, additive to** 18 **polygenic risk**

19 In 26421 participants in UK Biobank (median follow-up for mortality 13.7 years, 25th-75th
20 percentile 13.0-14.5 years), we found a broad relation of the MASLD score with MASLD and
21 MASLD-relevant clinical outcomes (**Figure 4A**). In addition to all-cause and cause-specific
22 mortality (specifically cancer), we observed very strong relations between the MASLD score and
23 chronic non-alcoholic liver diseases (an electronic health record surrogate of MASLD; adjusted
24 HR 0.59, 95% 0.52-0.67; $P=7.0\times 10^{-16}$) and diabetes (adjusted HR 0.50, 95% 0.47-0.54;
25 $P=1.7\times 10^{-86}$; adjustments in **Methods**). In sensitivity analyses, these associations were robust to

1 additional adjustment for AST, ALT, and hemoglobin A1c (chronic non-alcoholic liver disease
2 HR 0.63, diabetes HR 0.53; **Supplemental Data File SD06**) and were retained down to a 21-
3 protein Olink panel (chronic nonalcoholic liver disease HR 0.64, diabetes HR 0.57; **Figure 4C**).
4 In addition to association with clinical outcomes, addition of the MASLD score provided marked
5 improvements in discrimination and net reclassification above clinical models (including age,
6 sex, race, BMI, systolic blood pressure, diabetes [removed for models of diabetes], Townsend
7 Deprivation Index, smoking, alcohol use, and low-density lipoprotein) for both diabetes (C-
8 statistic 0.79 vs. 0.84, $P=6.6 \times 10^{-20}$) and chronic non-alcoholic liver disease (C-statistic 0.69 vs.
9 0.74, $P=1.1 \times 10^{-5}$; **Figure 4A**). In a sensitivity analysis for incident chronic non-alcoholic liver
10 disease we further adjusted for AST, ALT and A1c and still observed a significant improvement
11 in model performance (C-statistic 0.72 vs 0.74, $P=0.009$). Given the widespread use of
12 polygenic liability of diabetes for clinical risk prediction³³, we tested the relative effect of the
13 MASLD score and a diabetes polygenic risk score (PRS), demonstrating a largely additive effect
14 with minimal interaction (weak PRS-by-proteomic interaction beta = 0.07, $P=0.001$; model
15 predicted hazard ratios presented in **Figure 4B**). These results suggested a strong proteomic
16 liability to incident MASLD and MASLD-related disorders, with significant addition to polygenic
17 risk estimates for risk stratification.

18

19 **The MASLD proteome exhibits directionally consistent, cell-specific, and spatially** 20 **localized expression in areas of steatosis in early MASLD with dynamicity during MASLD** 21 **progression**

22 To elucidate the spatial and cellular organization of prioritized targets identified in
23 proteomic analyses, we conducted a comprehensive mapping of proteins implicated in
24 regression in CARDIA to human liver tissue in early MASLD. Leveraging published single-cell
25 and single-nuclear RNA sequencing (scRNA-seq) and spatial transcriptomics³⁴ (N=21; 62%
26 women, mean age 59 years, mean BMI 32 kg/m²), we investigated the expression profiles of

1 198 MASLD-associated protein-gene overlaps represented in both the scRNA-seq and spatial
2 transcriptomics datasets using individual gene expression and a composite expression score
3 (**Figure 5**, see **Methods**) to facilitate the broad identification of cell types and spatial distribution
4 of targets prioritized by the MASLD proteome within the human liver. This analysis showed a
5 predominant cell-specific expression pattern of implicated targets, primarily observed in
6 hepatocytes. However, a subset of targets implicated by the MASLD proteome show
7 heterogenous expression patterns across cell, including fibroblasts, cholangiocytes, endothelial
8 cells, and immune cells (**Figure 5A** and **Supplemental Figure 5A**). Employing a composite
9 expression score, we showed higher gene activity of implicated targets within steatotic tissue
10 and in the mid-central liver zonation. These zones were previously shown to correspond with
11 higher hepatocyte expression signature using the same dataset (**Figure 5B-F**, **Supplemental**
12 **Figure 5B**)³⁴.

13 We subsequently assessed whether implicated targets identified by the CARDIA model
14 exhibit differential expression in single-cell data. To account for the confounding effects of cell
15 isolation technique on the liver cell atlas cell types, we specifically utilized the spatial
16 transcriptomics dataset for this analysis. Of the 198 genes encoding implicated proteins
17 expressed in the spatial transcriptomics Visium dataset, 33 were differentially expressed
18 between healthy and steatotic tissue (minimum expression > 10% of spots; adjusted $P < 0.05$;
19 $|\log_2 \text{fold-change}| > 0.25$ & $|\text{minimum \% expressed spot difference}| > 10\%$; **Figure 5H**). Of the
20 33 genes, 30 were upregulated in steatotic tissue and only 3 genes (*IGFBP2*, *IL1RAP* and
21 *SHBG*) were downregulated. These genes exhibited distinct expression patterns in human liver
22 tissue with and without steatosis (refer to **Figure 5G-H**, select targets shown in **Supplemental**
23 **Figure 5C-D**). Moreover, we observed high biological concordance between their change in
24 expression and the clinical effect estimate of the circulating protein corresponding to this gene
25 with liver attenuation in CARDIA (**Figure 5G**). One notable example with high effect size is
26 *IGFBP2*—dynamic during metabolic intervention³⁵ and regeneration³⁶ with liver-enriched

1 expression³⁷—which was increased in healthy versus steatotic cell populations (transcriptional
2 level, **Supplemental Figure 5D**) and associated with lesser steatosis in CARDIA (population
3 proteomic level), consistent with smaller reports³⁸. Most observed population association-tissue
4 concordance consisted of metabolic genes up-regulated in steatotic liver that displayed greater
5 circulating protein abundance in individuals with steatosis (**Figure 5G-H**), many of which had
6 established mechanistic relevance in model systems of MASLD and its progression (e.g., *ENO3*
7 and ferroptosis³⁹; *UGDH* and fibrosis/redox status²²; *CTS2* and epithelial-mesenchymal
8 transition⁴⁰; *CDH1* and lipogenesis⁴¹; *CDH1* and *PPAR/PGC1a* signaling⁴²; among others).

9 Next, we matched these 33 differentially expressed genes to bulk transcriptomic data
10 across NASH-CRN defined stages of hepatic steatosis to investigate potential dynamicity
11 across individuals with increasing severity of histopathologic phenotype (**Figure 6**). Of 523
12 SteatoSITE participants with biopsy samples, there were 489 with MASLD (54% women, mean
13 age 52 years, mean BMI 32 kg/m², 75% diabetes) compared to 34 control samples. Of the 33
14 genes passed forward for assessment in bulk transcriptomics, 12 were not significantly
15 expressed in any of the stages of steatosis (by adjusted p-value < 0.05) and were not included
16 in visualization. We observed several genes with high effect size differences by steatosis grade,
17 concordant with circulating proteomic and spatial relations (e.g., *IGFBP2*, *IL1RAP*, *SHBG*,
18 *ENO3*, *DEGB1*, *ME1*). Across the genes prioritized by proteomic and spatial studies, we
19 observed two types of discordant findings: (1) genes with a directionality consistent with our
20 proteomic-spatial (but not bulk) results (e.g., *SERPINE1/PAI-1*⁴³⁻⁴⁵, *HSPA1B*⁴⁶); (2) genes
21 consistent with the bulk (but not proteomic-spatial) directionality (e.g., *PSAT1*⁴⁷, *UDGH*²²,
22 *ACO1*⁴⁸). Several factors—technical (sequencing methodologies, bulk versus single cell, limited
23 sample size in this previously published scRNAseq dataset³⁴, participant-level, and biological
24 (steatosis as one component of the MASLD phenotype, in addition to inflammation, ballooning,
25 fibrosis)—may account for these differences. Nevertheless, these findings collectively highlight

1 the potential for proteo-transcriptional target mapping for human MASLD amidst context-
2 dependent heterogeneity and complexity of integrating multiple approaches.

3

4 **Humanized liver-on-a-chip recapitulates an early phase human MASLD phenotype** 5 **consistent with proteo-transcriptional MASLD targets from human studies**

6 The experimental design for the liver-on-a-chip (LOC) studies is shown in **Figures 7A-B**.
7 Microscopy, immunofluorescence, and gene expression studies after administration of fatty
8 acids (oleic and palmitic) known to promote a steatosis phenotype⁴⁹ were consistent with
9 morphologic and transcriptomic induction of a MASLD phenotype (decreased *IRS1*^{50,51}, *IRS2*⁵¹,
10 *PPARa*^{52,53}; increased *SREBP1c*⁵⁴, *PPARg*⁵⁵, *FABP4*⁵⁶; **Figure 7C-D**). Given limited cDNA yield
11 from the LOC experiments, we prioritized 13 of 33 targets identified across proteomic and
12 spatial transcriptional studies (**Supplemental Data File SD07**) for assessment on the LOC in
13 two ways: (1) top 5 (*HMGCS1*, *SERPINE1*, *HSPA1B*, *ENO3*, *HSPA1A*) and bottom 5 (*CDA*,
14 *PYGL*, *IL1RAP*, *SHBG*, *IGFBP2*) differentially expressed targets, ranked by log₂-fold difference
15 in steatotic and non-steatotic livers by spatial transcriptomics (include all 3 of which were
16 downregulated in steatotic livers); (2) three additional targets differentially expressed in spatial
17 human liver studies but with prominent expression in non-parenchymal (non-hepatocyte; NPCs)
18 cells (*ME1*, *CTSZ*, *DEFB1*).

19 We observed broadly directionally consistent results between circulating proteomic,
20 tissue transcriptional, and LOC experiments, with increased expression of genes implicated in
21 hepatic lipid metabolism, stress, and non-canonical pathways (e.g., ferroptosis) across
22 hepatocytes and NPCs (*PYGL*, *HMGCS1*, *SERPINE1*, *ENO3*, *HSPA1B*; **Figure 7E**).
23 Furthermore, while *ME1*, *CSTZ*, and *DEFB1* were not expressed in LOC hepatocytes
24 (consistent with spatial human studies), the expression of these genes was increased in NPCs
25 in the LOC (**Figure 7E**). Taken together with bulk data for *DEFB1* and *ME1* in bulk
26 transcriptional data (**Figure 6**), these results suggest the increased expression of these genes

1 are predominantly driven by cells of non-hepatocyte origin. Of note, several genes did not
2 exhibit the expected directionality from proteomic, spatial, or bulk studies (non-significant:
3 *IGFBP2*, *SHBG*; opposite directionality, *IL1RAP*, *CDA*; **Figure 7E**), potentially owing to
4 biological heterogeneity between model systems.

6 **DISCUSSION**

7 Given the growing worldwide prevalence of MASLD and its broad impact on human
8 health, there is increasing interest to identify actionable biomarkers that reflect clinical risk and
9 mechanism underlying MASLD pathobiology. We identified MASLD-associated proteins across
10 a broad proteomic space (7230 aptamers) in \approx 5000 individuals from two large prevention
11 populations at dramatically lower metabolic risk relative to recent reports^{6,8,57} to address clinical
12 generalizability. Using both single and multivariable regression modeling, the MASLD proteome
13 prioritized broad metabolic, inflammatory, and fibrosis pathways, with multivariable instruments
14 offering significant augmentation in MASLD discrimination beyond clinical factors in both
15 CARDIA and UK Biobank (despite heterogeneity in the assessments of proteins, hepatic
16 steatosis, and social-environmental measures). Our approach was strengthened by using direct
17 measures of hepatic steatosis (CT, ultrasound, and MRI), as opposed to the use of diagnosis
18 codes⁵⁸, to define proteomic relations to MASLD, resulting in multivariable MASLD proteomic
19 instruments associated with hepatic steatosis in populations with elevated metabolic risk
20 (CCHC). The range in C-statistic for composite proteomic-phenotype risk models in both cohorts
21 (CARDIA: 0.94; UK Biobank: 0.83) were consistent with prior reports in markedly higher risk
22 individuals from Wood and colleagues (validation N=134 adults; BMI 49 ± 8 kg/m², 41% diabetes,
23 84% female; proteomic C-statistic: 0.864)⁵⁷ and Govaere and colleagues (validation N=115
24 adults; BMI 32 ± 6 kg/m², 52% diabetes, 44% female; proteomic C-statistic: 0.8)⁸. The
25 multivariable MASLD proteome forecasted prospective disease risk in >26000 UK Biobank
26 participants for incident chronic non-alcoholic liver disease, diabetes, and other outcomes linked

1 to MASLD (ischemic heart disease, cancer mortality, and all-cause mortality) and mitigated high
2 polygenic liability for diabetes in UK Biobank. These results extend reports in high-risk,
3 established disease populations⁶ into a prevention context with high clinical impact of
4 intervention, underscoring the translatability of this approach for MASLD- and MASLD-related
5 disease personalization.

6 While liver imaging enjoys discriminative accuracy for MASLD (e.g., CT, MRI)^{59,60}, these
7 studies have some limitations, including clinical availability and cost in widespread screening
8 efforts and sensitivity for early disease where intervention may have the greatest clinical impact.
9 Consequently, the clinical field in MASLD has directed focus toward identifying rapidly
10 measured biosignatures of dynamic hepatic states that are prognostically, diagnostically, and
11 biologically relevant. Candidate biomarkers have included human genetics^{58,61-66}, gene and
12 protein expression^{6,58,67,68}, and metabolism. Nevertheless, while key genetic susceptibility loci
13 critical to MASLD biology have been identified in large biobanks (e.g., *PNPLA3*), polygenic
14 liability for MASLD in a population context is limited⁶⁹ by more prominent non-genetic influences
15 (e.g., diet, alcohol or drug exposure, obesity, inactivity, environmental exposures). In non-
16 genomic studies across a broader array of biomarkers, sample sizes have generally been small
17 with population bias (e.g., measured in established risk, like obesity⁶¹). Of note, recent
18 innovative efforts to map a circulating snapshot of metabolic biology (via the human proteome)
19 into hepatic transcriptional states has been successful, albeit in a small sample with high
20 metabolic disease prevalence⁸. Indeed, the value of transcriptional indexing of the human
21 proteome across broad at-risk populations has recently been highlighted⁷⁰ (though not yet in
22 MASLD).

23 Beyond disease personalization, our central hypothesis included mapping of the
24 circulating human proteome into tissue MASLD to identify relevant, dynamic markers of interest.
25 We observed a very high enrichment of transcript expression of genes encoding the MASLD
26 proteome in the liver (far beyond any other tissue), supportive of its hepatic origin. Moreover,

1 identified protein targets specified broad pathways central to MASLD, including regulation of
2 hepatocyte regeneration (*EGFR*²³), injury, apoptosis (*MET*²⁵), inflammation (*CXCL2*²⁶, *CRP*,
3 *SERPINE1*), metabolism (*ACY1*^{6,11}, *FAH*¹², *ADH1A*^{13,14}, *ALDOB*, *SORD*¹⁵, *AKR1D1*¹⁶,
4 *AKR1C4*^{17,18}), and fibrosis (*IGF-1*²⁴). Given that these tissue references are from “normal” banks
5 in bulk resolution (e.g., Human Protein Atlas/GTEX), we further explored our MASLD-associated
6 proteomic targets at single nuclear and spatial resolution in human liver at an early MASLD
7 stage³⁴. A prime finding from this approach was the striking concordance of the circulating
8 proteomic effect size in our large, at-risk population (CARDIA) and the fold-difference between
9 healthy and fatty liver. Plasma proteins that were more abundant in patients with lower degree
10 of hepatic steatosis corresponded to mRNAs that were also higher in expression in non-
11 steatotic livers (and vice versa). Gene activity (as defined by a gene expression score across
12 198 MASLD-associated proteins) mapped primarily to histopathologic areas of steatosis, with a
13 predominant hepatocyte expression and zonation pattern. Interestingly, we also observed a
14 greater expression of *CTSZ* in macrophages, and migratory and conventional dendritic cells,
15 which is in line with recent reports linking inflammatory cells to MASLD pathogenesis in a
16 murine model⁷¹. Several genes differentially expressed in spatial transcription were dynamic
17 across MASLD stages with consistent directionality with spatial transcriptomic and circulating
18 proteomics⁷, further suggesting validity (and prioritization) of these targets.

19 Our studies were augmented by the development of a MASLD model in a humanized
20 LOC platform to determine direct causality between induction of steatosis and changes in
21 mRNA transcripts identified in the human transcriptional studies. While the LOC model used in
22 this study has been validated for recapitulating key aspects of human liver physiology⁷²⁻⁷⁴, it has
23 mostly been used as a drug screen for hepatotoxicity⁷⁵⁻⁷⁸. While the use of such platforms to
24 model steatosis is emerging⁷⁹, we successfully developed a model that includes both
25 hepatocytes and non-hepatocyte cells (Kupffer cells, stellate cells, and endothelial cells)
26 subjected to treatment with a cocktail of fatty acids. Importantly, while admittedly far less

1 complex than human MASLD, the model recapitulates key histological and transcriptional
2 features of human MASLD, allowing us to query for direct changes (in response to steatosis
3 induction) in mRNA transcripts in both hepatocyte and non-hepatocyte cell types derived from
4 our human studies. These data not only validate a primary role for intrahepatic steatosis in the
5 transcriptional changes, but also provide a possible platform to test novel therapeutic targets
6 (that may reverse these changes) in the future.

7 Overall, the shared findings across biologically plausible pathways relevant to MASLD
8 pathogenesis—and the association of these mediators in circulation with long-term outcomes—
9 supports a broad validity from human populations to individual cells. Several caveats merit
10 mention. We recognize the diagnosis of MASLD requires additional metabolic dysfunction
11 (independent of other causes, e.g., viral, alcohol) associated with steatosis, and different
12 cohorts had different distributions of these influences. Nevertheless, our CARDIA derivation
13 necessarily excluded individuals with viral or drug/toxin-induced liver disease as possible in an
14 epidemiologic setting and generated proteomic signatures associated with hepatic steatosis
15 across BMI (a key determinant of metabolic dysfunction). Broad validation across cohorts
16 (despite heterogeneity in hepatic steatosis ascertainment) is further evidence of external
17 validity. The use of the transcriptome for tissue deconvolution of the human proteome is an
18 emerging concept^{8,80}, traditionally countered by concerns over protein-to-transcript concordance
19 (including epigenetic effects that may disconnect them) and tissue ubiquity of the transcripts
20 corresponding to the circulating proteome. Nevertheless, recent work in high metabolic risk
21 individuals has suggested a largely positive (though variable) correlation between tissue RNA
22 and circulating protein levels for those targets associated with MASLD stage⁸, consistent with
23 our findings of a high liver transcriptional enrichment for prioritized proteins. Consistent cross-
24 platform imaging and prognostic findings (e.g., aptamer- to antibody-based proteomics) lend
25 validity. While the relation between the MASLD score and MRI based hepatic steatosis in UK
26 Biobank were temporally separated, the association was similar as in CARDIA. We are limited

1 in not having conducted spatial transcriptomics across variable disease states over time or with
2 therapy (e.g., weight reduction), as well as gain- and loss-of-function studies for targets that
3 survived our tiered approach. While this was not the scope of the current work, we envision that
4 the approach here can be extended to a serial context to examine tissue-circulating
5 concordance to further hone targets for mechanistic study. Finally, we noted some discordances
6 between proteomic, spatial, bulk, and LOC data. Beyond technical considerations (e.g., bulk vs.
7 spatial transcriptional approaches), we hypothesize variability in results may owe to population
8 and disease heterogeneity across cohorts, including MASLD phenotypic heterogeneity
9 (ballooning, inflammation, fibrosis, steatosis that can co-existent to different extent).
10 Nevertheless, targets that filter across all approaches are more likely to be involved in human
11 liver disease, and larger, integrative approaches will continue to add to this science.

12 In conclusion, across \approx 5000 participants with clinical, imaging, and biochemical data, we
13 define a proteomic architecture of MASLD with replication and diagnostic stratification across a
14 MASLD spectrum (from early- to high-risk metabolic cohorts) and strong association with
15 MASLD-related disease (beyond modern human genetic approaches) in >26,000 individuals.
16 Proteins implicated by these population-based approaches were highly enriched at a
17 transcriptional level in human liver and specified canonical and novel pathways of MASLD
18 progression. We observed spatially enriched activity of these genes in areas of steatosis and by
19 liver zonation, with concordance between the circulating proteomic effects on liver fat and the
20 fold differences between healthy and fatty liver by spatial transcription. Several targets
21 additionally demonstrated concordant changes during evolution of MASLD across histologically
22 defined stages and within a humanized “liver-on-a-chip” model system. These results
23 contextualize the promise of multi-level discovery—across broad clinical populations, proteome,
24 and tissue studies—to discern biologically relevant, spatially enriched targets in MASLD for
25 downstream mechanistic, diagnostic, and prognostic work.

26

1 METHODS

2 **Data Availability:** Data for reproduction of this analysis may be obtained from the
3 CARDIA Coordinating Center (www.cardia.dopm.uab.edu), CCHC Coordinating Center
4 (<https://sph.uth.edu/research/centers/hispanic-health/>), SteatoSITE (<https://steatosite.com>),
5 Liver Cell Atlas (<https://livercellatlas.org/index.php>), and UK Biobank
6 (<https://www.ukbiobank.ac.uk>). Analyses in the UK Biobank were performed under proposal
7 number 57492.

9 Study samples

10 The study involved multiple samples: (1) the Coronary Artery Risk Development in
11 Young Adults study (CARDIA, N=2679; proteomic discovery and validation of MASLD-related
12 proteins; characteristics in **Supplementary Table 1**, study design in reference); (2) the UK
13 Biobank study (N=26421; second validation of MASLD-related proteins; assessment of clinical
14 prognostic value against incident MASLD-related diseases; characteristics in **Supplementary**
15 **Table 2**); (3) Cameron County Hispanic Cohort (CCHC; N=206 with ultrasound-based measures
16 of liver structure and circulating proteomics; characteristics in **Supplementary Table 3**) (4)
17 published bulk RNA sequencing study (SteatoSITE, N=618 liver biopsies across stages of
18 MASLD)⁷; (5) spatial transcriptomic study in human liver (N=5 liver biopsies for spatial
19 transcriptomics and N=19 for single nuclear RNA-sequencing)³⁴. All study participants provided
20 written and informed consent, and all study protocols were approved by the Institutional Review
21 Boards of the respective studies.

22 **CARDIA:** The Coronary Artery Risk Development in Young Adults (CARDIA) study
23 started recruitment in 1985-1986 across 4 cities in the U.S. (Birmingham, AL; Chicago, IL;
24 Minneapolis, MN; and Oakland, CA) to study coronary risk factor development longitudinally
25 beginning in young adulthood⁸¹⁻⁸⁴. Our study used data from the Year 25 exam where 2977
26 participants had proteomics quantified. We excluded 275 participants with other potential

1 causes for hepatic steatosis (>14 alcoholic drinks/week, hepatitis C, cirrhosis, HIV, and use of
2 amiodarone, valproic acid, methotrexate, tamoxifen, or diltiazem)¹⁰. We excluded 11 participants
3 missing hepatic steatosis measurements, and 12 participants for missing data on BMI or
4 drinks/week. CARDIA participants were randomly split into derivation (70%) and validation
5 (30%) samples, balanced by computed tomography-based measurement of hepatic steatosis.

6 **UK Biobank:** The UK Biobank is a population-based study of >500000 participants who
7 were aged 40-69 when recruited between 2006-2010 across the United Kingdom. Proteomics
8 data from the initial assessment (instance 0) using the Olink Explore panel is available on
9 ≈54000 UK Biobank participants⁸⁰. We included 26429 participants with complete data for the
10 proteins used to calculate a proteomic score of hepatic steatosis, of which 8 participants were
11 excluded from analyses for having a proteomic score >5 SDs away from the mean. A subset of
12 2111 had hepatic steatosis quantified by MRI at the imaging visit (2014 and later; instance 2)

13 **Cameron County Hispanic Cohort (CCHC):** The CCHC is a community-based
14 prospective observational cohort study of 5122 individuals (age 8-90) from a low-income
15 Hispanic/Latino population at the Texas/Mexico border. The study design has been previously
16 described⁸⁵. We included 206 individuals who had abdominal ultrasound to measure controlled
17 attenuation parameter (CAP), a quantitative measure of hepatic steatosis⁸⁶, and simultaneous
18 circulating proteomics.

19

20 **Hepatic steatosis assessment**

21 In CARDIA, hepatic steatosis was measured as liver attenuation on computed
22 tomography as previously described, where lower levels of liver attenuation are associated with
23 greater steatosis¹⁰. MASLD was defined as liver attenuation <40 Hounsfield units. In the UK
24 Biobank, hepatic steatosis was measured by magnetic resonance imaging in a subset of
25 participants at instance 2 using the iterative decomposition of water and fat with echo
26 asymmetric and least-squares estimation (IDEAL) protocol, as previously described⁸⁷. MASLD

1 was defined as a proton density fat fraction $>5.5\%$ ³². In CCHC, vibration-controlled transient
2 elastography was used to measure CAP (FibroScan 502 Touch or FibroScan 530 Compact,
3 Echosens; automatic probe selection) for 10 valid measures with the median used in analysis,
4 as described⁸⁸.

5

6 **Proteomics**

7 **CARDIA:** Quantification of the circulating proteome was performed using aptamer-based
8 technology (Somalogic, Boulder, CO) which measured 7596 aptamers. Sixty-eight participants
9 had >1 samples measured and we averaged their proteomic data for analysis. We excluded
10 non-human proteins (N=72) and proteins with a coefficient of variation $>20\%$ (N=58). We tested
11 for batch effect and participant outliers using principal component analysis and identified
12 neither. Proteins were log-transformed and standardized (mean 0, variance 1) prior to use in
13 models.

14 **UK Biobank:** Recently released proteomic data from the Olink Explore platform (Olink,
15 Uppsala, Sweden) measured from the instance 0 visit were used in this study⁸⁰. Of the 1463
16 proteins measured, we excluded 130 proteins where $>40\%$ of reported measurements were
17 below the limit of detection and another 3 proteins where $>20\%$ of reported measurements were
18 missing. Proteins were standardized (mean 0, variance 1) prior to use in models.

19 **CCHC:** We performed proteomics in CCHC participants using the Olink Explore 1536
20 platform. Proteins were standardized (mean 0, variance 1) prior to use in models.

21

22 **Spatial, single nuclear, and bulk transcriptomics in human liver**

23 **Single nuclear and spatial transcriptomics:** To assess cell-specific and spatial
24 expression patterns of implicated protein targets, we harnessed integrated single cell and single
25 nuclear RNA-sequencing (scRNA seq, total N= 19; fatty = 7; non-fatty = 11; unknown = 1) and
26 Visium spatial transcriptomics data (total N= 4; fatty = 2; non-fatty = 2) previously published

1 from our collaborative group³⁴. Expression patterns of implicated proteins were assessed by
2 mapping significant model proteins to their corresponding gene symbol that were expressed in
3 both the scRNA-seq and Visium data, resulting in total of 198 genes represented across the
4 three datasets. Activity of these genes were then measured for their activity using single gene
5 expression measures as wells an expression composite score that represent the transcriptional
6 signature of all model genes in each individual cell (snRNAseq) or spot (Visium data).
7 Expression composite score was generated using the *AddModuleScore* function (implemented
8 in Seurat v5). To identify differential expression of nominated targets in the liver we compared
9 healthy samples to early steatotic samples using Visium data where both healthy and early
10 steatotic samples were available. Differentially expressed genes were assessed using negative
11 binomial model implemented in the *FindMarkers* function (Seurat). Only target genes
12 expressed in at least 10% of the spots were included in the analysis (198 genes) for differential
13 expression. We defined differential expression as adjusted p-value < 0.05 and $|\log_2\text{fold change}$
14 $> 0.25|$ and a minimum difference in expressed spots > 10% between fatty and non-fatty. We
15 confirmed the effect size estimates of our differential expression analysis via negative binomial
16 mixed models with sample as random effect (generalized linear mixed models are more
17 sensitive to the dispersion in single-cell data compared to generalized linear models), with high
18 agreement between \log_2 fold change and the negative binomial mixed model coefficients for all
19 198 model targets (Pearson $r = 0.86$) and for the 33 differentially expressed genes (Pearson r
20 $= 0.99$; **Supplemental Figure 6**).

21 **Transcriptional differences across MASLD stages:** We explored the pattern of
22 expression across the 33 genes prioritized by the spatial data analysis above (33 differentially
23 expressed genes between healthy and steatotic tissue out of 198 genes tested, see **Results**) in
24 the SteatoSITE cohort (523 samples; 34 controls, 489 with MASLD) categorized based on
25 NAFLD activity score (NAS) for steatosis (only those samples with scores 1, 2 and 3 were
26 chosen) and compared to control samples⁷. We excluded those individuals with NASH-CRN

1 stage F4 fibrosis, given differences in expression patterns detected in our initial study⁷ and
2 differences in physiology with advanced fibrosis (including paradoxical loss of hepatic fat⁸⁹).
3 Reads were normalised using the weighted trimmed mean of M values method⁹⁰. Differential
4 gene expression analysis was performed using *limma-voom* (v3.28.14) with the protein-coding
5 genes using an FDR of 5% (Benjamini-Hochberg)⁹¹. Of the 33 genes passed forward for
6 assessment in bulk transcriptomics, 12 were not significantly expressed in any of the stages of
7 steatosis (by adjusted p-value < 0.05) and were not included in visualization.

8 **Humanized liver-on-a-chip MASLD model:** The goal of “liver-on-a-chip” technology is
9 to simulate the liver microenvironment which retains key characteristics of native liver function
10 over long-term in culture. The quad culture was set up following the manufacturer’s protocol.
11 The methods below are reproduced from our recent work⁹² directly for rigor and reproducibility,
12 and this citation provides scientific attribution for this. Briefly, by design, each
13 polydimethylsiloxane (PDMS) chip (Chip-S1; Emulate) includes hepatocytes in the apical
14 channel and non-parenchymal cells [NPCs: Kupffer, Stellate and Liver Sinusoidal Endothelial
15 Cells (LSECs)] in the basal channel (**Supplementary Table 4**). These two channels are
16 separated by a porous membrane, coated by hepatic extracellular matrix (ECM). This setting
17 allows the cell-to-cell interaction mimicking the *in vivo* system. The top channel was seeded with
18 hepatocytes at a concentration of 3.5×10^6 cells/mL, followed by overlay with matrigel on the next
19 day. The day after hepatocyte overlay, cell suspensions of three NPCs were mixed in a 1:1:1
20 ratio (v/v/v) to generate the bottom channel tri-cell mixture. The final seeding density of each cell
21 types in the bottom channel were: LSECs: 3×10^6 cells/mL; Stellate cells: 0.1×10^6 cells/mL;
22 Kupffer cells: 0.5×10^6 cells/mL. Chips were maintained for another 96 hours at this condition
23 before treating with fatty acids (FAs). We mimicked an early phase of MASLD by treating both
24 channels of the LOC either with vehicle control (1% BSA) or a combination of FAs (oleic acid
25 $300 \mu\text{M}$: $300 \mu\text{M}$ palmitic acid bound with 1% BSA) for 5 consecutive days.

1 Hepatocytes and NPCs treated with either vehicle control or FAs were imaged directly
2 under brightfield microscope (BioRad). Chips were fixed with 4% paraformaldehyde (4%PFA)
3 followed by permeabilization of both channels with 0.1% Triton® X-100 before staining.
4 Permeabilized cells in both channels were incubated with LipidSpot™ for 10 minutes. The chips
5 were examined under a fluorescence microscope (ECHO Revolve microscope).

6 After 5 days of dosing with FAs, the chips were disconnected, washed with 1X PBS, and
7 filled with RNAlater (Invitrogen) to preserve cells for RNA extraction. The PureLink RNA Mini Kit
8 (Thermo Fischer Scientific) was used following the manufacturer's protocol. Total RNA was
9 eluted in 20µL, treated with DNase, and "cleaned-up" using RNA Clean & Concentrator-5 with
10 DNase I (Zymo Research) following manufacturer's protocol. Final RNA concentration was
11 quantified by spectrophotometry (Nanodrop 2000, Thermo Fischer Scientific). The High-
12 Capacity cDNA Reverse Transcription Kit (Thermo Fischer Scientific) was used for cDNA
13 synthesis from RNA. For amplification and quantification of selected genes (*HMGCS1*,
14 *SERPINE1*, *HSPA1B*, *ENO3*, *HSPA1A*, *PYGL*, *CDA*, *SHBG*, *IL1RAP*, *IGFBP2*, *ME1*, *CTSZ*,
15 *DEFB1*, *IRS1*, *IRS2*, *FABP4*, *SREBP1c*, *PPARα*, *PPARγ*, and β-actin) the ExiLENT SYBR®
16 Green master mix (Exiqon, Vedbæk) was used on a Quant Studio 6 Flex Real-Time PCR
17 System up to 40 amplification cycles. Any amplification cycle (Ct) greater than or equal to 40
18 was assigned as a "negative threshold", which means the corresponding genes were not
19 expressed above the limit of detection of the qRT-PCR assay and therefore those genes were
20 not included in our calculations. For **Fig 7E** absolute gene expression was quantified by $2^{-\Delta Ct}$
21 method after normalization of genes of interest to the internal control β-ACTIN, whereas relative
22 gene expression was used for **Fig 7D**. All qRT-PCR primer sequences are summarized in
23 **Supplementary Table 5**.

24

25 **Statistical methods**

1 **Relating the circulating proteome to hepatic steatosis to identify biological**
2 **pathways of steatosis and development of a diagnostic biomarker panel:** Relations of
3 individual aptamers with hepatic steatosis were examined via regression with aptamers as the
4 predictors adjusted for age, sex, race, and BMI with a false discovery rate of 5% (Benjamini-
5 Hochberg) in the CARDIA cohort using a derivation (70%) and validation (30%) split design
6 balanced on CT liver attenuation. To generate a multivariable protein score of hepatic steatosis
7 (referred to as “MASLD score”), we used least absolute shrinkage and selection operator
8 (LASSO) with non-penalized adjustments for age, sex, race, and BMI in the CARDIA derivation
9 sample, and replicated its relation in the CARDIA validation sample. This MASLD score was
10 then recalibrated for use in UK Biobank and CCHC (which used Olink proteomics platforms in
11 contrast to CARDIA, which used a SomaScan platform), using LASSO regression with the
12 original MASLD score as the dependent variable and all overlapping proteins (matching
13 between the Olink and SomaScan platforms on UniProt identifier) as the independent variables.
14 Pathway analysis was performed on proteins that were significant in both CARDIA derivation
15 and validation subsamples (FDR<5%) using R package ClusterProfiler⁹³ on KEGG and
16 Reactome database. Hypergeometric tests were used to evaluate enrichment level for each
17 pathway utilizing all proteins on the SomaScan platform as background. The top 10 most
18 enriched pathways in both KEGG and Reactome were visualized together via lollipop plots. To
19 identify hub genes, Protein-protein interactions for 235 significant genes were retrieved from the
20 STRING database⁹⁴. Hub genes were determined as any protein with more than 5 high-
21 confidence interactions (score>700) and hub genes and all interactions were visualized using
22 Cytoscape⁹⁵. Tissue-specific gene expression enrichment was performed by R package
23 TissueEnrich⁹⁶ based on tissue expression patterns in Human Protein Atlas database⁹⁷. To
24 examine the clinical utility of the MASLD score as a diagnostic marker of MASLD, we compared
25 a clinical model of MASLD (age, sex, race, BMI, alcoholic drinks per week, AST, ALT,

1 hemoglobin A1c) to a model with the MASLD score using receiver operator curve area under
2 the curve analysis in both the CARDIA validation sample and UK Biobank.

3 **Testing the association between a protein score of MASLD with development of**

4 **MASLD and clinical outcomes:** In UK Biobank, Cox regression was used to examine the
5 relation of the MASLD score with clinical endpoints. Death and type of death (cardiovascular
6 death, cancer death, respiratory death) were defined by using death registry data (UK Biobank
7 Data Field 40000) in conjunction with the primary cause of death International Classification of
8 Disease (ICD) 10 code (UK Biobank Data Field 40001). Translating ICD10 codes to type of
9 death was conducted as previously reported⁹⁸. Censoring for clinical endpoints was determined
10 by region-specific censor dates for each participant based on the location of initial assessment
11 (UK Biobank Data Field 54). Deaths were censored on 30 November 2022 for all participants.
12 Non-death outcomes in UK Biobank were defined by ICD10 diagnosis codes grouped into
13 relevant “phecodes” via the PheWAS package⁹⁹. For each phecode, we generated a case,
14 control, and excluded status for each subject. Time to event for phecodes was defined as the
15 date of the earliest relevant ICD10 was documented. Prevalent conditions were defined by self-
16 report or physician diagnosis (Data Fields 20002, 2443, 6150). Sequential models with
17 increasing adjustments were created 1) unadjusted 2) age, sex, race, BMI 3) age, sex, race,
18 BMI, Townsend Deprivation Index, diabetes, smoking, alcohol use, systolic blood pressure, and
19 LDL. We conducted a sensitivity analysis including further adjustment for AST, ALT, and
20 hemoglobin A1c. We compared adjusted models (age, sex, race, BMI, diabetes [removed from
21 models for diabetes], smoking, alcohol use, systolic blood pressure, LDL) with and without the
22 MASLD score to compare differences in C-statistics and net reclassification index (NRI;
23 calculated at the 75th percentile for NRI for events). To investigate the effect of the MASLD
24 score on polygenic risk, we examined the relationship between the MASLD score with the
25 standard polygenic risk score (PRS) of diabetes (Data Field 26285) in a Cox model for incident
26 type 2 diabetes as a function of the MASLD score and the PRS for type 2 diabetes, including an

1 interaction term, with adjustments for age, sex, race, and the top 4 principal components of
2 genetic ancestry.

3

4 **ACKNOWLEDGEMENTS**

5 CARDIA is conducted and supported by the NHLBI in collaboration with the University of
6 Alabama at Birmingham (HHSN268201800005I & HHSN268201800007I), Northwestern
7 University (HHSN268201800003I), University of Minnesota (HHSN268201800006I), and Kaiser
8 Foundation Research Institute (HHSN268201800004I). Proteomics quantification was funded by
9 the NHLBI (HL122477; PI Kalhan). This manuscript has been reviewed by CARDIA for scientific
10 content. The views expressed in this manuscript are those of the authors and do not necessarily
11 represent the views of the NHLBI; the NIH; or the U.S. Department of Health and Human
12 Services. The authors would like to thank the CCHC cohort team, particularly Rocío Uribe who
13 recruited and interviewed the participants. Marcela Morris, BS, and Hugo Soriano and their
14 teams for laboratory and data support respectively; Norma Pérez-Olazarán, BBA, and Christina
15 Villarreal, BA for administrative support; Valley Baptist Medical Center, Brownsville, Texas, for
16 providing us space for our Center for Clinical and Translational Science Clinical Research Unit
17 is located; and the community of Brownsville and the participants who so willingly participated in
18 this study in their city. This study was funded in part by Center for Clinical and Translational
19 Sciences, National Institutes of Health Clinical and Translational Award grant no. UL1
20 TR000371 from the National Center for Advancing Translational Sciences.

21

22 **DISCLOSURES**

23 R.S. is supported in part by grants from the National Institutes of Health (NIH) and the American
24 Heart Association (AHA). R.S. has served for a consultant for Amgen and Cytokinetics. R.S. is a
25 co-inventor on a patent for ex-RNAs signatures of cardiac remodeling (not relevant to the
26 current work). A.S.P. is supported by the AHA Strategically Focused Research Network in

1 Cardiometabolic Disease. R.S., J.B., A.S.P. have filed for a patent relevant to the findings in this
2 manuscript. J.F.K.S. and G.M. are employees of Emulate Inc. (a maker of the liver-on-a-chip)
3 and may hold equity interest in Emulate, Inc. S.D. holds a research grant from Bristol Myers
4 Squibb, is a founder and holds equity in Switch Therapeutics, and is a founder and consultant
5 and holds equity for Thryv Therapeutics. N.B. receives consulting fees from Deepcell. J.R.K.
6 has served as a consultant to Gilead, Merck, ViiV Healthcare and Janssen and also received
7 research support from Gilead Sciences and Merck. R.K. is supported in part by grants from the
8 NIH, has received grants from AstraZeneca, PneumRx/BTG, and Spiration, has received
9 consulting fees from CVS Caremark, AstraZeneca, GlaxoSmithKline, and CSA Medical, and has
10 received speaking fees from GlaxoSmithKline, AstraZeneca, and Boehringer Ingelheim. K.A. is
11 supported by an AHA Career Development Award (#929347). J.A.F. serves as a consultant or
12 advisory board member for Kynos Therapeutics, Resolution Therapeutics, Ipsen, River 2 Renal
13 Corp., Stimuliver, Galecto Biotech, Global Clinical Trial Partners, and Guidepoint and has
14 received research grant funding from Intercept Pharmaceuticals and Genentech. T.J.K.
15 undertakes consultancy work for Perspectum, Clinnovate Health, Kynos Therapeutics, Fibrofind,
16 HistoIndex, Concept Life Sciences, and Resolution Therapeutics, and has received speaker's
17 fees from Incyte Corporation and Servier Laboratories. K.V.K.J. is a member of the scientific
18 advisory board at Dyrnamix. J.J.C. receives project funding from GE Healthcare, Siemens
19 Healthineers, TheraTech, and the NIH. M.N. has received speaking honoraria from Cytokinetics.
20 The other authors report no relevant financial disclosures.

21

1 REFERENCES:

- 2 1. Riazi K, Azhari H, Charette JH, Underwood FE, King JA, Afshar EE, Swain MG,
3 Congly SE, Kaplan GG, Shaheen AA. The prevalence and incidence of NAFLD
4 worldwide: a systematic review and meta-analysis. *Lancet Gastroenterol*
5 *Hepatol.* 2022;7:851-861. doi: 10.1016/S2468-1253(22)00165-0
- 6 2. Anstee QM, Reeves HL, Kotsiliti E, Govaere O, Heikenwalder M. From NASH to
7 HCC: current concepts and future challenges. *Nat Rev Gastroenterol Hepatol.*
8 2019;16:411-428. doi: 10.1038/s41575-019-0145-7
- 9 3. Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, Boerwinkle
10 E, Cohen JC, Hobbs HH. Genetic variation in PNPLA3 confers susceptibility to
11 nonalcoholic fatty liver disease. *Nature genetics.* 2008;40:1461-1465. doi:
12 10.1038/ng.257
- 13 4. Luo Y, Wadhawan S, Greenfield A, Decato BE, Oseini AM, Collen R, Shevell DE,
14 Thompson J, Jarai G, Charles ED, Sanyal AJ. SOMAscan Proteomics Identifies
15 Serum Biomarkers Associated With Liver Fibrosis in Patients With NASH.
16 *Hepatol Commun.* 2021;5:760-773. doi: 10.1002/hep4.1670
- 17 5. Corey KE, Pitts R, Lai M, Loureiro J, Masia R, Osganian SA, Gustafson JL,
18 Hutter MM, Gee DW, Meireles OR, et al. ADAMTSL2 protein and a soluble
19 biomarker signature identify at-risk non-alcoholic steatohepatitis and fibrosis in
20 adults with NAFLD. *Journal of hepatology.* 2022;76:25-33. doi:
21 10.1016/j.jhep.2021.09.026
- 22 6. Sanyal AJ, Williams SA, Lavine JE, Neuschwander-Tetri BA, Alexander L, Ostroff
23 R, Biegel H, Kowdley KV, Chalasani N, Dasarathy S, et al. Defining the serum
24 proteomic signature of hepatic steatosis, inflammation, ballooning and fibrosis in
25 non-alcoholic fatty liver disease. *Journal of hepatology.* 2023;78:693-703. doi:
26 10.1016/j.jhep.2022.11.029
- 27 7. Kendall TJ, Jimenez-Ramos M, Turner F, Ramachandran P, Minnier J,
28 McColgan MD, Alam M, Ellis H, Dunbar DR, Kohnen G, et al. An integrated
29 gene-to-outcome multimodal database for metabolic dysfunction-associated
30 steatotic liver disease. *Nat Med.* 2023;29:2939-2953. doi: 10.1038/s41591-023-
31 02602-2
- 32 8. Govaere O, Hasoon M, Alexander L, Cockell S, Tiniakos D, Ekstedt M,
33 Schattenberg JM, Boursier J, Bugianesi E, Ratziu V, et al. A proteo-
34 transcriptomic map of non-alcoholic fatty liver disease signatures. *Nat Metab.*
35 2023;5:572-578. doi: 10.1038/s42255-023-00775-1
- 36 9. Niu L, Thiele M, Geyer PE, Rasmussen DN, Webel HE, Santos A, Gupta R,
37 Meier F, Strauss M, Kjaergaard M, et al. Noninvasive proteomic biomarkers for
38 alcohol-related liver disease. *Nature medicine.* 2022;28:1277-1287. doi:
39 10.1038/s41591-022-01850-y
- 40 10. VanWagner LB, Wilcox JE, Ning H, Lewis CE, Carr JJ, Rinella ME, Shah SJ,
41 Lima JAC, Lloyd-Jones DM. Longitudinal Association of Non-Alcoholic Fatty Liver
42 Disease With Changes in Myocardial Structure and Function: The CARDIA
43 Study. *J Am Heart Assoc.* 2020;9:e014279. doi: 10.1161/JAHA.119.014279
- 44 11. Pirola CJ, Sookoian S. Multiomics biomarkers for the prediction of nonalcoholic
45 fatty liver disease severity. *World journal of gastroenterology : WJG.*
46 2018;24:1601-1615. doi: 10.3748/wjg.v24.i15.1601

- 1 12. Ma J, Tan X, Kwon Y, Delgado ER, Zarnegar A, DeFrances MC, Duncan AW,
2 Zarnegar R. A Novel Humanized Model of NASH and Its Treatment With META4,
3 A Potent Agonist of MET. *Cell Mol Gastroenterol Hepatol*. 2022;13:565-582. doi:
4 10.1016/j.jcmgh.2021.10.007
- 5 13. Li H, Toth E, Cherrington NJ. Alcohol Metabolism in the Progression of Human
6 Nonalcoholic Steatohepatitis. *Toxicol Sci*. 2018;164:428-438. doi:
7 10.1093/toxsci/kfy106
- 8 14. Aljomah G, Baker SS, Liu W, Kozielski R, Oluwole J, Lupu B, Baker RD, Zhu L.
9 Induction of CYP2E1 in non-alcoholic fatty liver diseases. *Exp Mol Pathol*.
10 2015;99:677-681. doi: 10.1016/j.yexmp.2015.11.008
- 11 15. Niu L, Geyer PE, Wewer Albrechtsen NJ, Gluud LL, Santos A, Doll S, Treit PV,
12 Holst JJ, Knop FK, Vilsboll T, et al. Plasma proteome profiling discovers novel
13 proteins associated with non-alcoholic fatty liver disease. *Mol Syst Biol*.
14 2019;15:e8793. doi: 10.15252/msb.20188793
- 15 16. Gathercole LL, Nikolaou N, Harris SE, Arvaniti A, Poolman TM, Hazlehurst JM,
16 Kratschmar DV, Todorcevic M, Moolla A, Dempster N, et al. AKR1D1 knockout
17 mice develop a sex-dependent metabolic phenotype. *J Endocrinol*. 2022;253:97-
18 113. doi: 10.1530/JOE-21-0280
- 19 17. Zeng CM, Chang LL, Ying MD, Cao J, He QJ, Zhu H, Yang B. Aldo-Keto
20 Reductase AKR1C1-AKR1C4: Functions, Regulation, and Intervention for Anti-
21 cancer Therapy. *Front Pharmacol*. 2017;8:119. doi: 10.3389/fphar.2017.00119
- 22 18. Lyall MJ, Cartier J, Thomson JP, Cameron K, Meseguer-Ripolles J, O'Duibhir E,
23 Szkolnicka D, Villarin BL, Wang Y, Blanco GR, et al. Modelling non-alcoholic fatty
24 liver disease in human hepatocyte-like cells. *Philos Trans R Soc Lond B Biol Sci*.
25 2018;373. doi: 10.1098/rstb.2017.0362
- 26 19. Gorce M, Lebigot E, Arion A, Brassier A, Cano A, De Lonlay P, Feillet F, Gay C,
27 Labarthe F, Nassogne MC, et al. Fructose-1,6-bisphosphatase deficiency causes
28 fatty liver disease and requires long-term hepatic follow-up. *J Inherit Metab Dis*.
29 2022;45:215-222. doi: 10.1002/jimd.12452
- 30 20. Coles BF, Kadlubar FF. Human alpha class glutathione S-transferases: genetic
31 polymorphism, expression, and susceptibility to disease. *Methods Enzymol*.
32 2005;401:9-42. doi: 10.1016/S0076-6879(05)01002-5
- 33 21. Nagamani SC, Erez A, Lee B. Argininosuccinate lyase deficiency. *Genet Med*.
34 2012;14:501-507. doi: 10.1038/gim.2011.1
- 35 22. Zhang T, Zhang N, Xing J, Zhang S, Chen Y, Xu D, Gu J. UDP-glucuronate
36 metabolism controls RIPK1-driven liver damage in nonalcoholic steatohepatitis.
37 *Nat Commun*. 2023;14:2715. doi: 10.1038/s41467-023-38371-2
- 38 23. Bhushan B, Banerjee S, Paranjpe S, Koral K, Mars WM, Stoops JW, Orr A,
39 Bowen WC, Locker J, Michalopoulos GK. Pharmacologic Inhibition of Epidermal
40 Growth Factor Receptor Suppresses Nonalcoholic Fatty Liver Disease in a
41 Murine Fast-Food Diet Model. *Hepatology*. 2019;70:1546-1563. doi:
42 10.1002/hep.30696
- 43 24. Ohkubo R, Mu WC, Wang CL, Song Z, Barthez M, Wang Y, Mitchener N,
44 Abdullayev R, Lee YR, Ma Y, et al. The hepatic integrated stress response
45 suppresses the somatotroph axis to control liver damage in nonalcoholic fatty
46 liver disease. *Cell reports*. 2022;41:111803. doi: 10.1016/j.celrep.2022.111803

- 1 25. Kroy DC, Schumacher F, Ramadori P, Hatting M, Bergheim I, Gassler N,
2 Boekschoten MV, Muller M, Streetz KL, Trautwein C. Hepatocyte specific
3 deletion of c-Met leads to the development of severe non-alcoholic
4 steatohepatitis in mice. *Journal of hepatology*. 2014;61:883-890. doi:
5 10.1016/j.jhep.2014.05.019
- 6 26. Xiong X, Kuang H, Ansari S, Liu T, Gong J, Wang S, Zhao XY, Ji Y, Li C, Guo L,
7 et al. Landscape of Intercellular Crosstalk in Healthy and NASH Liver Revealed
8 by Single-Cell Secretome Gene Analysis. *Molecular cell*. 2019;75:644-660 e645.
9 doi: 10.1016/j.molcel.2019.07.028
- 10 27. Del Ben M, Overi D, Polimeni L, Carpino G, Labbadia G, Baratta F, Pastori D,
11 Noce V, Gaudio E, Angelico F, Mancone C. Overexpression of the Vitronectin
12 V10 Subunit in Patients with Nonalcoholic Steatohepatitis: Implications for
13 Noninvasive Diagnosis of NASH. *International journal of molecular sciences*.
14 2018;19. doi: 10.3390/ijms19020603
- 15 28. Schwabe RF, Tabas I, Pajvani UB. Mechanisms of Fibrosis Development in
16 Nonalcoholic Steatohepatitis. *Gastroenterology*. 2020;158:1913-1928. doi:
17 10.1053/j.gastro.2019.11.311
- 18 29. Yang L, Sun Z, Li J, Pan X, Wen J, Yang J, Wang Q, Chen P. Genetic Variants
19 of Glycogen Metabolism Genes Were Associated With Liver PDFF Without
20 Increasing NAFLD Risk. *Frontiers in genetics*. 2022;13:830445. doi:
21 10.3389/fgene.2022.830445
- 22 30. Chella Krishnan K, Floyd RR, Sabir S, Jayasekera DW, Leon-Mimila PV, Jones
23 AE, Cortez AA, Shrivah V, Peterfy M, Stiles L, et al. Liver Pyruvate Kinase
24 Promotes NAFLD/NASH in Both Mice and Humans in a Sex-Specific Manner.
25 *Cell Mol Gastroenterol Hepatol*. 2021;11:389-406. doi:
26 10.1016/j.jcmgh.2020.09.004
- 27 31. Saed CT, Tabatabaei Dakhili SA, Ussher JR. Pyruvate Dehydrogenase as a
28 Therapeutic Target for Nonalcoholic Fatty Liver Disease. *ACS Pharmacol Transl*
29 *Sci*. 2021;4:582-588. doi: 10.1021/acspsci.0c00208
- 30 32. Browning JD, Szczepaniak LS, Dobbins R, Nuremberg P, Horton JD, Cohen JC,
31 Grundy SM, Hobbs HH. Prevalence of hepatic steatosis in an urban population in
32 the United States: impact of ethnicity. *Hepatology*. 2004;40:1387-1395. doi:
33 10.1002/hep.20466
- 34 33. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P,
35 Lander ES, Lubitz SA, Ellinor PT, Kathiresan S. Genome-wide polygenic scores
36 for common diseases identify individuals with risk equivalent to monogenic
37 mutations. *Nat Genet*. 2018;50:1219-1224. doi: 10.1038/s41588-018-0183-z
- 38 34. Guilliams M, Bonnardel J, Haest B, Vanderborght B, Wagner C, Remmerie A,
39 Bujko A, Martens L, Thone T, Browaeys R, et al. Spatial proteogenomics reveals
40 distinct and evolutionarily conserved hepatic macrophage niches. *Cell*.
41 2022;185:379-396 e338. doi: 10.1016/j.cell.2021.12.018
- 42 35. Faramia J, Hao Z, Mumphrey MB, Townsend RL, Miard S, Carreau AM, Nadeau
43 M, Frisch F, Baraboi ED, Grenier-Larouche T, et al. IGFBP-2 partly mediates the
44 early metabolic improvements caused by bariatric surgery. *Cell Rep Med*.
45 2021;2:100248. doi: 10.1016/j.xcrm.2021.100248

- 1 36. Lin YH, Wei Y, Zeng Q, Wang Y, Pagani CA, Li L, Zhu M, Wang Z, Hsieh MH,
2 Corbitt N, et al. IGFBP2 expressing midlobular hepatocytes preferentially
3 contribute to liver homeostasis and regeneration. *Cell Stem Cell*. 2023;30:665-
4 676 e664. doi: 10.1016/j.stem.2023.04.007
- 5 37. Hedbacker K, Birsoy K, Wysocki RW, Asilmaz E, Ahima RS, Farooqi IS,
6 Friedman JM. Antidiabetic effects of IGFBP2, a leptin-regulated gene. *Cell*
7 *metabolism*. 2010;11:11-22. doi: 10.1016/j.cmet.2009.11.007
- 8 38. Chen X, Tang Y, Chen S, Ling W, Wang Q. IGFBP-2 as a biomarker in NAFLD
9 improves hepatic steatosis: an integrated bioinformatics and experimental study.
10 *Endocrine connections*. 2021;10:1315-1325. doi: 10.1530/EC-21-0353
- 11 39. Lu D, Xia Q, Yang Z, Gao S, Sun S, Luo X, Li Z, Zhang X, Han S, Li X, Cao M.
12 ENO3 promoted the progression of NASH by negatively regulating ferroptosis via
13 elevation of GPX4 expression and lipid accumulation. *Annals of translational*
14 *medicine*. 2021;9:661. doi: 10.21037/atm-21-471
- 15 40. Wang J, Chen L, Li Y, Guan XY. Overexpression of cathepsin Z contributes to
16 tumor metastasis by inducing epithelial-mesenchymal transition in hepatocellular
17 carcinoma. *PLoS One*. 2011;6:e24967. doi: 10.1371/journal.pone.0024967
- 18 41. Zhang J, Fan N, Peng Y. Heat shock protein 70 promotes lipogenesis in HepG2
19 cells. *Lipids in health and disease*. 2018;17:73. doi: 10.1186/s12944-018-0722-8
- 20 42. Chen D, Dong X, Chen D, Lin J, Lu T, Shen J, Ye H. Cdh1 plays a protective role
21 in nonalcoholic fatty liver disease by regulating PPAR/PGC-1alpha signaling
22 pathway. *Biochemical and biophysical research communications*. 2023;681:13-
23 19. doi: 10.1016/j.bbrc.2023.09.038
- 24 43. Henkel AS, Khan SS, Olivares S, Miyata T, Vaughan DE. Inhibition of
25 Plasminogen Activator Inhibitor 1 Attenuates Hepatic Steatosis but Does Not
26 Prevent Progressive Nonalcoholic Steatohepatitis in Mice. *Hepatology*.
27 2018;2:1479-1492. doi: 10.1002/hep4.1259
- 28 44. Lee SM, Dorotea D, Jung I, Nakabayashi T, Miyata T, Ha H. TM5441, a
29 plasminogen activator inhibitor-1 inhibitor, protects against high fat diet-induced
30 non-alcoholic fatty liver disease. *Oncotarget*. 2017;8:89746-89760. doi:
31 10.18632/oncotarget.21120
- 32 45. Day K, Seale LA, Graham RM, Cardoso BR. Selenotranscriptome Network in
33 Non-alcoholic Fatty Liver Disease. *Frontiers in nutrition*. 2021;8:744825. doi:
34 10.3389/fnut.2021.744825
- 35 46. Wang L, Zhou K, Wu Q, Zhu L, Hu Y, Yang X, Li D. Microanatomy of the
36 metabolic associated fatty liver disease (MAFLD) by single-cell transcriptomics. *J*
37 *Drug Target*. 2023;31:421-432. doi: 10.1080/1061186X.2023.2185626
- 38 47. Sim WC, Lee W, Sim H, Lee KY, Jung SH, Choi YJ, Kim HY, Kang KW, Lee JY,
39 Choi YJ, et al. Downregulation of PHGDH expression and hepatic serine level
40 contribute to the development of fatty liver disease. *Metabolism*.
41 2020;102:154000. doi: 10.1016/j.metabol.2019.154000
- 42 48. Kuwashiro S, Terai S, Oishi T, Fujisawa K, Matsumoto T, Nishina H, Sakaida I.
43 Telmisartan improves nonalcoholic steatohepatitis in medaka (*Oryzias latipes*) by
44 reducing macrophage infiltration and fat accumulation. *Cell and tissue research*.
45 2011;344:125-134. doi: 10.1007/s00441-011-1132-7

- 1 49. Moravcova A, Cervinkova Z, Kucera O, Mezera V, Rychtrmoc D, Lotkova H. The
2 effect of oleic and palmitic acid on induction of steatosis and cytotoxicity on rat
3 hepatocytes in primary culture. *Physiol Res*. 2015;64:S627-636. doi:
4 10.33549/physiolres.933224
- 5 50. Enooku K, Kondo M, Fujiwara N, Sasako T, Shibahara J, Kado A, Okushin K,
6 Fujinaga H, Tsutsumi T, Nakagomi R, et al. Hepatic IRS1 and ss-catenin
7 expression is associated with histological progression and overt diabetes
8 emergence in NAFLD patients. *J Gastroenterol*. 2018;53:1261-1275. doi:
9 10.1007/s00535-018-1472-0
- 10 51. Honma M, Sawada S, Ueno Y, Murakami K, Yamada T, Gao J, Kodama S, Izumi
11 T, Takahashi K, Tsukita S, et al. Selective insulin resistance with differential
12 expressions of IRS-1 and IRS-2 in human NAFLD livers. *Int J Obes (Lond)*.
13 2018;42:1544-1555. doi: 10.1038/s41366-018-0062-9
- 14 52. Liss KH, Finck BN. PPARs and nonalcoholic fatty liver disease. *Biochimie*.
15 2017;136:65-74. doi: 10.1016/j.biochi.2016.11.009
- 16 53. Montagner A, Polizzi A, Fouche E, Ducheix S, Lippi Y, Lasserre F, Barquissau V,
17 Regnier M, Lukowicz C, Benhamed F, et al. Liver PPARalpha is crucial for
18 whole-body fatty acid homeostasis and is protective against NAFLD. *Gut*.
19 2016;65:1202-1214. doi: 10.1136/gutjnl-2015-310798
- 20 54. Ferre P, Foufelle F. Hepatic steatosis: a role for de novo lipogenesis and the
21 transcription factor SREBP-1c. *Diabetes Obes Metab*. 2010;12 Suppl 2:83-92.
22 doi: 10.1111/j.1463-1326.2010.01275.x
- 23 55. Pettinelli P, Videla LA. Up-regulation of PPAR-gamma mRNA expression in the
24 liver of obese patients: an additional reinforcing lipogenic mechanism to SREBP-
25 1c induction. *J Clin Endocrinol Metab*. 2011;96:1424-1430. doi: 10.1210/jc.2010-
26 2129
- 27 56. Moreno-Vedia J, Girona J, Ibarretxe D, Masana L, Rodriguez-Calvo R. Unveiling
28 the Role of the Fatty Acid Binding Protein 4 in the Metabolic-Associated Fatty
29 Liver Disease. *Biomedicines*. 2022;10. doi: 10.3390/biomedicines10010197
- 30 57. Wood GC, Chu X, Argyropoulos G, Benotti P, Rolston D, Mirshahi T, Petrick A,
31 Gabrielson J, Carey DJ, DiStefano JK, et al. A multi-component classifier for
32 nonalcoholic fatty liver disease (NAFLD) based on genomic, proteomic, and
33 phenomic data domains. *Scientific reports*. 2017;7:43238. doi:
34 10.1038/srep43238
- 35 58. Sveinbjornsson G, Ulfarsson MO, Thorolfsdottir RB, Jonsson BA, Einarsson E,
36 Gunnlaugsson G, Rognvaldsson S, Arnar DO, Baldvinsson M, Bjarnason RG, et
37 al. Multiomics study of nonalcoholic fatty liver disease. *Nature genetics*.
38 2022;54:1652-1663. doi: 10.1038/s41588-022-01199-5
- 39 59. Gu J, Liu S, Du S, Zhang Q, Xiao J, Dong Q, Xin Y. Diagnostic value of MRI-
40 PDFF for hepatic steatosis in patients with non-alcoholic fatty liver disease: a
41 meta-analysis. *European radiology*. 2019;29:3564-3573. doi: 10.1007/s00330-
42 019-06072-4
- 43 60. Bril F, Ortiz-Lopez C, Lomonaco R, Orsak B, Freckleton M, Chintapalli K, Hardies
44 J, Lai S, Solano F, Tio F, Cusi K. Clinical value of liver ultrasound for the
45 diagnosis of nonalcoholic fatty liver disease in overweight and obese patients.
46 *Liver Int*. 2015;35:2139-2146. doi: 10.1111/liv.12840

- 1 61. Perakakis N, Stefanakis K, Mantzoros CS. The role of omics in the
2 pathophysiology, diagnosis and treatment of non-alcoholic fatty liver disease.
3 *Metabolism*. 2020;111S:154320. doi: 10.1016/j.metabol.2020.154320
- 4 62. Kotronen A, Peltonen M, Hakkarainen A, Sevastianova K, Bergholm R,
5 Johansson LM, Lundbom N, Rissanen A, Ridderstrale M, Groop L, et al.
6 Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and
7 genetic factors. *Gastroenterology*. 2009;137:865-872. doi:
8 10.1053/j.gastro.2009.06.005
- 9 63. Walker RW, Belbin GM, Sorokin EP, Van Vleck T, Wojcik GL, Moscati A,
10 Gignoux CR, Cho J, Abul-Husn NS, Nadkarni G, et al. A common variant in
11 PNPLA3 is associated with age at diagnosis of NAFLD in patients from a multi-
12 ethnic biobank. *Journal of hepatology*. 2020;72:1070-1081. doi:
13 10.1016/j.jhep.2020.01.029
- 14 64. Corbin KD, Abdelmalek MF, Spencer MD, da Costa KA, Galanko JA, Sha W,
15 Suzuki A, Guy CD, Cardona DM, Torquati A, et al. Genetic signatures in choline
16 and 1-carbon metabolism are associated with the severity of hepatic steatosis.
17 *FASEB journal : official publication of the Federation of American Societies for*
18 *Experimental Biology*. 2013;27:1674-1689. doi: 10.1096/fj.12-219097
- 19 65. Fairfield CJ, Drake TM, Pius R, Bretherick AD, Campbell A, Clark DW,
20 Fallowfield JA, Hayward C, Henderson NC, Joshi PK, et al. Genome-Wide
21 Association Study of NAFLD Using Electronic Health Records. *Hepatol Commun*.
22 2022;6:297-308. doi: 10.1002/hep4.1805
- 23 66. Anstee QM, Darlay R, Cockell S, Meroni M, Govaere O, Tiniakos D, Burt AD,
24 Bedossa P, Palmer J, Liu YL, et al. Genome-wide association study of non-
25 alcoholic fatty liver and steatohepatitis in a histologically characterised cohort(☆).
26 *Journal of hepatology*. 2020;73:505-515. doi: 10.1016/j.jhep.2020.04.003
- 27 67. Tan Y, Ge G, Pan T, Wen D, Gan J. A pilot study of serum microRNAs panel as
28 potential biomarkers for diagnosis of nonalcoholic fatty liver disease. *PLoS One*.
29 2014;9:e105192. doi: 10.1371/journal.pone.0105192
- 30 68. Wu J, Zhang R, Shen F, Yang R, Zhou D, Cao H, Chen G, Pan Q, Fan J. Altered
31 DNA Methylation Sites in Peripheral Blood Leukocytes from Patients with Simple
32 Steatosis and Nonalcoholic Steatohepatitis (NASH). *Med Sci Monit*.
33 2018;24:6946-6967. doi: 10.12659/MSM.909747
- 34 69. Dessein A. Clinical utility of polygenic risk scores for predicting NAFLD disorders.
35 *Journal of hepatology*. 2021;74:769-770. doi: 10.1016/j.jhep.2021.02.005
- 36 70. Oh HS, Rutledge J, Nachun D, Palovics R, Abiose O, Moran-Losada P,
37 Channappa D, Urey DY, Kim K, Sung YJ, et al. Organ aging signatures in the
38 plasma proteome track health and disease. *Nature*. 2023;624:164-172. doi:
39 10.1038/s41586-023-06802-1
- 40 71. Deczkowska A, David E, Ramadori P, Pfister D, Safran M, Li B, Giladi A, Jaitin
41 DA, Barbov O, Cohen M, et al. XCR1(+) type 1 conventional dendritic cells drive
42 liver pathology in non-alcoholic steatohepatitis. *Nat Med*. 2021;27:1043-1054.
43 doi: 10.1038/s41591-021-01344-3
- 44 72. Rennert K, Steinborn S, Groger M, Ungerbock B, Jank AM, Ehgartner J,
45 Nietzsche S, Dinger J, Kiehntopf M, Funke H, et al. A microfluidically perfused

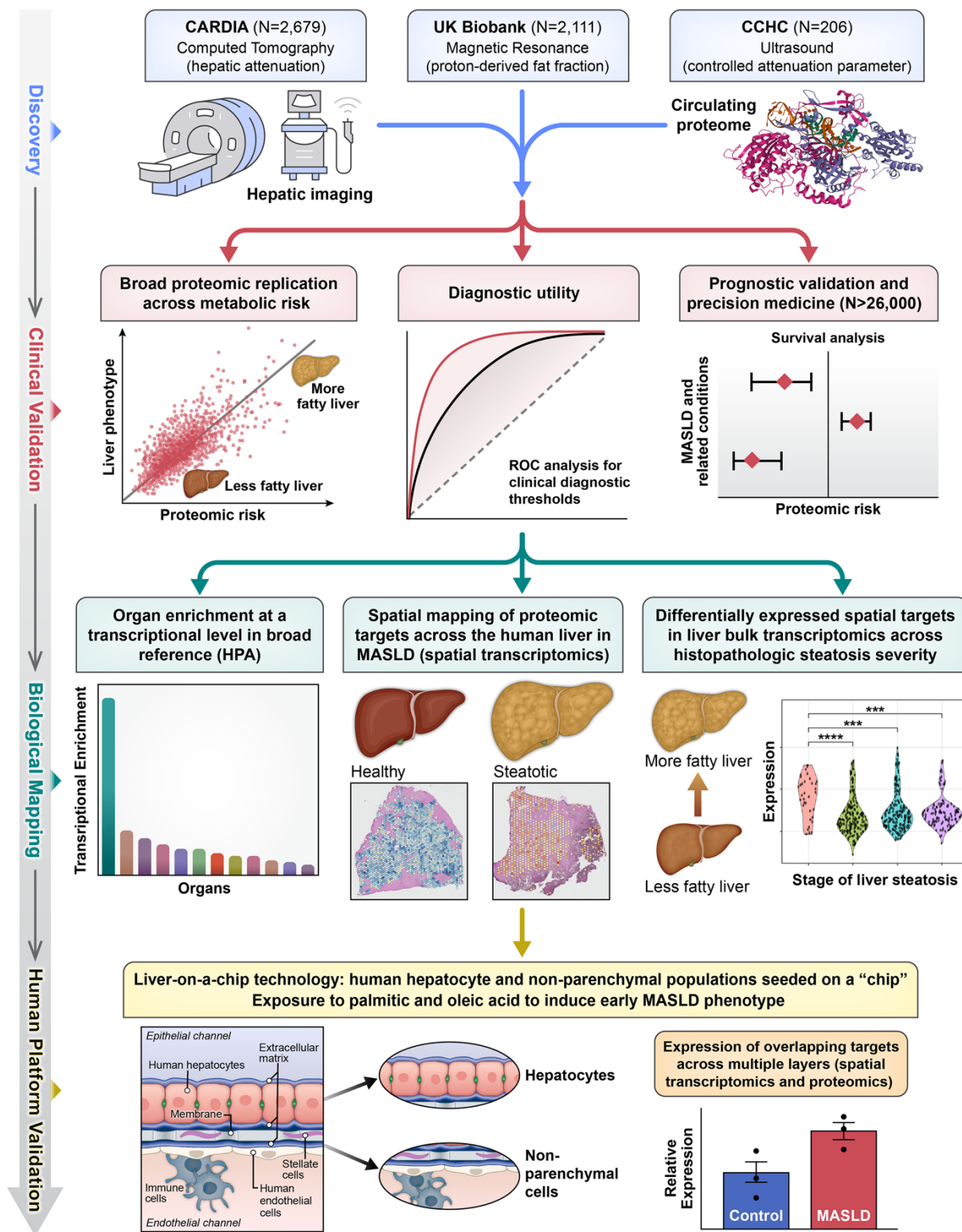
- 1 three dimensional human liver model. *Biomaterials*. 2015;71:119-131. doi:
2 10.1016/j.biomaterials.2015.08.043
- 3 73. Hassan S, Sebastian S, Maharjan S, Lesha A, Carpenter AM, Liu X, Xie X,
4 Livermore C, Zhang YS, Zarrinpar A. Liver-on-a-Chip Models of Fatty Liver
5 Disease. *Hepatology*. 2020;71:733-740. doi: 10.1002/hep.31106
- 6 74. Du K, Li S, Li C, Li P, Miao C, Luo T, Qiu B, Ding W. Modeling nonalcoholic fatty
7 liver disease on a liver lobule chip with dual blood supply. *Acta Biomater*.
8 2021;134:228-239. doi: 10.1016/j.actbio.2021.07.013
- 9 75. Levner D, Ewart L. Integrating Liver-Chip data into pharmaceutical decision-
10 making processes. *Expert Opin Drug Discov*. 2023;18:1313-1320. doi:
11 10.1080/17460441.2023.2255127
- 12 76. Ewart L, Apostolou A, Briggs SA, Carman CV, Chaff JT, Heng AR, Jadalannagari
13 S, Janardhanan J, Jang KJ, Joshipura SR, et al. Performance assessment and
14 economic analysis of a human Liver-Chip for predictive toxicology. *Commun Med*
15 *(Lond)*. 2022;2:154. doi: 10.1038/s43856-022-00209-1
- 16 77. Jang KJ, Otieno MA, Ronxhi J, Lim HK, Ewart L, Kodella KR, Petropolis DB,
17 Kulkarni G, Rubins JE, Conegliano D, et al. Reproducing human and cross-
18 species drug toxicities using a Liver-Chip. *Sci Transl Med*. 2019;11. doi:
19 10.1126/scitranslmed.aax5516
- 20 78. Cong Y, Han X, Wang Y, Chen Z, Lu Y, Liu T, Wu Z, Jin Y, Luo Y, Zhang X. Drug
21 Toxicity Evaluation Based on Organ-on-a-chip Technology: A Review.
22 *Micromachines (Basel)*. 2020;11. doi: 10.3390/mi11040381
- 23 79. Gori M, Simonelli MC, Giannitelli SM, Businaro L, Trombetta M, Rainer A.
24 Investigating Nonalcoholic Fatty Liver Disease in a Liver-on-a-Chip Microfluidic
25 Device. *PLoS One*. 2016;11:e0159729. doi: 10.1371/journal.pone.0159729
- 26 80. Sun BB, Chiou J, Traylor M, Benner C, Hsu YH, Richardson TG, Surendran P,
27 Mahajan A, Robins C, Vasquez-Grinnell SG, et al. Plasma proteomic
28 associations with genetics and health in the UK Biobank. *Nature*. 2023;622:329-
29 338. doi: 10.1038/s41586-023-06592-6
- 30 81. Wagenknecht LE, Perkins LL, Cutter GR, Sidney S, Burke GL, Manolio TA,
31 Jacobs DR, Jr., Liu KA, Friedman GD, Hughes GH, et al. Cigarette smoking
32 behavior is strongly related to educational status: the CARDIA study. *Preventive*
33 *medicine*. 1990;19:158-169.
- 34 82. Dyer AR, Cutter GR, Liu KQ, Armstrong MA, Friedman GD, Hughes GH, Dolce
35 JJ, Raczynski J, Burke G, Manolio T. Alcohol intake and blood pressure in young
36 adults: the CARDIA Study. *Journal of clinical epidemiology*. 1990;43:1-13.
- 37 83. Bild DE, Jacobs DR, Jr., Sidney S, Haskell WL, Anderssen N, Oberman A.
38 Physical activity in young black and white women. The CARDIA Study. *Ann*
39 *Epidemiol*. 1993;3:636-644.
- 40 84. Sidney S, Jacobs DR, Jr., Haskell WL, Armstrong MA, Dimicco A, Oberman A,
41 Savage PJ, Slattery ML, Sternfeld B, Van Horn L. Comparison of two methods of
42 assessing physical activity in the Coronary Artery Risk Development in Young
43 Adults (CARDIA) Study. *Am J Epidemiol*. 1991;133:1231-1245.
- 44 85. Jiao J, Watt GP, Lee M, Rahbar MH, Vatcheva KP, Pan JJ, McCormick JB,
45 Fisher-Hoch SP, Fallon MB, Beretta L. Cirrhosis and Advanced Fibrosis in

- 1 Hispanics in Texas: The Dominant Contribution of Central Obesity. *PLoS One*.
2 2016;11:e0150978. doi: 10.1371/journal.pone.0150978
- 3 86. de Ledingham V, Vergniol J, Capdepon M, Chermak F, Hiriart JB, Cassinotto C,
4 Merrouche W, Foucher J, Brigitte le B. Controlled attenuation parameter (CAP)
5 for the diagnosis of steatosis: a prospective study of 5323 examinations. *J*
6 *Hepatol*. 2014;60:1026-1031. doi: 10.1016/j.jhep.2013.12.018
- 7 87. Wilman HR, Kelly M, Garratt S, Matthews PM, Milanese M, Herlihy A, Gyngell M,
8 Neubauer S, Bell JD, Banerjee R, Thomas EL. Characterisation of liver fat in the
9 UK Biobank cohort. *PLoS One*. 2017;12:e0172921. doi:
10 10.1371/journal.pone.0172921
- 11 88. Watt GP, De La Cerda I, Pan JJ, Fallon MB, Beretta L, Loomba R, Lee M,
12 McCormick JB, Fisher-Hoch SP. Elevated Glycated Hemoglobin Is Associated
13 With Liver Fibrosis, as Assessed by Elastography, in a Population-Based Study
14 of Mexican Americans. *Hepatol Commun*. 2020;4:1793-1801. doi:
15 10.1002/hep4.1603
- 16 89. van der Poorten D, Samer CF, Ramezani-Moghadam M, Coulter S, Kacevska M,
17 Schrijnders D, Wu LE, McLeod D, Bugianesi E, Komuta M, et al. Hepatic fat loss
18 in advanced nonalcoholic steatohepatitis: are alterations in serum adiponectin
19 the cause? *Hepatology*. 2013;57:2180-2188. doi: 10.1002/hep.26072
- 20 90. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc*
21 *Natl Acad Sci U S A*. 2003;100:9440-9445. doi: 10.1073/pnas.1530509100
- 22 91. Robinson MD, Oshlack A. A scaling normalization method for differential
23 expression analysis of RNA-seq data. *Genome Biol*. 2010;11:R25. doi:
24 10.1186/gb-2010-11-3-r25
- 25 92. Chatterjee E, Rodosthenous RS, Kujala V, Gokulnath P, Spanos M, Lehmann HI,
26 de Oliveira GP, Shi M, Miller-Fleming TW, Li G, et al. Circulating extracellular
27 vesicles in human cardiorenal syndrome promote renal injury in a kidney-on-chip
28 system. *JCI Insight*. 2023;8. doi: 10.1172/jci.insight.165172
- 29 93. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing
30 biological themes among gene clusters. *OMICS*. 2012;16:284-287. doi:
31 10.1089/omi.2011.0118
- 32 94. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL,
33 Fang T, Doncheva NT, Pyysalo S, et al. The STRING database in 2023: protein-
34 protein association networks and functional enrichment analyses for any
35 sequenced genome of interest. *Nucleic Acids Res*. 2023;51:D638-D646. doi:
36 10.1093/nar/gkac1000
- 37 95. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N,
38 Schwikowski B, Ideker T. Cytoscape: a software environment for integrated
39 models of biomolecular interaction networks. *Genome Res*. 2003;13:2498-2504.
40 doi: 10.1101/gr.1239303
- 41 96. Jain A, Tuteja G. TissueEnrich: Tissue-specific gene enrichment analysis.
42 *Bioinformatics*. 2019;35:1966-1967. doi: 10.1093/bioinformatics/bty890
- 43 97. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A,
44 Sivertsson A, Kampf C, Sjostedt E, Asplund A, et al. Proteomics. Tissue-based
45 map of the human proteome. *Science*. 2015;347:1260419. doi:
46 10.1126/science.1260419

- 1 98. Gonzales TI, Westgate K, Strain T, Hollidge S, Jeon J, Christensen DL, Jensen
2 J, Wareham NJ, Brage S. Cardiorespiratory fitness assessment using risk-
3 stratified exercise testing and dose-response relationships with disease
4 outcomes. *Sci Rep.* 2021;11:15315. doi: 10.1038/s41598-021-94768-3
5 99. Wu P, Gifford A, Meng X, Li X, Campbell H, Varley T, Zhao J, Carroll R,
6 Bastarache L, Denny JC, et al. Mapping ICD-10 and ICD-10-CM Codes to
7 Phecodes: Workflow Development and Initial Evaluation. *JMIR Med Inform.*
8 2019;7:e14325. doi: 10.2196/14325
9
10
11

1 **Figures**

2 **Figure 1: Study diagram: Study design.**



3

Figure 2. Proteins related to hepatic steatosis are primarily expressed in the liver and identify pathways of metabolism. (A) Volcano plot of proteins associated with hepatic steatosis after adjustment for age, sex, race, and BMI. For visualization, proteins with an FDR<5% in CARDIA derivation subsample are visualized with the beta coefficient and p values presented coming from models using the CARDIA validation subsample. **(B)** Heatmap of the top 25 positively associated and top 25 negatively associated proteins with hepatic steatosis in the CARDIA validation sample. **(C)** Tissue expression analysis of the proteins related to hepatic steatosis in CARDIA using the full SomaScan 7k platform as the background, demonstrated enrichment of proteins expressed in the liver. **(D)** KEGG and Reactome pathway analysis. **(E)** Hub gene analysis of significant proteins associated with liver attenuation showing the hub genes (≥ 5 connections; circles) and all proteins with high confidence connections to the hub genes (rectangles).

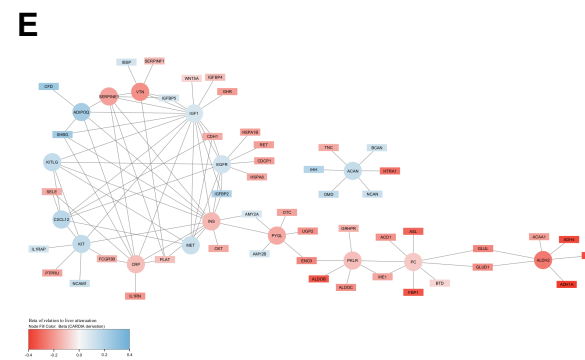
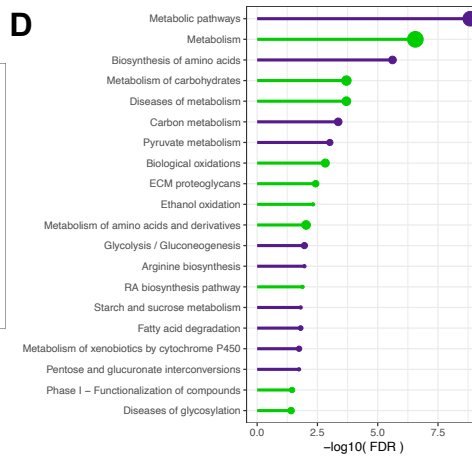
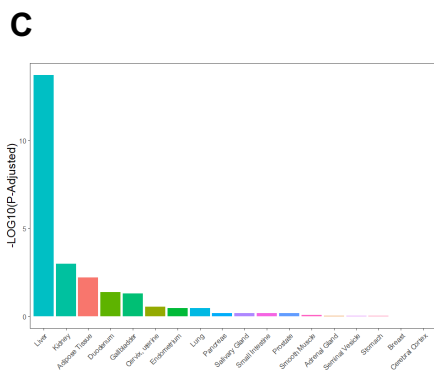
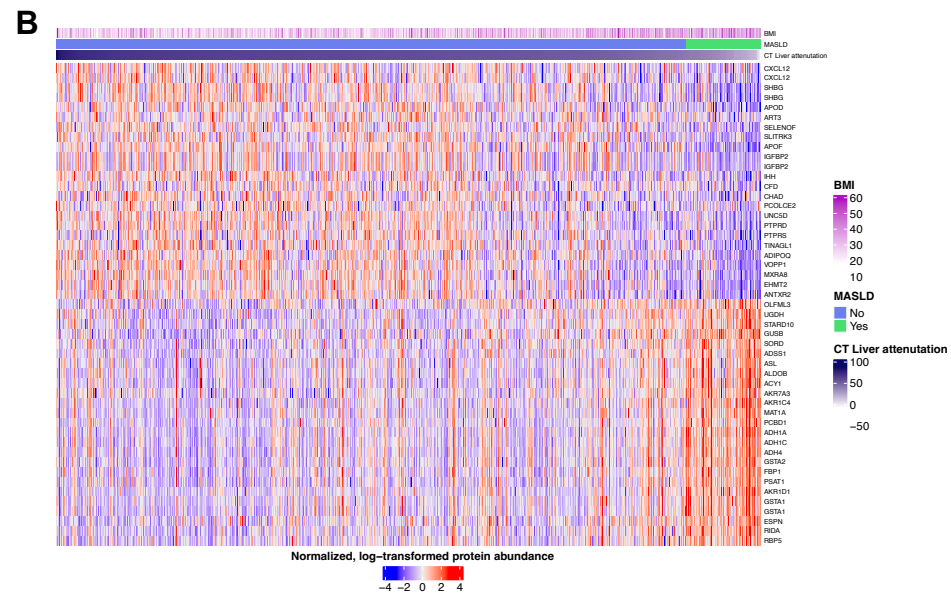
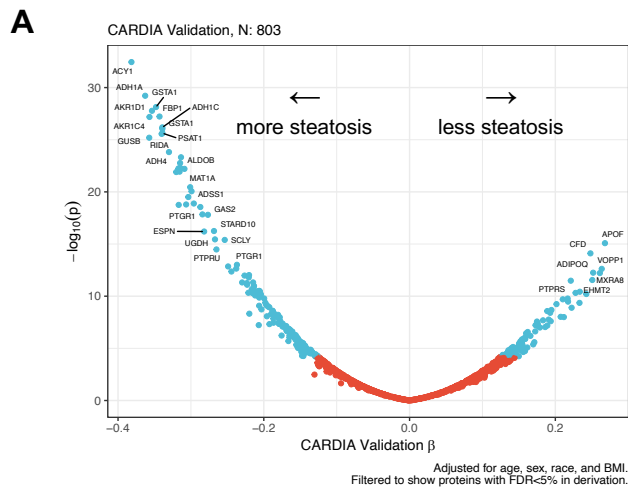


Figure 3: Development of a proteomic score of MASLD and its utility as a diagnostic test.

(A) A protein score of liver attenuation by CT (less attenuation ~ more steatosis) demonstrated moderate correlation with the parent variable in both CARDIA derivation and validation samples.

(B) The protein score distinguishes between MASLD and non-MASLD populations in CARDIA.

(C) Replication of the association between a protein score of liver attenuation and MRI-based measure of hepatic steatosis (proton density fat fraction: higher ~ more steatosis, opposite directionality as with CT based liver attenuation) in UK Biobank. (D) The protein score distinguishes between MASLD and non-MASLD populations in UK Biobank.

(E) The protein score is related to controlled attenuation parameter (higher value ~ more steatosis) in CCHC.

(F) Receiver operator curve analysis in the CARDIA validation sample comparing a clinical model of MASLD (age, sex, race, BMI, drinks/week, AST, ALT, A1c) to models including a

protein score of liver attenuation. (G) ROC analysis in UK Biobank.

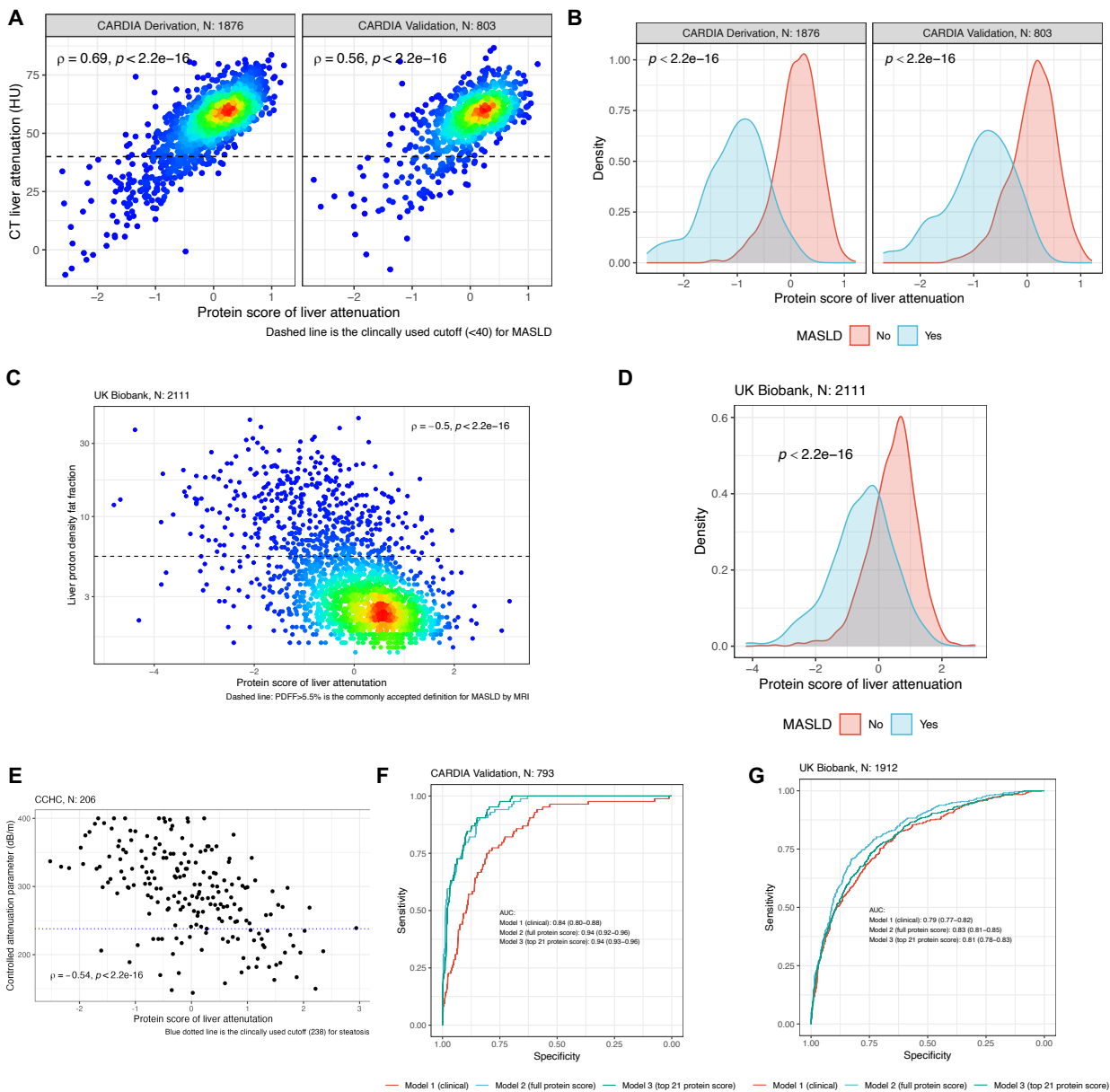
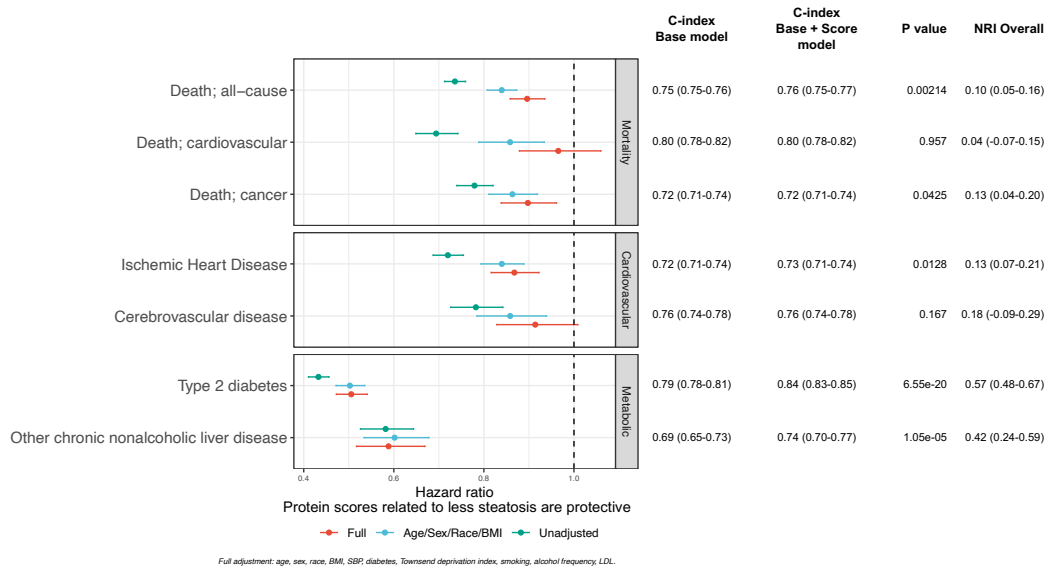
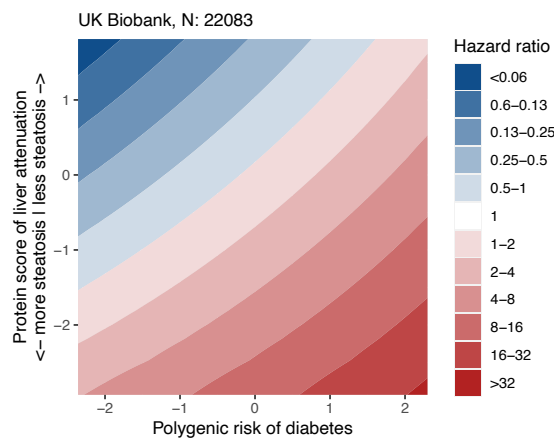


Figure 4: Protein score of liver attenuation is strongly associated with incident metabolic disease and additive to genetic risk. (A) Forest plot of associations with clinical outcomes in UK Biobank along with C-index comparisons of models with and without the protein score (see **Supplemental Data File SD06**). P values reported are for comparisons of C-indices. (B) We observed a weak interaction between the protein score of liver attenuation and polygenic risk for incident diabetes suggesting the effects of both are largely complementary and additive. Hazard ratios presented are from model predicted estimates. (C) Cox regression models using a clinically translatable 21-protein panel provides similar prognostication as the full protein score (see **Supplemental Data File SD06**). The shaded ellipses represent the standard error of the beta coefficient.

A



B



C

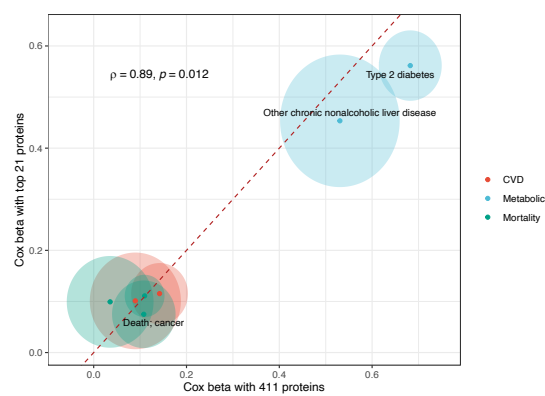


Figure 5: Single nuclear and spatial transcriptional architecture of the circulating MASLD proteome. (A) Uniform manifold approximation projection (UMAP) of single nuclear RNA sequencing in steatotic and healthy liver, colored by a composite expression score (derived from gene expression of implicated target proteins, see **Methods**). (B, C) UMAP of spatial transcriptomic data (Visium) in the liver colored by liver pathology diagnosis (healthy vs. fatty, **B**) and the composite expression score (**C**). (D) Violin plot comparing composite expression score across Visium spots by fatty vs. healthy state (Wilcoxon rank-sum). (E) Representative images of healthy and steatotic hematoxylin-eosin-stained liver tissue overlaid with Visium spots, colored by composite expression score, demonstrating increased activity of implicated targets in steatotic regions. All sections presented in our parent manuscript are shown in **Supplemental Figure 4**. (F) Composite expression across liver zonation groups. (G) Differential expression of implicated targets between healthy and fatty liver (Visium) versus circulating proteomic regression coefficient. A more positive proteomic coefficient indicates less liver fat, and a more positive \log_2 fold-change indicates greater expression of a given transcript in healthy (non-steatotic) liver. Highlighted in purple are targets that were considered as differentially expressed using spatial data (satisfied adjusted p-value < 0.05 & absolute \log_2 fold-change > 0.25). (H) Gene expression of significantly different implicated targets between healthy and steatotic regions (Visium), liver zonation (Visium) and across cell types (single-cell RNA sequencing).

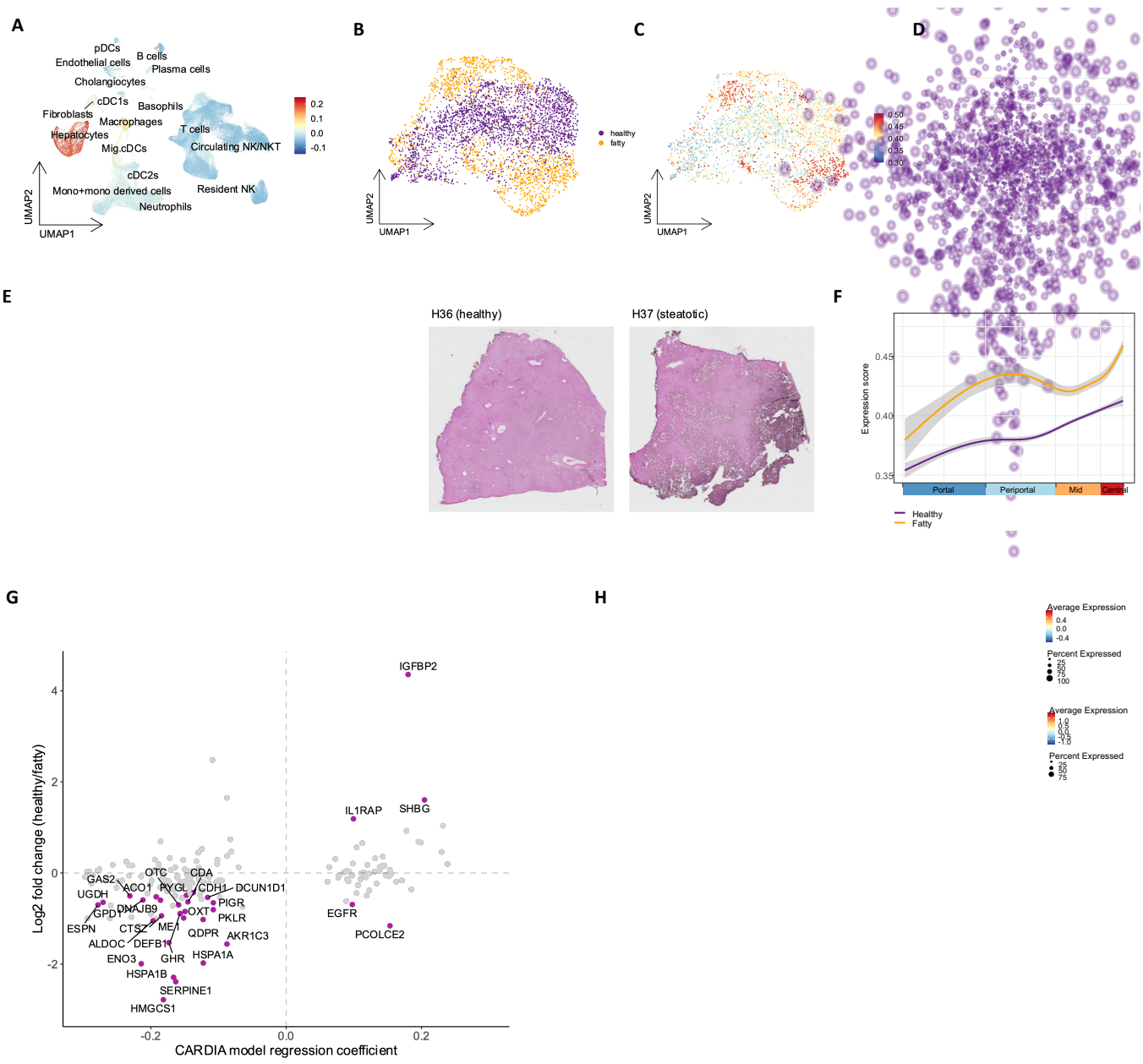


Figure 6: Transcriptional heterogeneity of spatial targets in human liver across steatosis stages. (A) Bulk transcript \log_2 fold change in human liver (over control samples without histologic steatosis) for those genes (among 33 significant on spatial studies) that were significantly differentially expressed in at least one comparison (stage 1 vs. control; stage 2 vs. control; stage 3 vs. control). Of the 33 genes passed forward for assessment in bulk transcriptomics, 12 were not significantly expressed in any of the stages of steatosis (by adjusted p-value < 0.05) and were not included in visualization. The “liver attenuation beta” represents the regression coefficient against liver attenuation in the CARDIA derivation sample. A positive coefficient (red) indicates a greater protein level is related to higher attenuation (lower steatosis); a negative coefficient (blue) indicates a greater protein level is related to lower attenuation (higher steatosis). This analysis excluded individuals with stage F4 fibrosis, given differences in hepatic physiology at this stage of decompensation. (B) Violin plots of example gene expression (in \log_2 counts per million) for genes that displayed a “concordant” directionality between the proteome and the bulk transcriptome (top and middle panel) and “discordant” directionality between proteome and bulk transcriptome.

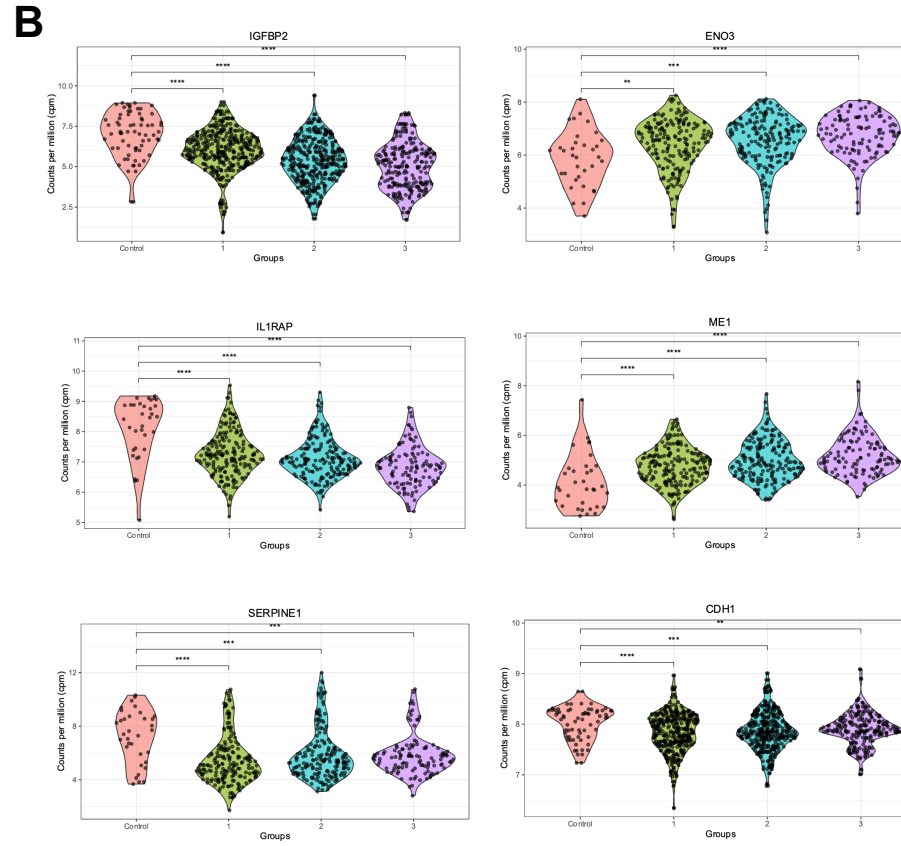
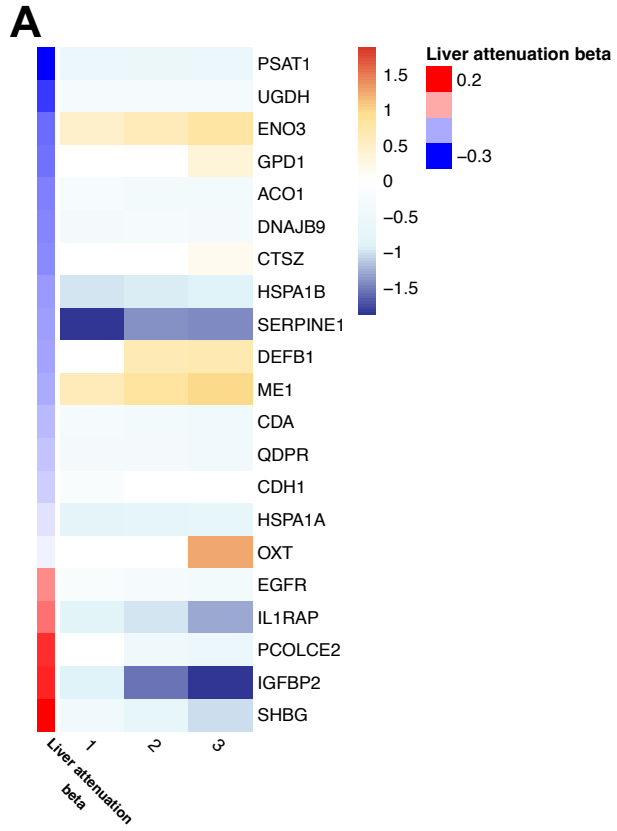
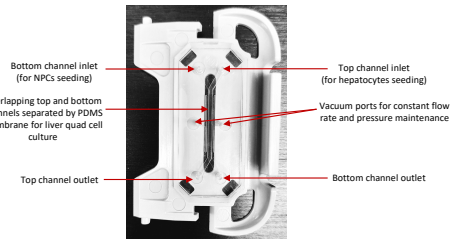
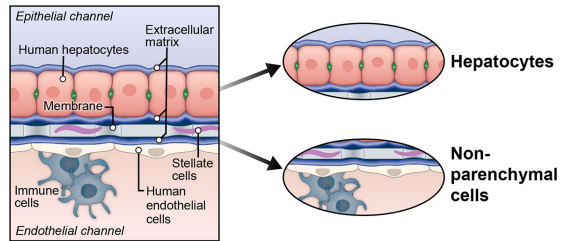
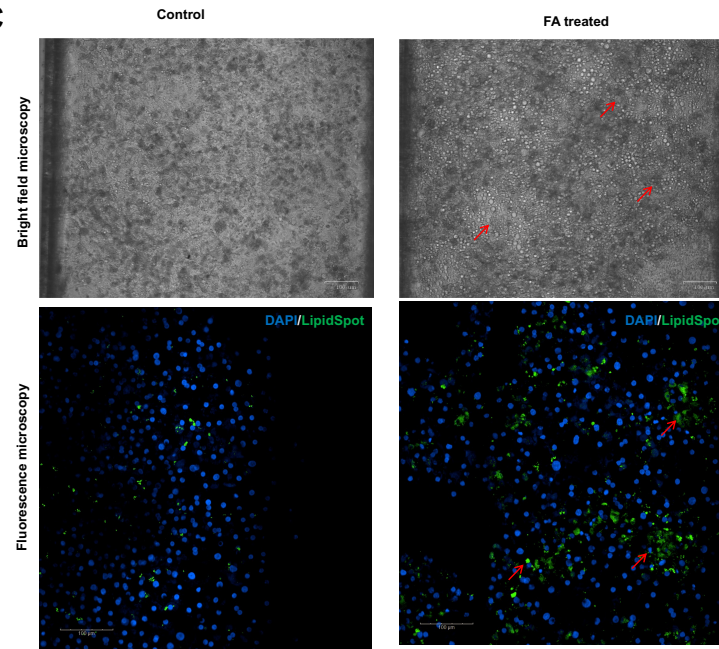
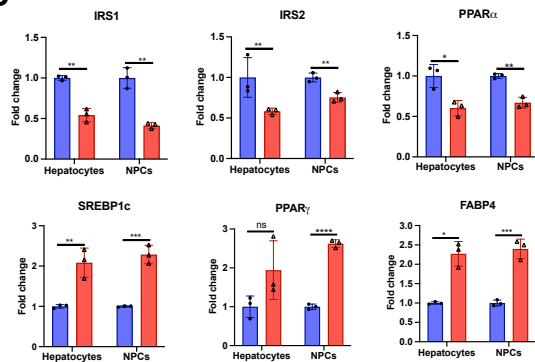


Figure 7. Transcriptional architecture of MASLD on a humanized liver-on-a-chip (LOC) largely replicates population and tissue findings. (A) and (B) show the structure and experimental design of MASLD induction on the LOC. (C) Successful MASLD model generation on a representative LOC. On the left, lipid droplet accumulation was visualized after 5-day treatment period of FAs. Representative brightfield and fluorescent confocal images of the LOC cells (scale bar = 100 μ m). DAPI (nuclear) and LipidSpot (lipid droplet) stains are shown. Red arrows represent lipid droplet accumulation. (D) mRNA expression of canonical genes implicated in steatosis demonstrate expression patterns consistent with MASLD in both hepatocytes and NPCs. A total of 6 chips were included (3 FA and 3 control). Results were analyzed by an unpaired *t* test and expressed as mean \pm standard error of 3 independent experiments. Each data point represents the average of 3 technical replicates. Control is in blue and FA treated is in red. Relative expression is shown as fold change (delta-delta CT) relative to control, normalized to beta-actin expression. (E) mRNA expression of top genes on the LOC that were prioritized by proteomic and transcriptomic studies (see text). Breaks in Y axis are presented given disparate expression of some genes (e.g., HMGCS1 had low expression in non-hepatocytes, while HSPA1A was expressed at low levels in hepatocytes). ME1, CTSZ and DEFB1 were not expressed in hepatocytes; CDA, SHBG, IL1RAP and IGFBP2 were not expressed in NPCs. See text for details. Abbreviations: *ne*, not expressed (raw Ct > 40); *ns*, non-significant; *, $p < 0.05$; **, $p < 0.01$; *, $p < 0.001$; ****, $p < 0.0001$.**

A**B****C****D****E**