Science Translational Medicine
AAAS

Supplementary Materials for

Peptide-based urinary monitoring of fibrotic nonalcoholic steatohepatitis by mass-barcoded activity-based sensors

Sophie C. Cazanave *et al.*

Corresponding author: Gabriel A. Kwong, gkwong@glympsebio.com; Sophie C. Cazanave, scazanave@glympsebio.com

Sci. Transl. Med. **13**, eabe8939 (2021)
DOI: 10.1126/scitranslmed.abe8939

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

MATERIALS AND METHODS

nCounter custom design, hybridization, data collection and analysis. A custom 800-member NanoString nCounter gene expression Codeset (Warren01_C4724), targeting 229 secreted and membrane-bound endo-proteases from the human genome and 500 NASH-related genes implicated in metabolic, inflammatory and fibrotic pathways, was designed under contract by NanoString Technologies using the Homo sapiens genome assembly (GRCh38.p12) and the Homo sapiens Annotation Release 109. 71 gene normalizers, including positive normalizers, were included in the gene set. Hybridization reactions were performed in sets of 12 samples per run, according to the manufacturer's instruction, on a total of 36 nCounter ChIP-String Assays. Briefly, hybridization buffer and RNAse-free water were added to specific-Reporter CodeSet reagent and this master mix was aliquoted into PCR reaction tubes. RNA was then added to each tube, followed by a Capture ProbeSet reagent. Samples were mixed and then added to each tube and incubated for 16 h at 65°C in a thermal cycler. Each set was then stored at 4°C until use on an nCounter cartridge and run in an nCounter SPRINT profiler for data collection at the Whitehead Institute for Biomedical Research (Cambridge, MA). nCounter data were processed using nSolverAnalysisSoftware 3.0 (NanoString, Seattle, WA, USA). After quality control checks on individual RCC files, raw counts across samples were normalized to the geometric mean counts of synthetic DNA positive controls included in the hybridization reactions to mitigate platform-associated sources of variation. No background subtraction or thresholding was performed at this stage. Reference genes were selected using the geNorm algorithm within the nCounter Advanced Analysis (nCAA) module (version 2.0.115, NanoString). Five genes were used to normalize the data: SDHA, CLTC, TUBB, PGK1, and GUSB. For each sample, normalization was performed by dividing counts for each gene by the geometric mean of the counts for the five reference genes. nCAA was used to calculate the differential gene expression with a significance determined by a Bonferroni-corrected ANOVA adjusted p-value ≤ 0.05. Bonferroni correction factor was k= 229 based on the total number of secreted and membrane-bound proteases under consideration.

RNA-seq data. RNA-seq analysis were performed by BGI (Cambridge, MA). Total RNAs were extracted by BGI from RNAlater-frozen mouse liver samples and RNA quality and concentration was measured using an Agilent 2100 Bioanalyser. RNA Integrity Number cutoff was ≥ 8.5. Total RNA samples are first treated with DNase I to degrade any possible DNA contamination; and mRNA is enriched by using the oligo (dT) magnetic beads. Mixed with the fragmentation buffer, the mRNA is fragmented into short fragments. Then the first strand of cDNA is synthesized by using random hexamer-primer. Buffer, dNTPs, RNase H and DNA polymerase I are added to synthesize the second strand. The double strand cDNA is purified with magnetic beads. End reparation and 3'-end single nucleotide A (adenine) addition is then performed. Finally, sequencing adaptors are ligated to the fragments. The fragments are enriched by PCR amplification. During the QC step, Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR System are used to qualify and quantify of the sample library. The library products are ready for sequencing (50 bp, single reads and 7 bases index read) via Illumina HiSeqTM 2000 instrument. We also combined our in-house RNA-seq data with publicly available NAFL datasets (GSE138945 and GSE138946) from C57BL/6 male mice kept on the same HFD for 20 weeks (n= 3) using the same sequencing platform (*42*). Differential expression was calculated between Control ASO HFD 20W group (GSE138945) and Saline Healthy CD group (GSE138946).

Recombinant protease screening. Fluorogenic substrates were assayed with a 7-point dilution, from 6 uM to 0.09 uM, with 3 protease concentrations – 500 nM, 50 nM, and 5 nM. The substrate plate was prepared on a 384-well black Greiner microclear assay plate, with 25 uL of substrate per well. A baseline read was performed using the SpectraMax i3 kinetic plate reader, at 535 nm with a 485 nm excitation. The plate was immediately spun down at 200xg for 30 seconds and read on the SpectraMax i3 for a kinetic read of 90 minutes.

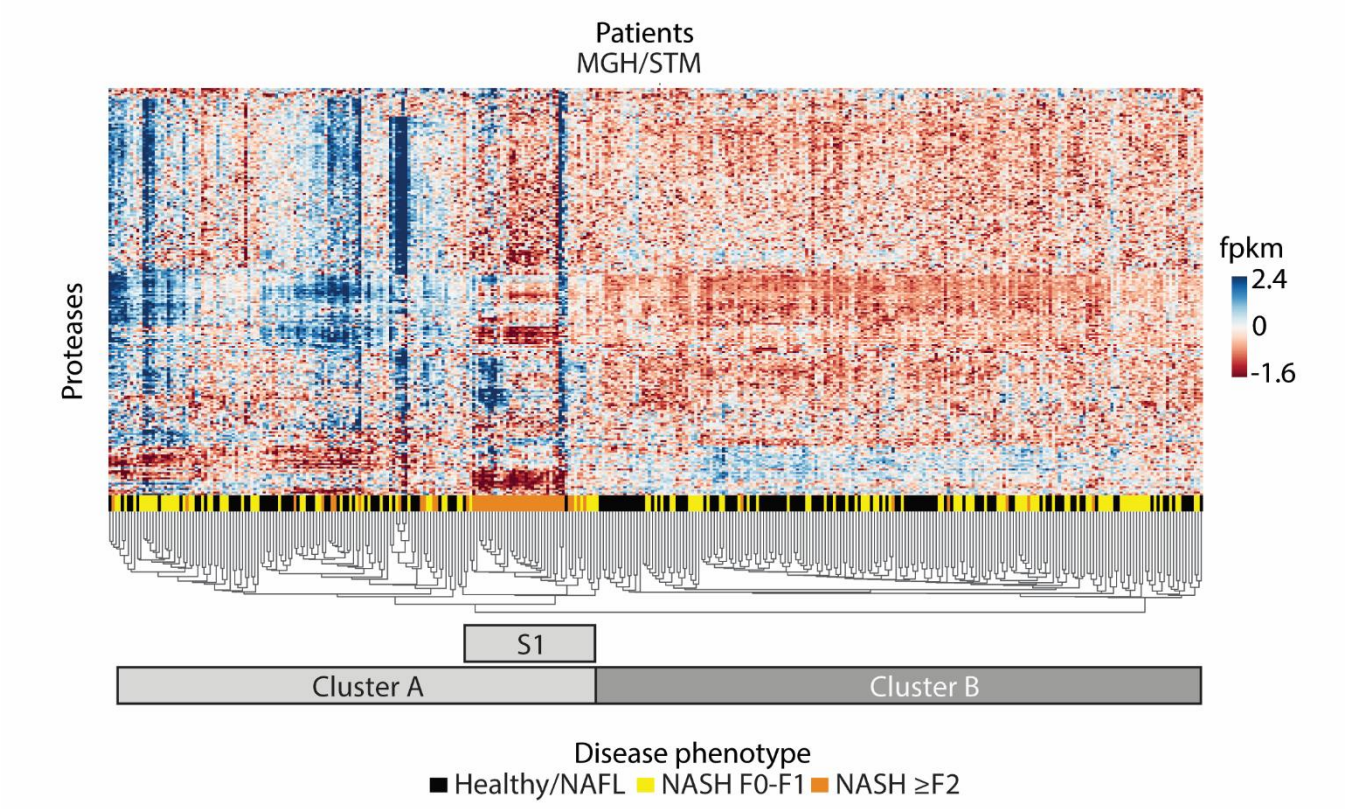
Blood biomarkers. Blood was collected in EDTA-coated plasma tubes (BD vacutainer) and spun at 1,000 × g for 20 min to pellet blood cells. Triglyceride, cholesterol, ALT and AST plasma levels were measured by Charles River Clinical Pathology Laboratories (Shrewsbury, MA). Plasma HA concentrations were determined in duplicate using a commercially available HA Test Kit (Corgenix, Inc., Broomfield, CO, Cat# 029-001). Plasma PIIINP concentrations were determined in duplicate using a commercially available rat Procollagen III N-Terminal Propeptide (PIIINP) ELISA Kit (Biomatik, Wilmington, DE, Cat# EKU06788).

Histology analysis. Liver and kidney tissues were fixed in 10% neutral-buffered formalin for at least 72 h and then stored into 70% ethanol prior to embedding into paraffin. 5-µm sections were processed for staining with hematoxylin and eosin (H&E), Masson's trichrome (MT) and Picro Sirius Red (PSR) by Mass Histology (Worcester, MA). CD68 (Abcam ab31630 [ED1], Ms mAb, 2ug/mL) and αSMA (Biocare CM001 [1A1], Ms mAb, 0.12ug/mL) IHC stainings were performed on deparaffinated liver slides using DAB peroxidase substrate. Quantitative morphometry of PSR, CD68 and αSMA stainings were performed on the entire liver section and expressed as percent PSR, CD68 and αSMA-positive area of total parenchymal area. All liver slides were scored by the blinded review of a certified pathologist (Pierre Bedossa, Liverpat, Paris, France) according to the NASH CRN criteria and disease was classified following the FLIP algorithm (*64*). Following this algorithm definition, NASH was characterized by scores for steatosis, lobular inflammation and hepatocyte ballooning concomitantly equal or higher than 1. NAFL mice were characterized by score for steatosis equal or higher than 1 without the concomitant presence of lobular inflammation or hepatocyte ballooning. Tubular degenerative histology change was scored as Normal, Mild, Moderate or Severe when present in less than 25, 25, 25-50 or more than 50%, respectively, of the whole-slide kidney section. Digital images of H&E, MT and PSR were captured using a Pannoramic 250 Flash III digital slide scanner (3DHISTECH Ltd.).

LC-MS/MS reporter analysis. Liquid chromatography/tandem mass spectrometry analysis was performed at Syneos Health- Inventiv Health using a SCIEX 6500 triple quadrupole instrument. Briefly, urine was collected and stored at -80°C. Photolysis of the photolabile group 3-Amino-3-(2-nitro-phenyl)propionic Acid (ANP) linker was performed under UV light for 2h using a CL-1000 ultraviolet crosslinker (UVP Inc.) to yield the Glu-Fib reporter from residual peptide fragments. Samples were extracted by solid phase extraction and analyzed by multiple reaction monitoring by LC-MS/MS to quantify the concentration of each 19 Glub-Fib mass barcode. An internal control reporter (PC) was spiked into the urine prior to analysis. Reporter concentrations were calculated from a standard curve using peak area ratio (PAR) to the internal standard. PAR values were mean normalized across all reporters in a given urine sample prior to further statistical analysis to account for mouse-to-mouse differences in urine concentration.

GTBS-NASH Logistic Regression Classifiers: FiND, FiNR, GBTS-NASH ≥F2 and FiND_BTBR classifiers were built based on animal urine MS signals after administration of GBTS-NASH and inputted into a regularized logistic regression. Classifier performance was evaluated with 100 rounds of cross validation (randomized 80% train and 20% validation splits). More specifically, FiND classifier was trained and tested on 48 healthy (CD 9 weeks) animals and 48 NASH (CDAHFD 9 weeks). The output probability of having NASH was calculated based on the following equation: S (-0.71 - 0.09*N18 + 0.41*N11 - 0.47*N17 + 0.03*N09 - 0.87*N16 - 0.40*N14 - 0.06*N01 + 0.09*N19 + 0.38*N07 - 0.16*N10 - 0.06*N04 - 0.06*N05 - 0.32*N06 + 0.12*N08 + 0.20*N13 + 0.11*N12 - 0.34*N03 - 0.05*N02 - 0.24*N15). NAFL cohorts (HFD 16 Weeks, n= 15) or additional cohorts of CD and CDAHFD mice (AM, PM, R1, R2, Fed, Fasted, Vehicle, FA cohorts, n= 14-20 per group and per cohort) were all applied naively to the FiND classifier. FiNR classifier was built on 32 NASH baseline 9 Weeks and 15 Early Regression animals (9W+1W) (80/20 split) or on 32 NASH baseline 9 Weeks and 17 Late Regression (9W+3W). The output probability of NASH regression was calculated based on the following equation: S (1.85 + 0.13*N18 - 0.01*N11 + 0.13*N17 + 1.03*N09 - 0.26*N16 + 0.02*N14 - 0.42*N01 + 0.44*N19 + 1.31*N07 - 0.34*N10 - 0.21*N04 - 0.13*N05 - 0.13*N06 + 0.01*N08 + 0.12*N13

+ 1.02*N12 - 0.21*N03 - 0.42*N02 - 0.68*N15). GBTS-NASH ≥F2 classifier was trained and tested on rat cohorts ii (n= 10, NASH 6W), iii (n= 15, NASH 6W + vehicle 6W) and iv (n= 15, NASH 6W + TRIPLE 6W) for which histology scoring was available. The output probability of having ≥F2 was calculated based on the following equation: S (-0.07 + 0.29*N17 + 0.30*N18 - 0.39*N14 - 0.38*N03 + 0.06*N02 + 0.13*N15 + 1.62*N11 + 0.02*N19 - 0.46*N07 + 0.11*N10 + 0.36*N01 - 0.06*N04 + 0.49*N05 - 0.43*N06 + 0.25*N08 + 0.03*N09 - 0.85*N16 + 0.10*N13 - 0.25*N12). Urine samples from these same animals at 7, 8 and 10 weeks or chow diet group were then applied naively into the ≥F2 classifier. FiND_BTBR classifier was trained and tested on BTBR WT CD (n= 20) and CDAHFD 9 Weeks (n= 19) (80/20 split). The output probability of having NASH was calculated based on the following equation: S (-1.53 - 0.39*N18 + 0.87*N11 - 0.62*N16 + 0.07*N14 - 0.01*N01 + 0.51*N19 + 0.63*N07 - 0.02*N10 + 0.05*N04 - 0.01*N05 - 0.15*N06 + 0.83*N08 - 0.96*N13 + 0.22*N12 - 0.59*N03 - 0.07*N02 - 0.67*N15). Urine samples from BTBR ob/ob CD 9 weeks (n= 19) mice were applied naively into that classifier. For all these classifiers, S is a sigmoid function, relative reporter concentrations are L1-norm, then Log d and finally *z*-scored. All coefficients were rounded to the second decimal point.



SUPPLEMENTAL FIGURES

Fig. S1. Unsupervised hierarchical clustering using NanoString normalized count data for 229 protease genes in MGH/STM cohort. Color assignments correspond to *z*-scores per gene that were computed from normalized count data. The distance map based on the Nanostring data indicates the distribution of clinicopathological features. Two distinct clusters, A and B, were observed. Cluster A displayed a subcluster (S1) of ≥F2 patients with NASH.

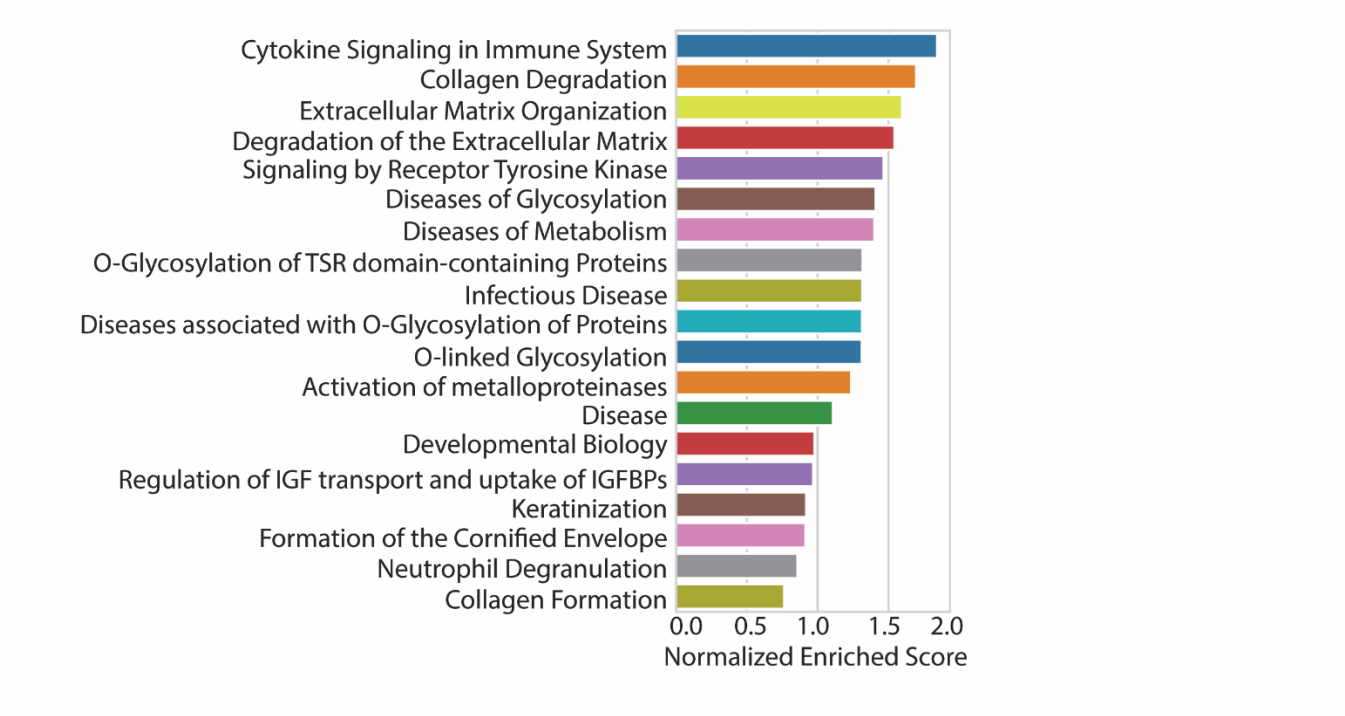


Fig. S2. Non-protease genes related to inflammation, metabolism and fibrosis pathways are dysregulated in ≥F2 *vs* F0-F1. NanoString transcriptome analysis for 570 non-protease genes from the combined Massachusetts General Hospital (MGH) and St. Mary's weight loss surgery clinic (STM) NAFLD cohort. Integrated pathway enrichment analysis (*z*-score or normalized enriched score) of differentially expressed, non-protease genes in ≥F2 *vs* F0-F1 patients was performed according to the pre-ranked algorithm (min size= 10, max size= 500) according to GSEA 4.1.0 with the Reactome v7.2 pathway database.

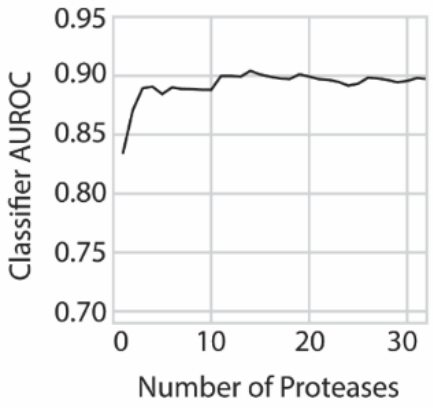


Fig.S3. Fifteen NASH-upregulated proteases are sufficient to classify ≥F2 *vs* F0-F1 with high accuracy. AUC for classifiers trained with an increasing number of proteases (from top 1 to 32, ranked by ANOVA) within our list of upregulated proteases in NASH.

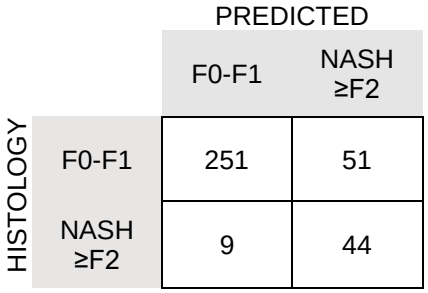


Fig. S4. Performance of the 13-NASH gene classifier using the MGH/STM cohort with fraction of correctly assigned histological score, false positive and false negative predictions. Confusion matrix showing correct classification of 295/355 (83.1%) patients as determined by histological scores using the binary classification F0–F1 *vs* ≥F2. Specificity and sensitivity were 83.1% and 83.0%, respectively.

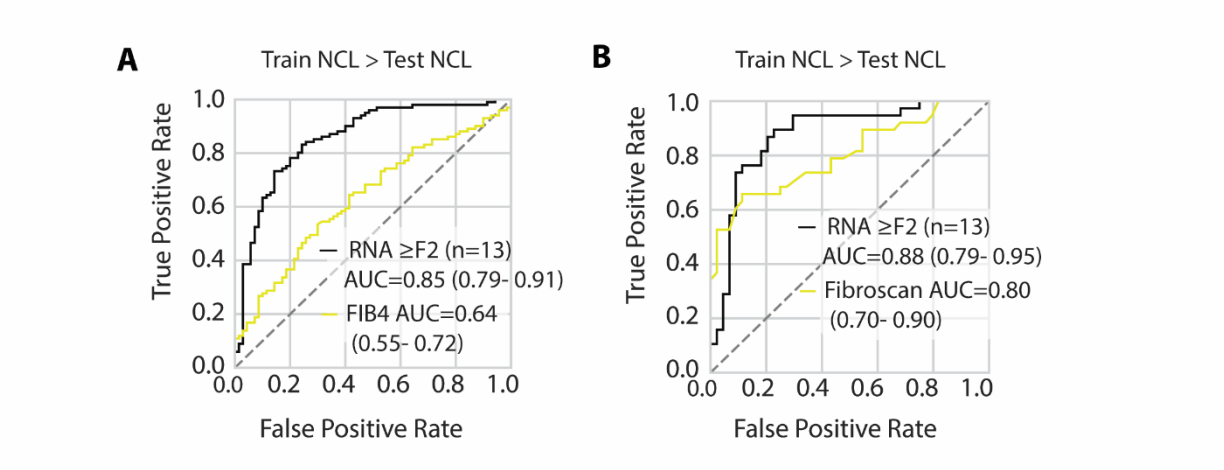


Fig. S5. RNA ≥F2 gene classifier comparison to FIB4 and liver stiffness (Fibroscan) to predict ≥F2 *vs* F0-F1 using the Newcastle NASH cohort. (A, B) RNA from 146 liver tissue samples from a Newcastle cohort of NAFLD (NCL) was analyzed with a custom NanoString panel. Area under the receiver operating characteristic (AUROC) for binary classification of fibrosis stage ≥F2 versus F0–F1 with a regularized logistic regression classifier trained either with RNA ≥F2 gene classifier (comprising a subset of 13 NASH protease gene counts) or calculated FIB4 score (**A**) or liver stiffness values (Fibroscan) (**B**) in randomized 80% training and 20% validation sets. For FIB4 classifier, eight of the patients did not have platelet measurements necessary to calculate the score and were excluded from this analysis. For Fibroscan classifier, data values were available for only 82 patients (19 NAFL and 63 NASH: 25 F0-1, 10 F2, 25 F3 and 3 F4); and RNA ≥F2 was retrained in the same patients for comparison.

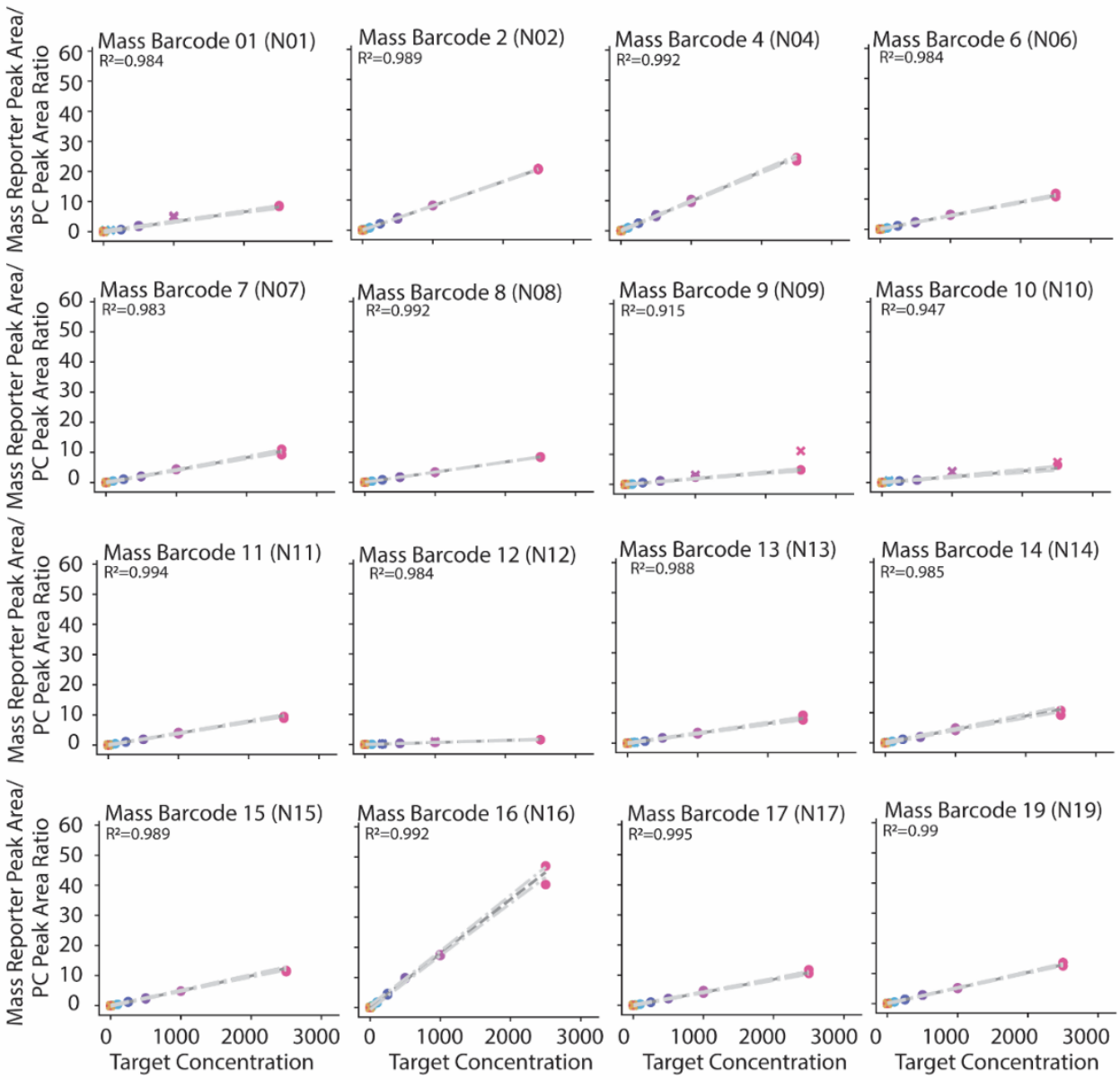


Fig. S6. Quality Control for GBTS-based mass reporters by mass spectrometry. Calibration curves from 16 GBTS mass-barcoded protease substrates upon UV cleavage to release mass barcode measured by LC-MS/MS. Peak area from each mass barcode is divided by peak area of spiked-in isotopically-labelled internal standard (PC).

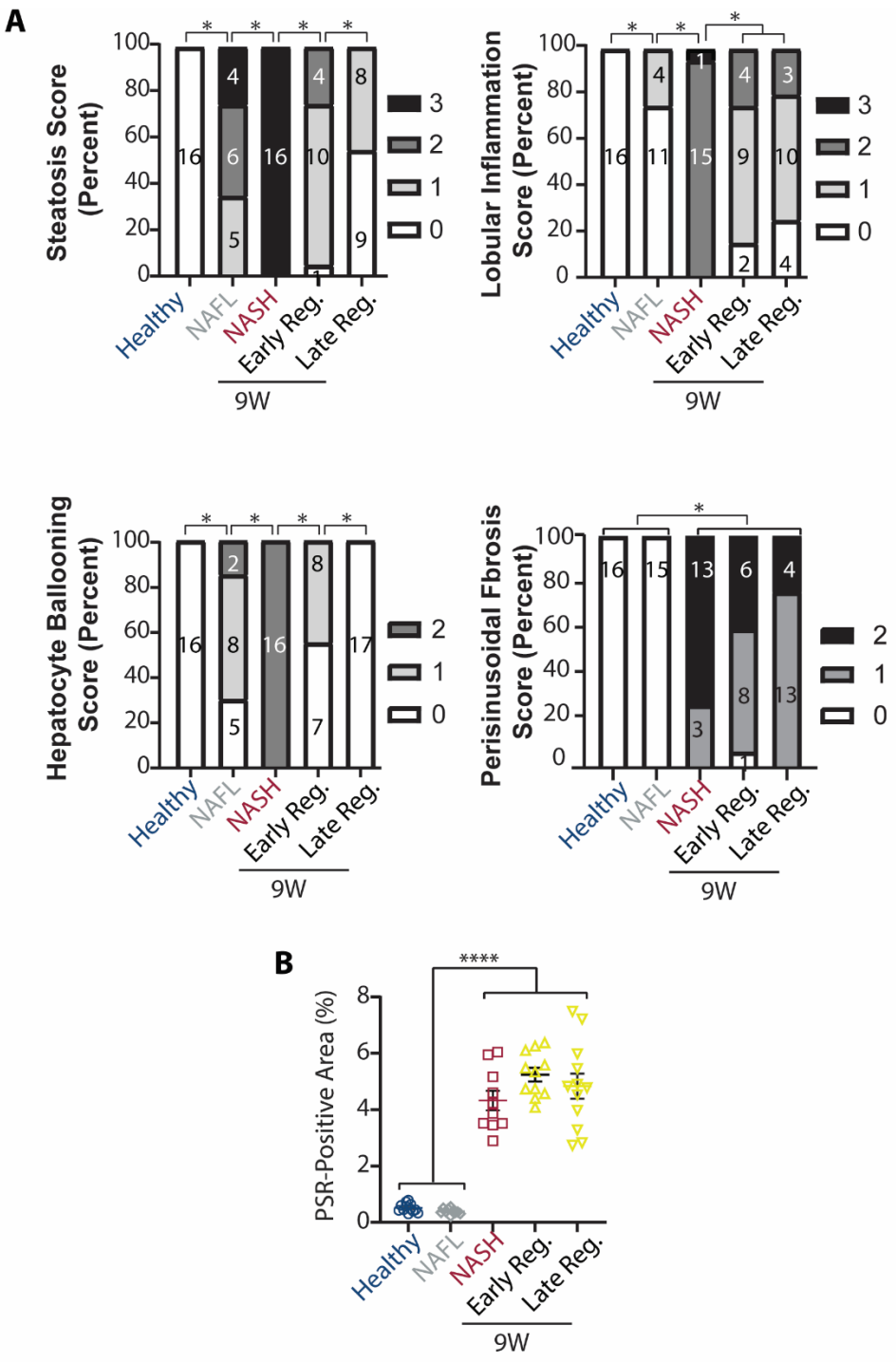


Fig. S7: Histological scores during NASH progression and diet regression. (**A**) Steatosis, lobular inflammation, hepatocyte ballooning, fibrosis and perisinusoidal fibrosis scores were measured in 16 healthy (CD 9 weeks), 15 NAFL (HFD 16 weeks), 16 NASH (CDAHFD 9 weeks), 15 early regression ((NASH 9 weeks + CD 1 week) and 17 late regression (CDAHFD 9 weeks + CD 3 weeks). The percentage of animals per group with a specific assigned NAS or fibrosis score are displayed. The number of animals for each scoring category is also shown within bars. **Chi Square p≤0.05*. (**B**) Quantitative morphometry of PSR staining was performed on the entire liver section of 12 healthy, 10 NAFL, 10 NASH, 11 early regression and 12 late regression and expressed as percent PSR-positive area of total parenchymal area. *****p≤0.0001*.

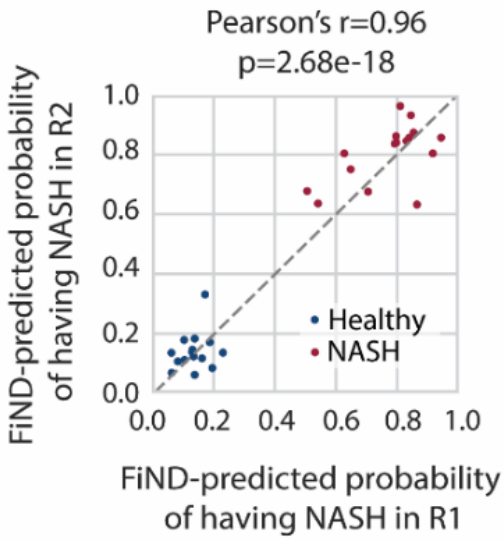


Fig. S8: Mouse reproducibility of FiND-predicted output probability of being NASH. Pearson correlation analysis of FiND-predicted output probability of being NASH between NASH F1-F2 (n= 16) and healthy (n= 15) mice injected with GBTS-NASH at week 9 (R1) and 10 (R2).

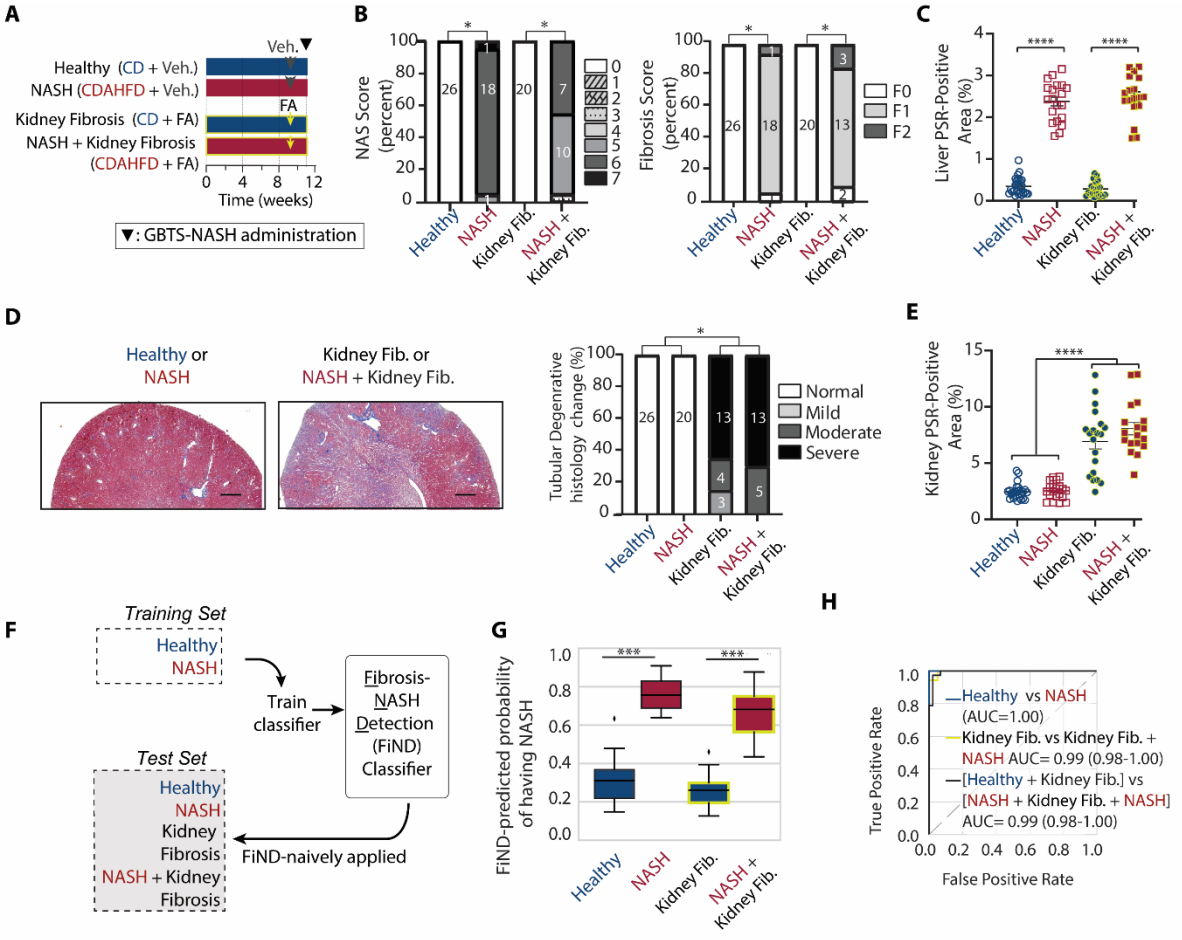


Fig. S9: Acute kidney injury (AKI) does not affect FiND classifier performance. (**A**) Acute kidney injury was modeled using single dose i.p. administration of folic acid (FA; 250mg/kg body weight) to healthy (Kidney Fibrosis, CD + FA, n= 20) or NASH (NASH + Kidney Fibrosis, CDAHFD + FA, n= 18) mice at 9 weeks. Additional cohorts of healthy (Healthy, CD + Veh., n= 25) and NASH (NASH, CDAHFD + Veh., n= 20) mice at 9 weeks receiving vehicle (0.3 mM Sodium bicarbonate) were also included. Two weeks later, all mice were injected with GBTS-NASH (1 nmole/peptide) and plasma, kidney and liver tissues were collected. (**B**) Liver histology NAS and fibrosis scores. The percentage of animals per group with a specific assigned NAS or fibrosis score are displayed. The number of animals for each scoring category is also shown within bars. *Chi Square *p≤0.05.* (**C**) Quantitative morphometry of PSR staining was performed on the entire liver section and expressed as percent PSR-positive area of total parenchymal area. (**D**) Representative histology picture from Masson's trichrome-stained kidney slides. Scale bar is 0.5 mm. Tubular degenerative histology change was scored as Normal, Mild, Moderate or Severe when absent or present in less than 25, 25-50 or more than 50%, respectively, of the whole-slide kidney section. The percentage of animal per group with a specific assigned score are displayed. The number of animals for each scoring category is also shown within bars, **Chi Square p≤0.05.* (**E**) Quantitative morphometry of PSR staining was performed on the entire kidney section and expressed as percent PSR-positive area of total parenchymal area. (**F**) MS urine outputs from Healthy, NASH, Kidney Fibrosis and NASH + Kidney Fibrosis were applied naively to the FiND classifier (training on 48 CD and 48 CDAHFD 9 weeks). (**G**) FiND-predicted probability of having NASH. (**H**) AUROC curves for Healthy vs NASH, Kidney Fibrosis vs NASH + Kidney Fibrosis and combined

Healthy and Kidney Fibrosis vs NASH and NASH + Kidney Fibrosis animals; ***p≤0.01, ***p≤0.001, ****p≤0.0001*.

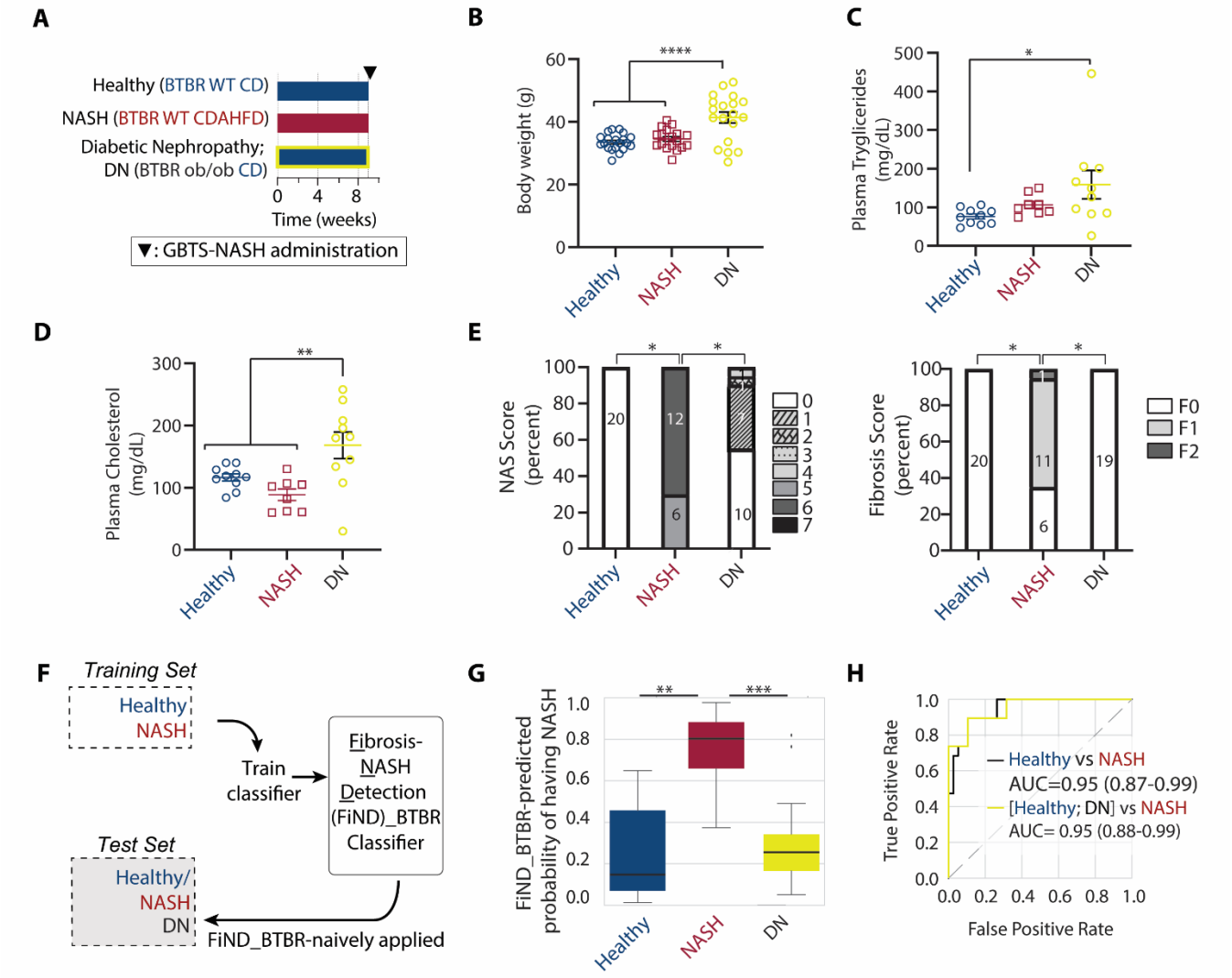


Fig. S10: FiND classifier trained on a BTBR mouse model of NASH was robust with respect to obesity, T2DM, and diabetic nephropathy. (**A**) BTBR WT mice were fed a CD (Healthy, n= 20) or a CDAHFD (NASH, n= 19) for 9 weeks. Age-matched BTBR *ob/ob* (Diabetic Nephropathy, DN, n= 19) male mice were fed a CD for 9 weeks. All mice were injected with GBTS-NASH (1 nmole/peptide). (**B**) Body weight in grams. (**C-D**) Plasma triglyceride and cholesterol in mg/dL, n= 9-10 per group (**E**) Histology NAS and fibrosis scores. The percentage of animals per group with a specific assigned NAS or fibrosis score are displayed. The number of animals for each scoring category is also shown within bars. *Chi Square *p≤0.05*. (**F**) FiND_BTBR machine learning classifier was generated as followed: mass barcode urinary concentrations measured by LC-MS/MS from Healthy and NASH BTBR WT mice were used as input to a machine learning algorithm to generate a disease classifier (logistic regression, 80/20 cross validation). Mass barcode urinary concentrations from Diabetic Nephropathy animals (BTBR ob/ob CD) were applied only into the test cohort. (**G**) FiND_BTBR predicted-output probability of having NASH. (**H**) AUROC curves for Healthy vs NASH and combined Healthy and Diabetic Nephropathy vs NASH; **p≤0.05, **p≤0.01, ****p≤0.0001*.

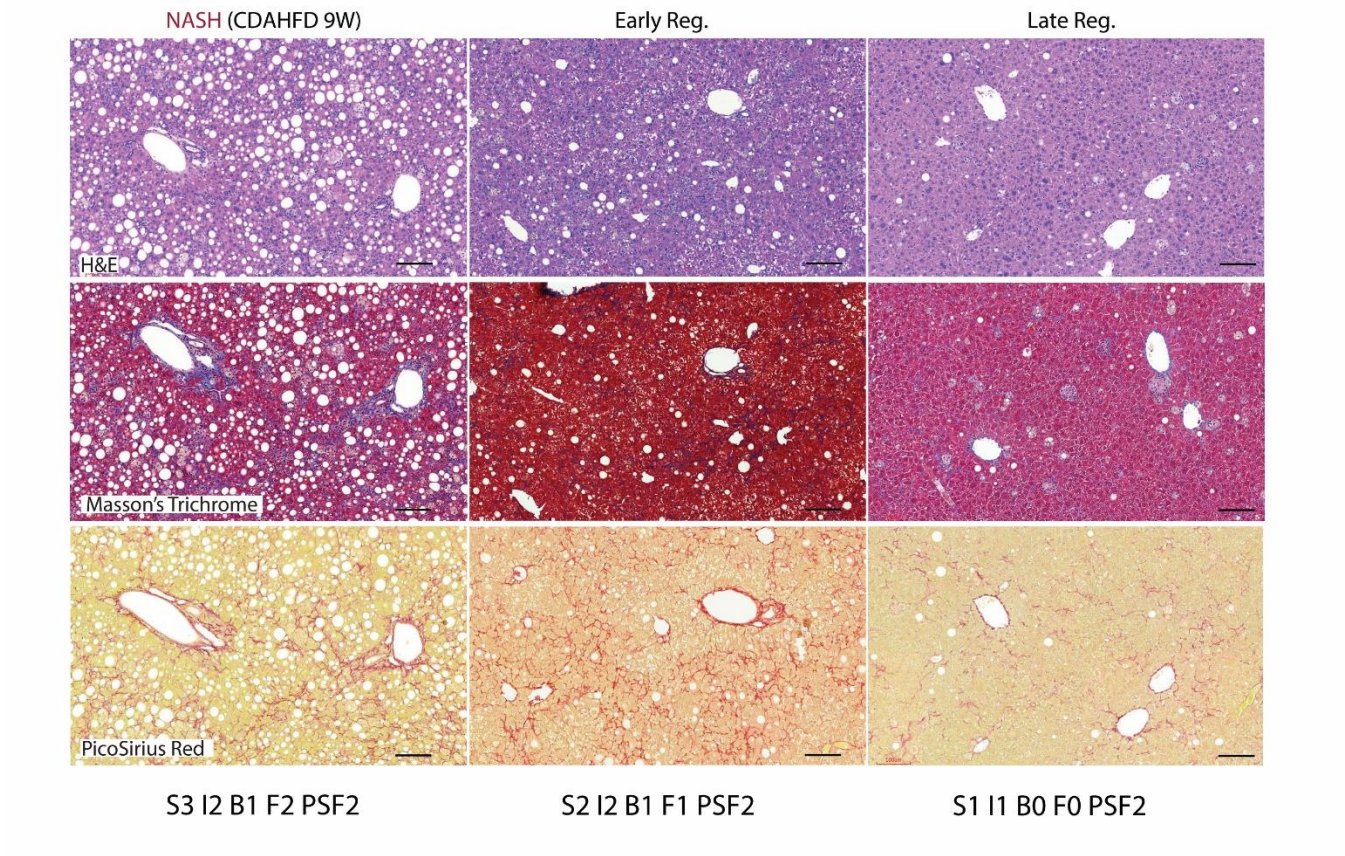


Fig. S11: Liver histology in diet-induced early and late regression. Representative histology images from H/E, Masson's trichrome and PicroSirius Red stained liver slides from NASH (CDAHFD) 9 weeks or in Early or Late regression by switching CDAHFD mice at 9 weeks (NASH) to a CD for 1 or 3 weeks. Liver histology scores are depicted at the bottom with S for Steatosis, I for Lobular Inflammation, B for Hepatocyte Ballooning, F for Fibrosis and PSF for PeriSinusoidal Fibrosis. Scale bar, 200 µM.

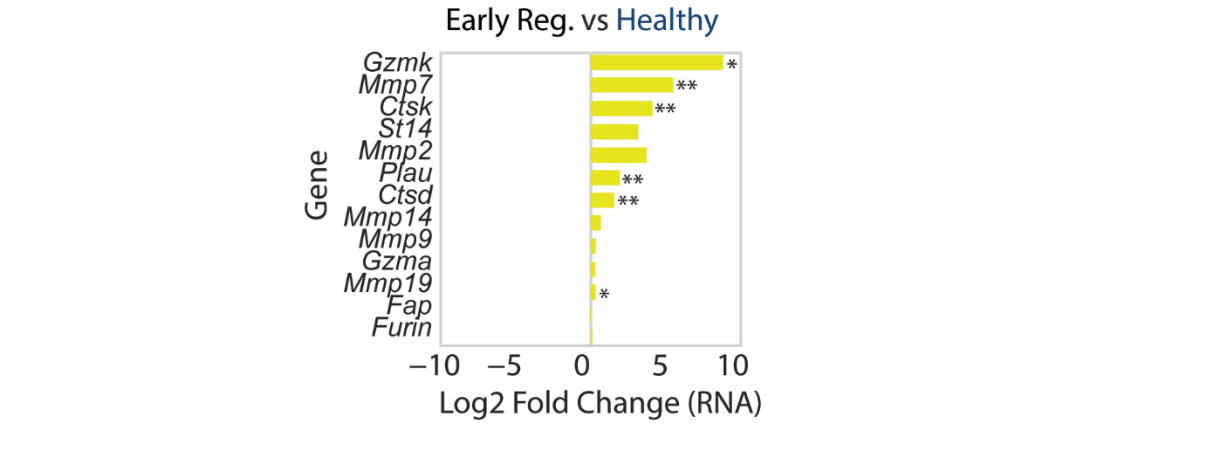


Fig. S12: NASH protease gene expression comparison in early regression and healthy controls. Differential expression (Log_2 Fold Change) from RNA-seq data from early diet-regression animals (NASH 9 weeks switched to a CD for 1 week) and healthy CD controls (n= 4 per group).

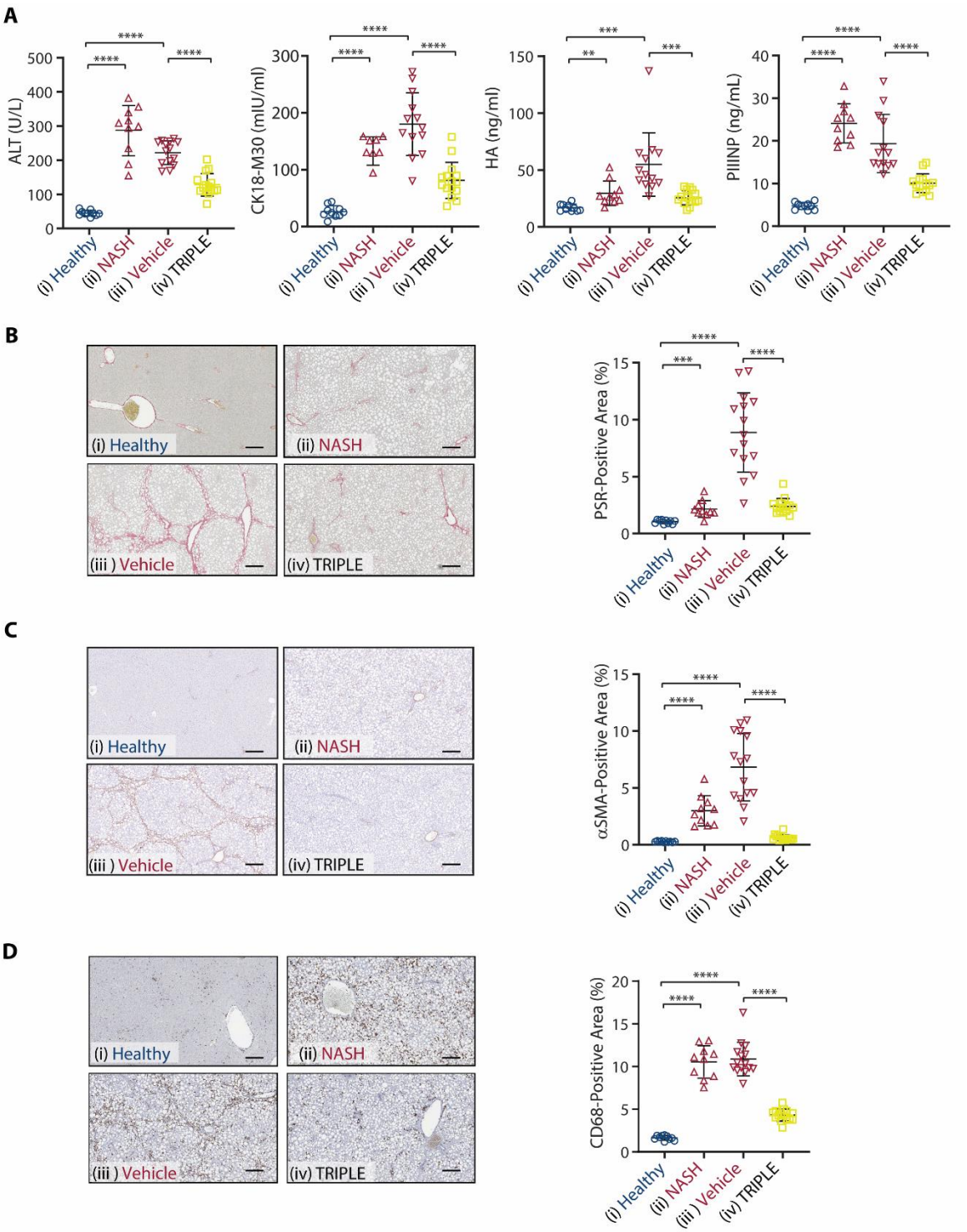


Fig. S13: TRIPLE protects from inflammation and advanced fibrosis in rats fed a CDAHFD. Plasma and liver tissues were collected from male Wistar Han rats healthy 12 weeks (i, n= 10), CDAHFD 6 weeks (ii, NASH, n= 10), CDAHFD 12 weeks receiving either vehicle (iii, Vehicle, n= 15) or TRIPLE treatment combination (iv, TRIPLE, n= 15) from weeks 6-12. (**A**) Plasma ALT in U/L, Cytokeratin (CK)18-M30 in mU/mL, Hyaluronic Acid (HA) and PIIINP in ng/mL. (**B-D**) *left*: Representative histology images and *Right*: quantitation expressed as percent positive area of total parenchymal area from PicroSirius Red (PSR, **B**) and α-smooth muscle actin (αSMA, **C**) and CD68 (**D**) stained liver slides from each cohort of rats. Scale bar, 100 µM; ***p≤0.01, ****p≤0.001*, *****p≤0.0001*.

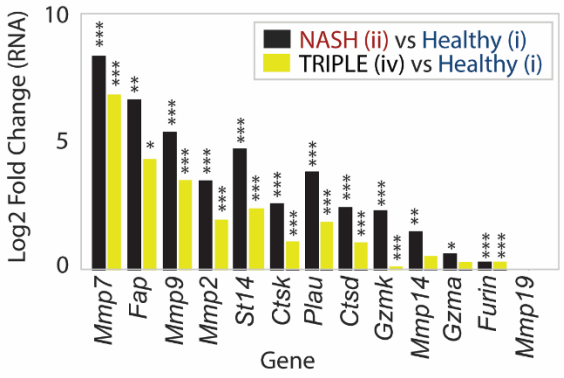


Fig. S14: NASH protease gene expression is upregulated in rat NASH and is decreased upon TRIPLE treatment. Differential expression (Log_2 Fold Change) from RNA-seq data from rats fed a CDAHFD for 6 weeks (NASH, cohort ii) when compared to healthy CD rats at 12 weeks (cohort i) (black bar), or from TRIPLE-treated rats at 12 weeks (cohort iv) when compared to healthy (green bar).

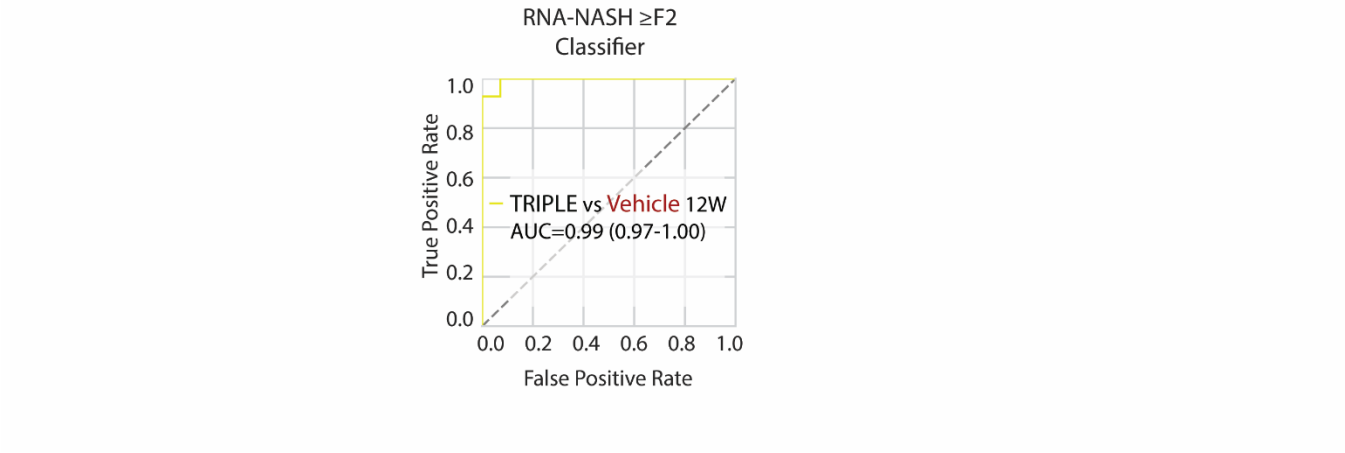


Fig. S15. Training a classifier using biopsy-derived NASH protease gene expression data allowed for discrimination of TRIPLETRIPLE-treated animals. RNA-seq was performed on liver RNA extracted from male Wistar Han rats from healthy (CD 12 weeks, n= 10), NASH (CDAHFD 6 weeks, n= 10), Vehicle (CDAHFD 6 weeks+ Vehicle 6 weeks, n= 15) or TRIPLE (CDAHFD 6 weeks+ TRIPLE 6 weeks, n= 15) groups. RNA-NASH ≥F2 machine learning classifier: NASH protease genes expression (fpkm) was used as input to machine learning algorithms (logistic regression, 80/20 cross validation) trained on endpoint liver and urine samples, respectively and relative to histology fibrosis score. The classifier efficiently discriminated regressed (TRIPLE) *vs* stable/progressing (vehicle) rats at 12 weeks.

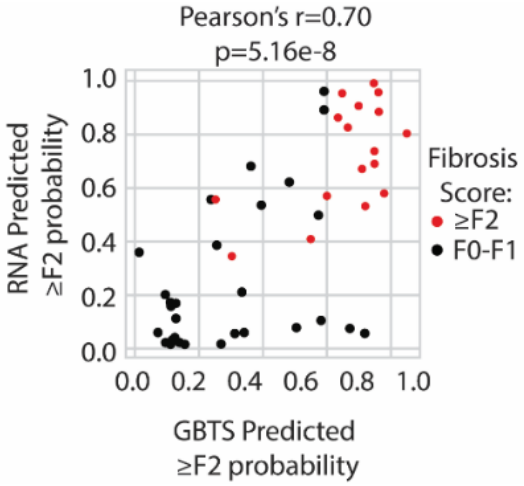


Fig. S16. Direct correlation between GBTS- and RNA-predicted output ≥F2 probability in rat NASH model. Pearson correlation analysis between RNA- and GBTS- predicted output probability of being ≥F2. NASH protease gene expression (fpkm) measured by RNA-seq or mass barcode urinary concentrations measured by LC-MS/MS were used as input to machine learning algorithms (logistic regression, 80/20 cross validation) trained on endpoint liver and urine samples, respectively (Healthy, n= 10; NASH, n= 10; Vehicle, n= 15; TRIPLE, n= 15) and relative to histology fibrosis score.

SUPPLEMENTAL TABLES

Parameters	Mean ± SEM (Min-Max)	Mean ± SEM (Min-Max)	*p* value
Age	44.50 ± 0.65 (16-74)	55.71 ± 0.92 (25-75)	9.14e-19
BMI	46.08 ± 0.40 (26-69)	34.06 ± 0.50 (22.79-58.82)	2.69e-45
ALT (UI/L)	40.08 ± 1.63 (8-288)	68.45 ± 4.20 (16.00-328.00)	8.43e-19
AST (UI/L	27.15 ± 1.16 (0-202)	47.33 ± 2.39 (10.00-171.00)	5.84e-26
Creatinine (mg/dL)	0.82 ± 0.03 (0-3)	-	
	Number of patients	**Number of patients**	
CAD (Yes:No:Unknown)	21:334:0	0:0:146	
OSA (Yes:No:Unknown)	133:222:0	0:0:146	
Weight Loss Surgery (Yes:No)	339:16	0:146	
NAS (0:1:2:3:4:5:6:7)	77:74:49:37:34:44:28:12	0:6:10:22:29:41:24:14	
Fibrosis (Sirius Red) (0:1:2:3:4)	240:62:34:13:6	33:34:24:46:9	
Steatosis (0:1:2:3)	77:150:113:15	0:41:53:52	
Lobular Inflammation (0:1:2:3)	210:84:49:12	16:75:54:1	
Ballooning (0:1:2)	167:124:64	28:70:48	
Diabetes (Yes:No:Unknown)	117:238:0	59:84:3	
Gender (Female:Male:Unknown)	260:95:0	95:60:1	
Ethnicity (Hispanic or Latino:Non-Hispanic or Latino:Refused or Unknown)	26:291:38	0:146:0	
Tissue source	118 MGH Weight Center:237 STM Repository	146 NCL	
Preservation method (RNAlater:Flash Frozen:FFPE)	245:82:28	0:120:26	
Sample type (Core Needle Biopsy:Wedge Biopsy)	265:90	146:0	

Table S1: Massachusetts General Hospital (MGH) and St. Mary's weight loss surgery clinic (STM) or Newcastle University (NCL) patient clinical and biochemical characteristics. BMI: Body Mass Index; CAD: Coronary Artery Disease; OSA: Obstructive Sleep Apnea; NAS: NAFLD Activity Score.

Protease	AUC
MMP9	0.84***
FAP	0.86***
MMP2	0.85***
MMP14	0.80***
ST14	0.84***
MMP19	0.80***
CTSK	0.79***
PLAU	0.81***
CTSD	0.81***
GZMK	0.77***
GZMA	0.76***
FURIN	0.71***
MMP7	0.73***

Table S2: Contribution of NASH proteases to gene-based classifier to discriminate ≥F2 *vs* F0-F1: AUC and significance. n.s.: non-significant, Student's t-test ≥F2 vs F0-F1; *p ≤ 0.05, **p ≤ 0.01, ***p ≤0.001.

	AUC (+ and -) Condition	AUC (+) Condition	AUC (-) Condition	Significance
MGH cohort:				
T2DM	0.90 (0.85 - 0.95)	0.90 (0.81 - 0.96)	0.90 (0.81 - 0.96)	0.977 (n.s.)
Obesity		0.90 (0.85 - 0.95)	N.A.	N.A.
NC cohort:				
T2DM	0.86 (0.80 - 0.91)	0.91 (0.84 - 0.96)	0.82 (0.71 - 0.91)	0.160 (n.s.)
Obesity		0.85 (0.77 - 0.92)	0.83 (0.69 - 0.94)	0.713 (n.s.)
MGH+NC cohort:				
T2DM	0.82 (0.78 – 0.86)	0.81 (0.75 - 0.86)	0.83 (0.77 - 0.88)	0.549 (n.s.)
Obesity		0.82 (0.78 - 0.86)	0.85 (0.72 - 0.95)	0.744 (n.s.)

Table S3: RNA ≥F2 classifier performance is unchanged with respect to NASH comorbid states: obesity or diabetes. RNA ≥F2 classifier AUC was calculated in MGH and NC patient populations with (+) and/or without (-) comorbidities (obesity as defined by BMI>30 or T2DM). AUCs were compared using the bootstrap method with 10,000 permuations. N.A.: Not Applicable, n.s.: not significant.