LIVER FIBROSIS

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

Sophie C. Cazanave[1]*, Andrew D. Warren[1†], Maciej Pacula[1†], Fayçal Touti[1], Anna Zagorska[2], Nil Gural[3], Eric K. Huang[1], Sarah Sherman, Mehar Cheema[1], Sabrina Ibarra[1], Jamie Bates[2], Andrew N. Billin[2], John T. Liles[2], Grant R. Budas[2], David G. Breckenridge[2], Dina Tiniakos[4,5], Vlad Ratziu[6], Ann K. Daly[4,5], Olivier Govaere[4,5], Quentin M. Anstee[4,5], Louis Gelrud[7], Jay Luther[8], Raymond T. Chung[8], Kathleen E. Corey[8], Wendy Winckler[1], Sangeeta Bhatia[3], Gabriel A. Kwong[1,9]*

Noninvasive detection of nonalcoholic steatohepatitis (NASH), the progressive form of nonalcoholic fatty liver disease, promises to improve patient screening, accelerate drug trials, and reduce health care costs. On the basis of protease dysregulation of the biological pathways of fibrotic NASH, we developed the Glympse Bio Test System (GBTS) for multiplexed quantification of liver protease activity. GBTS-NASH comprises a mixture of 19 mass-barcoded PEGylated peptides that is administered intravenously and senses liver protease activity by releasing mass-barcoded reporters into urine for analysis by mass spectrometry. To identify a protease signature of NASH, transcriptomic analysis of 355 human liver biopsies identified a 13-protease panel that discriminated clinically relevant NASH ≥F2 fibrosis from F0-F1 with high classification accuracy across two independent patient datasets. We screened 159 candidate substrates to identify a panel of 19 peptides that exhibited high activity for our 13-protease panel. In the choline-deficient, L-amino acid-defined, high-fat diet (CDAHFD) mouse model, binary classifiers trained on urine samples discriminated fibrotic NASH from simple steatosis and healthy controls across a range of nondisease conditions and indicated disease regression upon diet change [area under receiver operating characteristics (AUROCs) > 0.97]. Using a hepatoprotective triple combination treatment (FXR agonist, ACC and ASK1 inhibitors) in a rat model of NASH, urinary classification distinguished F0-F1 from ≥F2 animals and indicated therapeutic response as early as 1 week on treatment (AUROCs >0.91). Our results support GBTS-NASH to diagnose fibrotic NASH via an infusion of peptides, monitor changes in disease severity, and indicate early treatment response.

INTRODUCTION

Nonalcoholic fatty liver disease (NAFLD) is the most prevalent chronic liver disease worldwide (*1*) and has been dubbed the silent epidemic due to the lack of symptoms until liver disease has progressed to cirrhosis with hepatic decompensation. Although there is substantial interpatient variation in the natural history of disease and long-term outcome, the presence of nonalcoholic steatohepatitis (NASH)—the progressive form of NAFLD characterized by steatosis accompanied with inflammation and hepatocyte ballooning—may lead to hepatic fibrogenesis and ultimately to life-threatening conditions including cirrhosis, liver cancer, and organ failure (*1*, *2*). Currently, only weight reduction and lifestyle intervention are regularly used for treatment, and despite investigations involving more than 70 drugs in the development pipeline (*1*), none have been approved to date. The lack of accurate noninvasive biomarkers means that liver biopsy is used to diagnose NASH and select patients in the majority of NASH clinical trials. Liver biopsy is an invasive procedure that carries small risk of complications (*3*) and is subject to sampling heterogeneity and interobserver interpretation that can lead to errors in diagnosis and staging (*4*, *5*). In NASH drug trials, the requirement for histological end points by serial biopsies continues to be a major barrier by influencing their size, length, and cost (*6*, *7*). Moreover, liver biopsies are unsuitable for screening at a population level, which leaves many patients who are otherwise asymptomatic undiagnosed until later in the disease process. For these reasons, noninvasive methods are needed to diagnose liver disease and indicate patient response to drugs earlier during treatment.

Numerous approaches are being evaluated for diagnostic, monitoring, pharmacodynamic, and prognostic applications across different contexts of use (*8*). Magnetic resonance (MR) imaging–estimated proton density fat fraction is a quantitative imaging–based biomarker of hepatic steatosis that is used in clinical trials to report on response to drugs that have an antisteatotic mechanism of action (*9*). To diagnose liver fibrosis, FibroScan or MR elastography measures the propagation velocity of shear waves to quantify liver stiffness as a biomarker of advanced fibrosis (F3-F4), but its use

[1]Glympse Bio Inc., Cambridge, MA 02138, USA. [2]Gilead Sciences Inc., Foster City, CA 94404, USA. [3]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK. [5]Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne NE1 7RU, UK. [6]Sorbonne Université, ICAN (Institute of Cardiometabolism And Nutrition), Hôpital Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, Sorbonne University, INSERM UMRS 1138 CRC, Paris 75013, France. [7]Bon Secours St Mary's Hospital, Richmond VA 23226, USA. [8]Liver Center, GI Division, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA. [9]The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology & Emory University, Atlanta, GA 30332, USA.
*Corresponding author. Email: gkwong@glympsebio.com (G.A.K.); scazanave@glympsebio.com (S.C.C.)
†These authors contributed equally to this work.

may be limited because of body habitus in some patients (*10*). Serum-based protein assays such as those for apoptosis [cytokeratin (CK) 18 fragments (*11*)] and fibrogenesis, either as single biomarkers or as part of a multiparameter scoring system (for example, collagen neoepitope PRO-C3, FibroTest, and enhanced liver fibrosis or ELF), have been proposed to improve disease or advanced fibrosis detection (*12*, *13*); however, they lack accuracy at discerning patients with clinically relevant ≥F2 NASH fibrosis from those with F0-F1, leaving the largest part of the upcoming drug-treatable patients undiagnosed (*14*). Highlighting the pressing need for noninvasive diagnostics, several major international initiatives involving academic and pharmaceutical partners are underway to validate and develop biomarkers that diagnose the severity of NAFLD/NASH and monitor changes in disease severity.

We developed the Glympse Bio Test System (GBTS)–NASH as an injectable 19-plex library of mass-barcoded PEGylated peptides to measure protease dysregulation in NASH livers as activity-based biomarkers. Proteases such as matrix metalloproteinases (MMPs) and cathepsins are involved in the key pathways of progressive NASH including lipogenesis, apoptosis, inflammation (*15*–*19*), and fibrosis (*20*–*25*). After intravenous administration, GBTS-NASH is cleaved by liver-associated proteases, releasing mass-barcoded reporters that then renally filter into urine for quantification by tandem mass spectrometry (MS/MS). Compared to endogenous circulating biomarkers, synthetic biomarker concentrations in urine are amplified by protease turnover of peptides and renal enrichment of reporters, which may substantially improve diagnostic sensitivity compared to serum biomarkers (*26*–*29*). Moreover, the use of a 19-plex library of probes allows capture of high-dimensional data to train diagnostic classifiers to improve specificity compared to single biomarkers. We demonstrate GBTS-NASH in the key use cases of diagnosis in nutritional rodent models of NASH fibrosis, monitoring changes in disease regression by diet change, and pharmacodynamic response to a hepatoprotective triple treatment combination as early as 1 week on therapy.

RESULTS

Human transcriptome analysis identifies proteases dysregulated in fibrotic NASH

We first quantified the breadth of protease dysregulation by transcriptomic analysis of 355 RNA-later, frozen, or formalin-fixed, paraffin-embedded (FFPE)–preserved liver biopsy samples ($n = 76$ normal, $n = 90$ NAFL, and $n = 189$ NASH F0 to F4) from the Massachusetts General Hospital (MGH) and St. Mary's weight loss surgery clinic (STM) (Fig. 1A and table S1). Of the ~550 proteases encoded by the genome, we designed a custom NanoString panel to quantify all 229 secreted and membrane-bound endoproteases as well as 570 nonprotease genes linked to lipogenic, inflammatory, and fibrotic pathways (*30*). Exoproteases were excluded because our peptide sensors are chemically protected at the C and N termini. Unsupervised hierarchical clustering of protease gene expression stratified the 355 normal and NAFLD samples into two distinct clusters, A and B (fig. S1). Cluster A was characterized by more advanced fibrosis [$n = 47$ with moderate to severe fibrosis (≥F2) and $n = 113$ with no or mild fibrosis (F0-F1) versus $n = 6$ ≥F2 and $n = 189$ F0-F1 in cluster B, Fisher's exact test, $^{****}P = 1.8 \times 10^{-12}$] and a higher number of NASH diagnoses (64% versus 44%, Fisher's exact test, $^{****}P = 1.7 \times 10^{-4}$). Within cluster A, NASH ≥F2 separated as a subset

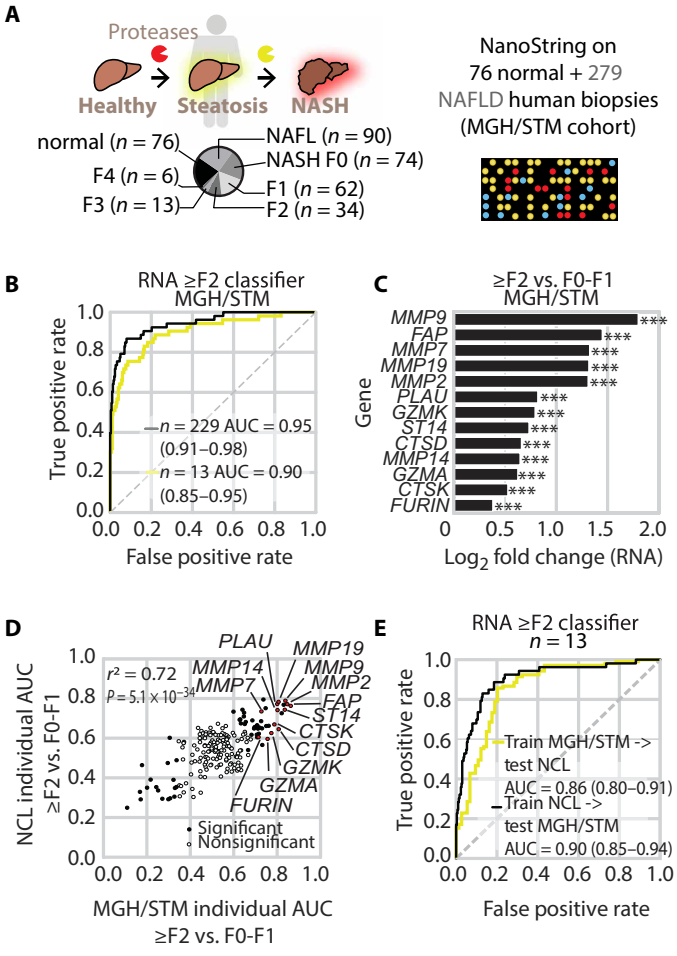

Fig. 1. Dysregulated protease expression classifies NASH ≥F2 from F0-F1 with high accuracy. (**A**) In silico identification of candidate proteases by analysis of liver RNA from MGH/STM human cohort including 76 normal individuals, 90 NAFL (steatosis), and 189 patients with NASH using a custom NanoString panel. Within the NASH cohort, the proportion of patients with stage 0, 1, 2, 3, and 4 NASH according to the CRN scoring system was $n = 74$ (39.2%), $n = 62$ (32.8%), $n = 34$ (18.0%), $n = 13$ (6.9%), and $n = 6$ (3.2%), respectively. (**B**) AUROC for binary classification of fibrosis stage ≥F2 vs. F0-F1 with a regularized logistic regression gene classifier trained with all 229 proteases (black) or a subset of 13 NASH proteases (green). (**C**) Differential expression of the 13 proteases in patients with NASH ≥F2 vs. F0-F1 plotted as $\log_2$ fold change ($^{***}P \leq 0.001$). (**D**) Pearson correlation of the AUROCs of 206 overlapping proteases to predict ≥F2 versus F0-F1 between the MGH/STM and NCL cohorts (filled circles: $^{*}P \leq 0.05$ in either dataset, Bonferroni-corrected). Thirteen-protease panel depicted as red circles. (**E**) AUROCs for binary classification of fibrosis stage ≥F2 versus F0-F1 by a 13-protease regularized logistic regression classifier trained on the MGH/STM cohort and validated with the independent Newcastle cohort (green) or *vice versa* (black).

(S1) from normal, NAFL, and NASH F0-F1 with minimum overlap (fig. S1). Integrated pathway cluster analysis of differentially expressed nonprotease genes confirmed that multiple metabolic, inflammatory, and extracellular matrix remodeling pathways were overrepresented in ≥F2 versus F0-F1 fibrosis (fig. S2).

To differentiate NASH by protease expression, we trained a regularized logistic regression classifier using all 229 protease counts as features and assessed its performance to distinguish NASH ≥F2 from F0-F1 with 100 rounds of randomized 80% training, 20% testing

(80/20) cross-validation. This binary cutoff was chosen because natural history studies and longitudinal monitoring by serial biopsies have highlighted NASH ≥F2 as associated with an increased risk of death and liver transplant (*31*) and as an important stage of clinical intervention (*32*). This 229-protease classifier distinguished ≥F2 from F0-F1 NASH with an area under the receiver operating characteristic (AUROC) of 0.95 [95% confidence interval (CI) 0.91 to 0.98] (Fig. 1B). We asked whether a subset of up-regulated proteases could distinguish patients with NASH ≥F2 by calculating AUROCs as a function of the top 1 to 32 significantly up-regulated proteases as ranked by analysis of variance (ANOVA) (n = 32, Bonferroni-corrected ANOVA, $^*P \leq 0.05$) (fig. S3). We found that classification accuracy depended on protease counts up to ~12 to 15, after which AUROCs negligibly increased. We therefore down-selected 229 proteases to a list of 13 that comprised four main families: metalloproteinases (MMP2, MMP7, MMP9, MMP14, and MMP19), cathepsins (CTSD and CTSK), granzymes (GZMA and GZMK), and other serine proteases (FAP, ST14, PLAU, and FURIN). A binary classifier trained on these 13 proteases accurately separating ≥F2 from F0-F1 with an AUROC of 0.90 (95% CI 0.85 to 0.95) (Fig. 1B) was statistically equivalent to the 229-protease classifier (P = 0.08) and correctly assigned the histological category of 295 of 355 patients (specificity = 83.1%, sensitivity = 83.0%) (fig. S4). All 13 proteases were significantly up-regulated in NASH ≥F2 compared to F0-F1 ($^{***}P \leq 0.001$, Bonferroni-corrected; Fig. 1C) with univariate AUROCs ranging from 0.70 to 0.80 ($^{***}P \leq 0.001$) (table S2).

As the MGH/STM patient samples were collected from a bariatric surgery clinic, potential differences in population distribution and clinical characteristics may bias protease selection. Therefore, we cross-validated our results with a natural history cohort of NAFLD (33 NAFL and 113 NASH) from Newcastle University (NCL, United Kingdom) and Hôpital Pitié Salpêtrière (France). NCL liver samples were previously quantified by a NanoString panel that contained 206 overlapping proteases with our MGH/STM panel (*33*). Despite significant differences in clinical features including body mass index (BMI), age, alanine aminotransferase (ALT), and aspartate aminotransferase (AST) ($^{***}P \leq 0.001$) (table S1), the performance of single proteases in distinguishing ≥F2 versus F0-F1 strongly correlated between the two cohorts (Pearson's $r^2 = 0.72$, $P = 5.1 \times 10^{-34}$; Fig. 1D). Moreover, a binary classifier trained on the same 13-protease panel accurately separated ≥F2 from F0-F1 with an AUROC of 0.85 (95% CI 0.79 to 0.91) with a sensitivity of 78.1% and a specificity of 79.7% (fig. S5). The ability to predict patients at stage ≥F2 did not depend on whether the classifier was trained on the MGH/STM and tested on the NCL cohort (AUROC = 0.86, 95% CI 0.80 to 0.91) or trained on the NCL and tested on the MGH/STM cohort (AUROC = 0.90, 95% CI 0.85 to 0.94; Fig. 1E). Both were statistically equivalent to classifiers trained and tested on the same dataset (P = 0.49 to 0.54). As obesity and T2DM are prevalent in NASH populations (*34*), we examined potential confounding signals from comorbidities. We calculated ≥F2 classifiers without retraining separately in obese versus nonobese or diabetic versus nondiabetic patients with NASH (table S3). In the MGH cohort, all patients were overweight (average BMI ~45) (table S1) and were pooled with the NCL cohort to allow for obese versus nonobese comparison. Across these two comorbid groups, AUROCs to predict ≥F2 were unchanged by obesity (MGH + NCL AUROC = 0.82 in obese versus AUROC = 0.85 in nonobese, P = 0.74, nonsignificant) or diabetic status [MGH AUROC = 0.90 in type 2 diabetes mellitus (T2DM) versus AUROC = 0.90 in non-T2DM, P = 0.98; NCL AUROC = 0.91 in T2DM versus AUROC = 0.82 in non-T2DM, P = 0.16, all nonsignificant].

Last, we compared the performance of our protease classifier with other noninvasive tests. Fibrosis-4 (FIB-4) is recommended by the American Association for the Study of Liver Disease (*35*) for detecting advanced fibrosis. In the NCL cohort, the FIB-4 index underperformed our protease RNA classifier to classify NASH ≥F2 with an AUROC of 0.64 (95% CI 0.55 to 0.72, $^{****}P \leq 0.0001$) (fig. S5). The sensitivity and specificity of FIB-4 for ≥F2 diagnosis were 65.3 and 54.3%, respectively, at a low cutoff value of 1.3, and of 11.9 and 98.6% at a high cutoff value of 2.67. Liver stiffness values by FibroScan for 82 patients from the NCL cohort (19 NAFL and 63 NASH: 25 F0-F1, 10 F2, 25 F3, and 3 F4), at a Youden cutoff value of 8.2 kPa (*36*), predicted ≥F2 (AUROC = 0.80, 95% CI 0.70 to 0.90, P = 0.12, nonsignificant) with similar performance but with significantly lower sensitivity and specificity (73.7 and 56.8%, respectively) when compared to our RNA ≥F2 (AUROC = 0.88, 95% CI 0.79 to 0.95, $^*P$ = 0.041). Together, these results indicated that transcript levels of a 13-protease panel classify patients with NASH ≥F2 with high accuracy across two independent datasets.

Design of activity-based sensors to detect NASH proteases

GBTS-NASH probes are mass-barcoded, PEGylated peptides comprising an eight-arm 40-kDa poly(ethylene glycol) (PEG) core conjugated to protease substrates labeled with a mass barcode (Fig. 2A). PEG has a longstanding history of clinical safety for U.S. Food and Drug and Administration–approved drugs (*37*) and a branched eight-arm structure allows multivalent presentation of peptides and extends sensor half-life in circulation to increase distribution to the liver (*38*). The hydrodynamic diameter of 40-kDa eight-arm PEG (~10 nm) is larger than the glomerulus pore size (~4 to 5 nm) to prevent surface-conjugated peptides from filtering into urine before protease cleavage, with less than 0.5% of injected dose filtering into urine (*28*).

To identify peptide substrates, we synthesized a library of 159 fluorogenic substrates ranging 4 to 12 amino acids in length flanked by a fluorophore (FAM) and quencher (CPQ2) using sequences from previously reported studies (*29*, *38*). This library was assayed with recombinant proteases from our 13-member panel by quantifying increases in sample fluorescence over time (Fig. 2B). To facilitate down-selection from 159 candidate substrates, we applied t-distributed stochastic neighbor embedding (t-SNE) to compare substrate cleavage activities (Fig. 2C). t-SNE revealed three major groups: Cluster 1 substrates were preferentially cleaved by MMPs and cathepsins, cluster 2 by cathepsins, and cluster 3 by different families of serine proteases. From our library, we selected N01, N02, N03, N04, N05, N06, N07, and N08 from cluster 1; N09, N10, N11, N12, and N13 from cluster 2; and N14, N15, N16, N17, N18, and N19 from cluster 3 (Fig. 2D). To verify substrate cleavage after conjugation to eight-arm PEG, we tested three sequences (N03, N05, and N18) across a broad range of peptide concentrations (0.09 to 6 μM) and observed similar cleavage kinetics between free substrates and their PEGylated counterparts (Fig. 2B).

To allow detection by mass spectrometry, each GBTS-NASH substrate was synthesized in tandem with a mass barcode composed of the reporter peptide Glu-Fibrinopeptide B (Glu-Fib) via a photocleavable residue (3-amino-3-(2-nitrophenyl)propionyl or ANP) (*29*). Upon protease cleavage, Glu-Fib mass reporters are released from the PEG carrier and accumulate in urine (*38*), allowing recovery of

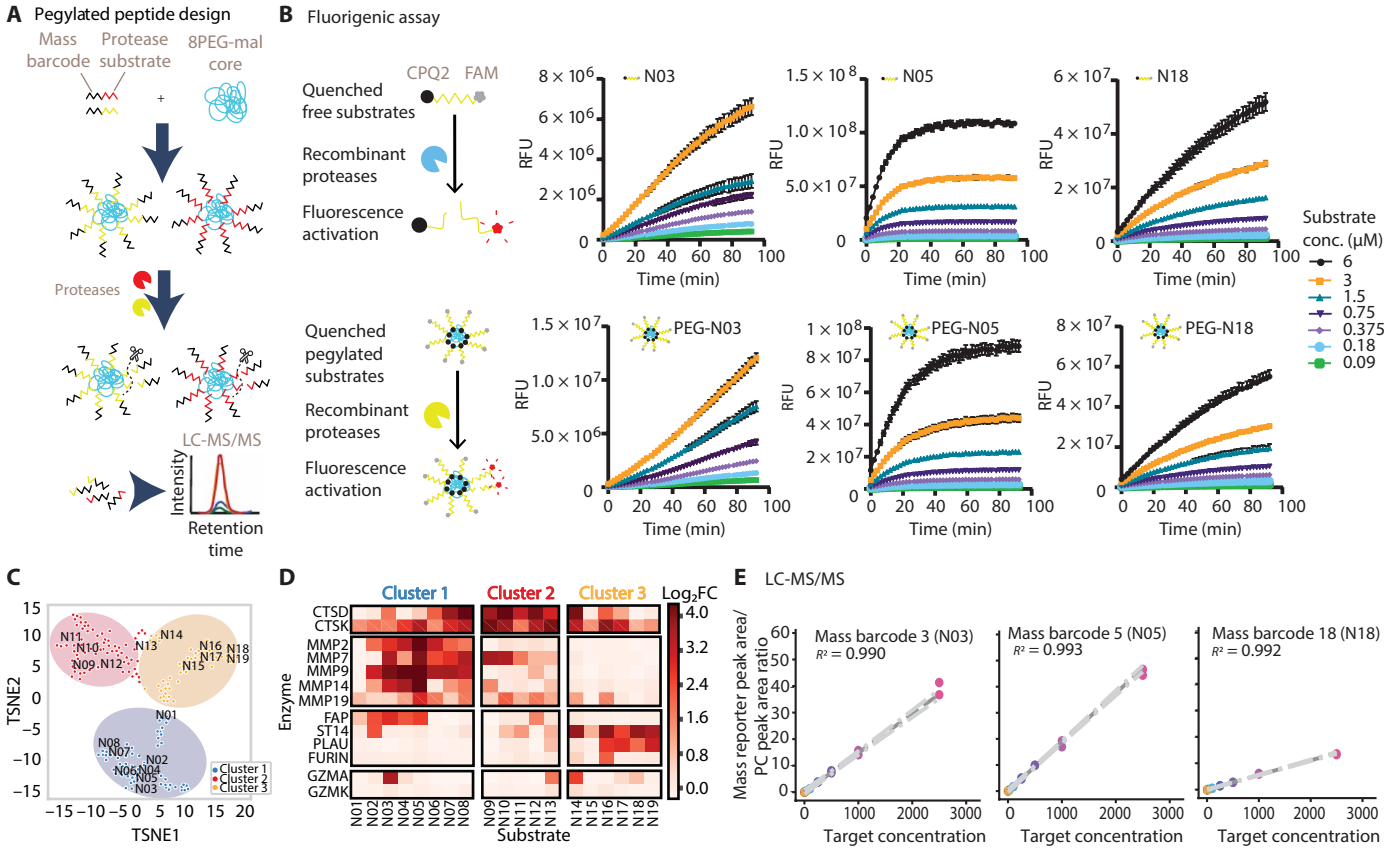

Fig. 2. Mass-barcoded PEGylated peptides for multiplexed detection of protease activity by tandem mass spectrometry. (**A**) Schematic of PEGylated peptides composed of a protease substrate conjugated to an eight-arm PEG. Each protease substrate was uniquely labeled with a mass barcode for quantification of cleavage fragments by liquid chromatography tandem mass spectrometry (LC-MS/MS). (**B**) Schematic of example peptide substrates (N03, N05, and N18) showing peptides flanked by a quencher (CPQ2) and fluorophore (FAM). Both free or PEGylated peptides were assayed at different substrate concentrations (0.09 to 6 μM) by fluorimetry after addition of their respective protease (MMP9, ST14, or MMP13 at 50 nM). *y* axes differ for visibility. (**C**) t-SNE map of our 159-substrate library cleavage profile. Each data point represents a unique protease substrate. (**D**) Hierarchical clustering of fluorescence fold changes at 90 min of 19 selected peptides cleaved by 13 NASH proteases. (**E**) Calibration curves of mass barcodes 3, 5, and 18 after ultraviolet-triggered release from PEGylated peptides N03, N05, and N18 as quantified by LC-MS/MS. Peak area from each mass barcode is divided by the peak area of an isotopically labeled internal standard (PC).

mass barcodes after photolysis. We synthesized Glu-Fib reporters with D-stereoisomer amino acids to resist peptidolysis. To enable multiplexed liquid chromatography MS/MS (LC-MS/MS) quantification, we adapted previously described isobaric encoding strategies (*29*, *38*) by enriching Glu-Fib peptides with stable heavy isotope–labeled amino acids, thereby extending the number of mass barcodes to 19. We verified the ability to resolve and quantify all 19 mass barcodes by LC-MS/MS from purified urine samples with R^2 values ranging from 0.915 to 0.995 at dose-relevant concentrations after peak area normalization by peak area from an internal control reporter (PC) spiked into the urine before analysis (Fig. 2E and fig. S6). Collectively, our data show that mass-barcoded PEGylated peptides are sensitive to cleavage by proteases up-regulated in NASH with fibrosis and allow multiplexed quantification by MS/MS.

GBTS-NASH enables noninvasive detection of NASH F1-F2 in a dietary preclinical model

We next tested the ability of 19-plex GBTS-NASH to noninvasively detect NASH from urine. We used the dietary model where C57BL/6 mice are fed a 60% choline deficient, L-amino acid-defined, high fat diet (CDAHFD) (Fig. 3A) (*39*). To confirm NASH and progressive fibrosis, we performed liver histology to quantify the NAFLD activity score [NAS, a nonweighted score for steatosis, lobular inflammation, and hepatocyte ballooning (*40*)] and extent of fibrosis by PicroSirius Red (PSR) staining of collagen fibrils (Fig. 3, B and C, and fig. S7, A and B). At 9 weeks on CDAHFD, both the NAS (100% NAS $\geq$ 5) and fibrosis stage were significantly increased (n = 16, 73% F1 and 27% F2, chi-square $^*P \leq 0.05$; PSR-positive area = 4.3 ± 0.3%, $^{****}P \leq 0.0001$) compared to healthy mice fed a standard chow diet (CD) (n = 16, 100% NAS 0, 100% F0; PSR-positive area = 0.5 ± 0.04%) (Fig. 3C and fig. S7B). By comparison, liver sections from mice fed a 60% high-fat diet (HFD) for 16 weeks to model simple steatosis (NAFL) and obesity (body weight: 27.0 ± 0.6 g in CD versus 44.1 ± 1.6 g in HFD, $^{****}P \leq 0.0001$) showed an increase in NAS (65% NAS ≤3 and 35% NAS >3) (Fig. 3C) with steatosis scores close to the CDAHFD group (30% steatosis score 1 and 70% steatosis score >2) but with little to no lobular inflammation (70% lobular inflammation score 0 and 30% lobular inflammation score 1) or fibrosis (100% F0, PSR-positive area = 0.4 ± 0.03%) (fig. S7B). To compare the extent of protease dysregulation between murine and human disease, we performed RNA sequencing (RNA-seq) on liver samples from mice with NASH F1-F2 (9 weeks of CDAHFD, n = 6),

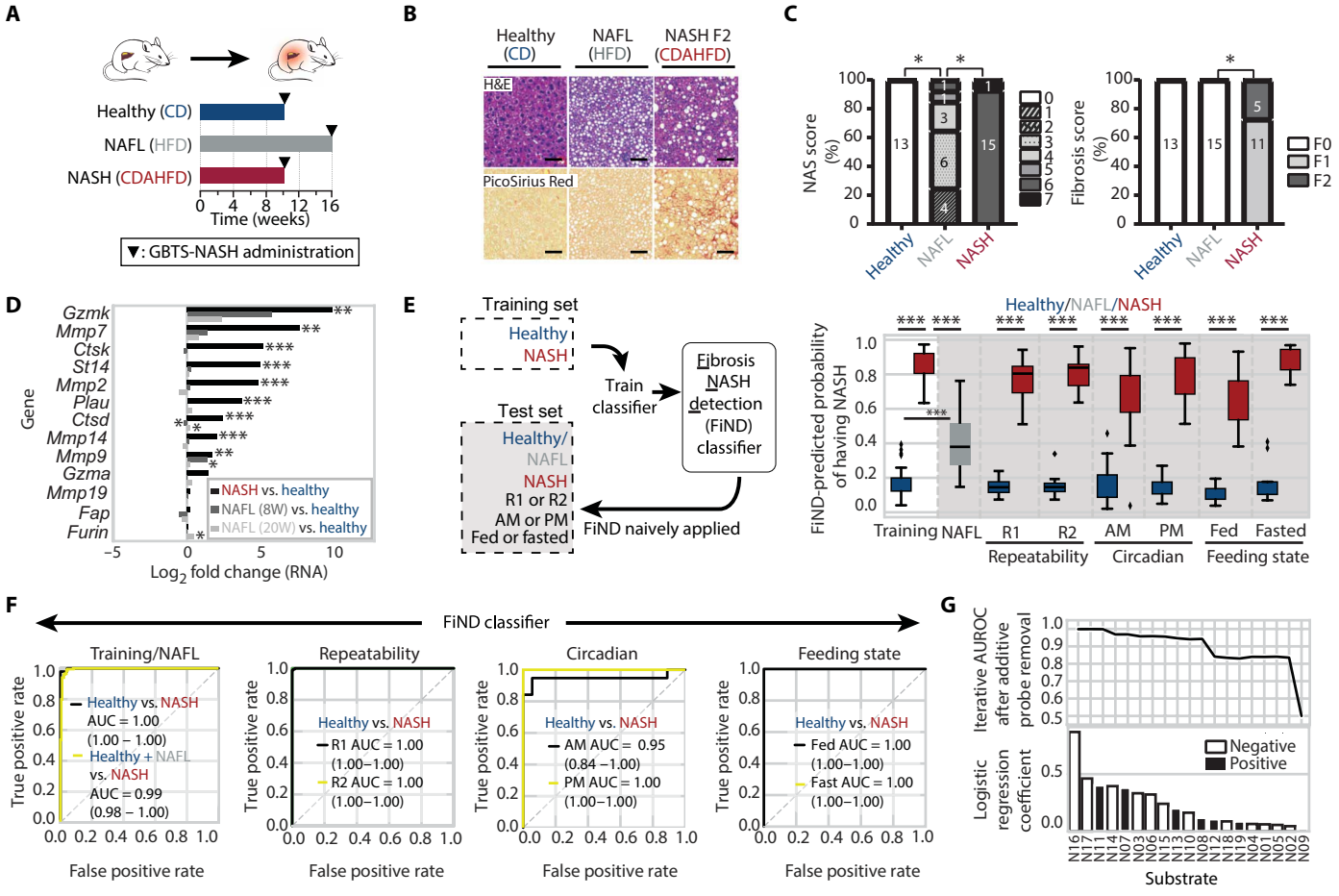

Fig. 3. Noninvasive urinary classification of NASH in a dietary mouse model by logistic regression. (**A**) NASH was modeled in mice using a choline-deficient amino acid–defined high-fat diet (CDAHFD) model for 9 weeks. Additional cohorts included mice fed a standard chow diet (CD) for 9 weeks as healthy controls or a 60% high-fat diet (HFD) for 16 weeks as simple steatosis NAFL group. In all experiments, mice received GBTS-NASH at a dose of 0.125 nmol per PEGylated peptide. (**B** and **C**) Representative histology images from hematoxylin and eosin (H&E)– and PicroSirius Red–stained liver slides from patients with NASH or NAFL or CD. Scale bars, 100 μm. (C) Percentage of animals per group with a specific assigned NAS or fibrosis score for healthy ($n = 16$, CD 9 weeks), NAFL ($n = 15$, HFD 16 weeks), and NASH cohorts ($n = 16$, 9 weeks of CDAHFD). The number of animals per group is shown within bars. Chi-square $^{*}P \leq 0.05$. (**D**) Thirteen NASH protease genes were assessed by RNA-seq from liver samples from mice with six NASH (9 weeks of CDAHFD), four NAFL (8 weeks of HFD), or four healthy controls (of CD). We also combined our in-house RNA-seq data with publicly available NAFL datasets (GSE138945 and GSE138946) from C57BL/6 mice kept on the same HFD for 20 weeks ($n = 3$) using the same sequencing platform. Student's t test, $^{**}P < 0.01$. (**E**) Left: Schematic of fibrosis NASH detection (FiND) classifier (using logistic regression and 80/20 cross-validation) generated using LC-MS/MS quantified concentrations of 19 mass-barcoded reporters from urine samples from CD healthy ($n = 48$) or CDAHFD ($n = 48$) mice. Only healthy and NASH mice were used to train the classifier. NAFL mice ($n = 15$) as well as separate independent cohorts of healthy and NASH mice used to assess MS assay repeatability (R1 at 9 weeks, R2 at 10 weeks, $n = 15$ CD and $n = 16$ CDAHFD), GBTS-NASH treatment time (AM, PM, $n = 18$–19 CD and 19–20 CDAHFD at 11/12 weeks), or feeding status (fed versus fasted, $n = 14$ CD and 15 CDAHFD at 20 weeks) were naively applied to the FiND classifier. Right: FiND-predicted probability of having NASH in each group and cohort. (**F**) AUROCs of GBTS-NASH staging classifier to discriminate NASH in CDAHFD animals from healthy CD mice from all cohorts tested (training, R1, R2, AM, PM, fed, and fasted) and from combined healthy + NAFL animals. (**G**) Logistic regression coefficients for probes included in the FiND classifier. Iterative feature elimination analysis (top) is shown. Probes were removed in order of importance to the classifier. The directionality of logistic regression coefficient is represented as back-filled bars for positive and white-filled bars for negative (bottom).

NAFL (8 weeks of HFD, $n = 4$), or healthy controls (9 weeks of CD, $n = 4$). We also combined our in-house RNA-seq data with publicly available NAFL datasets (GSE138945 and GSE138946) from C57BL/6 mice on HFD for 20 weeks ($n = 3$) (*41*). Of the 13 proteases identified from analysis of human liver samples, 9 were significantly up-regulated in murine NASH F1-F2 ($\log_2$ fold change ≥ 1.71, $^{**}P \leq 0.01$) and only 1 or 2 were significantly up-regulated in steatosis groups at 8 and 20 weeks ($^{*}P \leq 0.05$), respectively, compared to healthy liver controls (Fig. 3D).

To assess the potential of GBTS-NASH to detect NASH F1-F2 from urine, we administered GBTS-NASH by retro-orbital injection to cohorts of CDAHFD ($n = 48$) and CD mice ($n = 48$) and collected urine within the first 2 hours for analysis by mass spectrometry (Fig. 3A). Using z scores of normalized reporter concentrations, we trained a regularized logistic regression classifier to output a NASH probability for each urine sample. Our classifier—which we named the fibrosis NASH detection (FiND) classifier—assigned a high median NASH probability of 86% to urine samples from CDAHFD

mice and a low median probability of 18% to samples from CD mice (Fig. 3E). By ROC analysis of the predicted probabilities, the FiND classifier discriminated NASH F1-F2 from healthy urine samples with an average test set AUROC of 1.00 (specificity = 100.0% and sensitivity = 100.0%) (Fig. 3F). To assess the relative weight of each probe to classification, we performed iterative feature elimination based on the absolute magnitude of probe coefficient (Fig. 3G). We observed that AUROCs for discriminating NASH F1-F2 from healthy remained above 0.95 with as many as 7 probes excluded or above 0.90 with 10 excluded probes, indicating a redundant, multivariate F1-F2 NASH signature. We further asked whether our FiND classifier could discriminate NASH F1-F2 from NAFL without retraining and found that FiND assigned a median NASH probability of 37% to urine samples from NAFL mice (HFD $n = 15$), which was closer in value to the median probability assigned to healthy (18%) than to NASH F1-F2 (86%) samples (Fig. 3E). Moreover, when NAFL urine samples were grouped together with healthy urine samples, FiND discriminated NASH F1-F2 with a classification AUROC of 0.99 (95% CI 0.98 to 1.00, sensitivity = 100.0%, specificity = 85.7%) (Fig. 3F), indicating differentiation of NASH from all nonclinically relevant liver phenotypes.

We further tested the robustness of FiND across a wide range of non–disease-related variables. We applied the FiND classifier to urine samples collected from cohorts of mice to assess the effect of repeatability, GBTS-NASH treatment time, and fed-fasted states on classification, which was done without additional training. To test repeatability, we administered GBTS-NASH to NASH F1-F2 ($n = 16$) and healthy ($n = 15$) mice at week 9 (R1) and week 10 (R2) and found that the FiND output probability of having NASH was consistent for each individual animal between the two GBTS-NASH administrations (Pearson's $r = 0.96$, $^{****}P = 2.68 \times 10^{-18}$; fig. S8). FiND classification assigned a high median NASH probability of 84% to both cohorts and classified NASH F1-F2 and healthy mice with AUROCs of R1 and R2 = 1.00, specificity of R1 and R2 = 100.0%, sensitivity of R1 and R2 = 100.0% (R1 nonsignificantly different from R2) (Fig. 3, E and G). Likewise, FiND classification achieved similar accuracies regardless of whether GBTS-NASH was administered at 8 to 11 AM ($n = 18$ to 19) or 3 to 6 PM ($n = 19$ to 20) (AM AUROC = 0.95, 95% CI 0.84 to 1.00, specificity = 94.4% and sensitivity = 89.5% versus PM AUROC = 1.00, specificity = 100.0% and sensitivity = 100%, nonsignificant). FiND classification was also unaffected by whether animals were fed or fasted for 16 hours ($n = 14$ to 15 per group, AUROC of fed and fasted = 1.00, specificity of fed = 100.0% and fasted = 92.9%, sensitivity of fed = 93.3% and fasted = 100.0%, nonsignificant) (Fig. 3, E and G).

As NAFLD is a risk factor for chronic kidney disease (CKD) (*42*), we tested whether kidney fibrosis or T2DM-induced impaired kidney function could affect the performance of our urinary NASH test in two separate animal models. To model acute kidney injury (AKI) (*43*, *44*), we administered a single intraperitoneal dose of folic acid (FA) to C57BL/6 healthy CD 9 weeks (kidney fibrosis, $n = 20$) or NASH CDAHFD 9 weeks (NASH + kidney fibrosis, $n = 18$) mice (fig. S9A), after which animals were kept on their respective diets for an additional 2 weeks before GBTS-NASH injection. FA-treated animals were compared to age-matched cohorts of CD and CDAHFD mice injected with vehicle (healthy, $n = 25$ and NASH, $n = 20$). Both NASH and NASH + kidney fibrosis cohorts had similar liver NAS and fibrosis stage (100% NAS ≥ 5, 95% F ≥ 1, PSR-positive area = 2.4 ± 0.1% in NASH versus 95% NAS ≥ 5, 90% F ≥ 1, PSR-positive area = 2.5 ± 0.1% in NASH + kidney fibrosis, all nonsignificant) (fig. S9, B and C), and both healthy and kidney fibrosis animals had normal liver histology. Consistent with FA-directed nephrotoxicity, FA-treated animals displayed marked degenerative histology in tubular epithelial cells (fig. S9D) associated with excessive collagen deposition within kidney tissues (mean PSR-positive area > 6.9% in kidney fibrosis or NASH + kidney fibrosis versus mean PSR-positive area < 2.6% in healthy and NASH, $^{****}P \leq 0.0001$; fig. S9E). We prospectively applied our FiND classifier without retraining (fig. S9F) and found that FiND predicted NASH F1-F2 with similar accuracies whether mice had AKI or not (vehicle AUROC = 1.00, sensitivity = 100%, specificity = 88% versus FA AUROC = 0.99, 95% CI 0.98 to 1.00, sensitivity = 100%, specificity = 95%) (fig. S9, G and H). As AKI develops suddenly and more frequently in cirrhotic patients (*45*) whereas CKD develops gradually as a result of chronic illnesses such as diabetes (*42*), we further tested GBTS-NASH using the Black and Tan Brachyury (BTBR) *ob/ob* model of obesity and diabetes with mild features of diabetic nephropathy (DN) (*46*). These mice are insulin resistant, hyperinsulinemic, severely hyperglycemic, and obese with progressive proteinuria, polyuria, and glomerular hypertrophy (*46*, *47*). BTBR *ob/ob* mice fed CD for 9 weeks (DN, $n = 19$) displayed higher weight gain ($^{****}P \leq 0.0001$; fig. S10, A and B), plasma triglyceride ($^{*}P \leq 0.05$), and cholesterol ($^{**}P \leq 0.01$) (fig. S10, C and D) than control age- and strain-matched BTBR wild-type (WT) mice ($n = 20$). However, DN mice did not spontaneously develop a NASH phenotype (50% had mild NAFL phenotype without fibrosis) (fig. S10E). By contrast, BTBR WT mice developed NASH liver phenotype after a CDAHFD for 9 weeks (100% NAS >5 and 65% F ≥ 1 in BTBR WT CDAHFD, $n = 19$; fig. S10E). Because of strain difference, we retrained a classifier based on BTBR WT healthy and NASH mice (FiND_BTBR classifier; fig. S10F) with high NASH prediction accuracy (AUROC = 0.95, 95% CI 0.87 to 0.99, sensitivity = 89.5 specificity = 89.5). FiND_BTBR classified DN animals as healthy (median NASH probability = 25.6% for DN, 15.1% for BTBR WT CD, and 80.3% for BTBR WT CDAHFD; fig. S10G), and when urine samples from DN and healthy mice were grouped together, FiND_BTBR discriminated NASH without altering classification accuracy (AUROC = 0.95, 95% CI 0.88 to 0.99, sensitivity = 89.5%, specificity = 86.8%; fig. S10H). Together, these results demonstrated that multivariate classifiers trained on urinary reporters classify NASH F1-F2 with high accuracy under a wide range of comorbid and nondisease conditions.

GBTS-NASH sensors predict diet-induced fibrotic NASH regression

NASH and fibrosis activity and stage fluctuate over time as part of the natural history of the disease. Natural regressors or placebo-treated patients often show histological improvement in both fibrosis and disease activity after their enrollment in lifestyle-change programs at the time of recruitment into clinical trials (*48*, *49*). Therefore, we tested the ability of GBTS-NASH to report on NASH with fibrosis regression. C57BL/6 mice were fed CDAHFD for 9 weeks before being switched to CD for 1 or 3 weeks to model early and late regression, respectively (Fig. 4A). By liver histology, NAS scores progressively decreased from week 1 (80% of early regression, NAS ≤ 3, $n = 15$) to week 3 (100% of late regression, NAS ≤ 2, $n = 17$) on CD compared to mice on CDAHFD for 9 weeks (100% NAS ≥ 6, $n = 16$) (chi-square $^{*}P \leq 0.05$) (Fig. 4B and figs. S7A and S11). Fibrosis

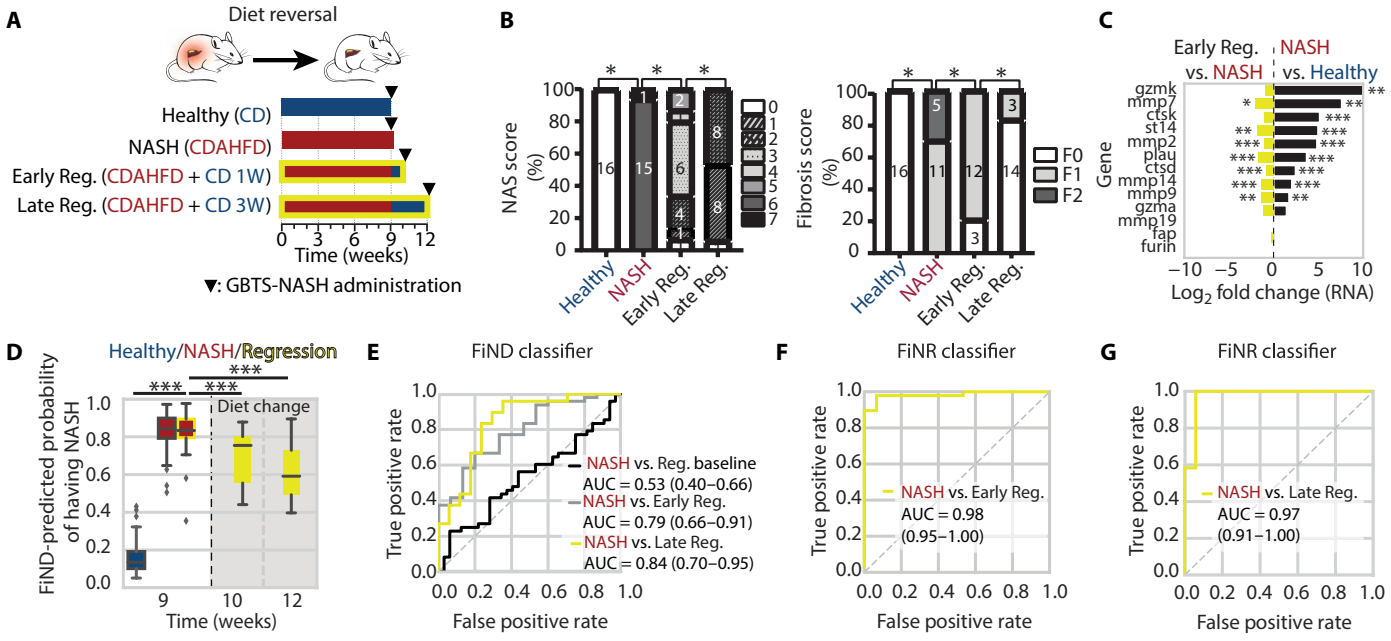

Fig. 4. GBTS-NASH protease sensors detect diet-induced early and late regression of NASH in mice. (**A**) NASH early or late regression (Reg.) was modeled by switching CDAHFD mice at 9 weeks (NASH) to a CD for 1 or 3 weeks (Early Reg. and Late Reg.). NASH animals received GBTS-NASH (0.125 nmol per PEGylated peptide) at baseline ($n = 32$) and then again at 1 week ($n = 15$) and 3 weeks ($n = 17$) after diet change (10 and 12 weeks total). (**B**) Histology NAS and fibrosis scores. Percentage of animals per group with a specific assigned NAS or fibrosis score is displayed. Number of animals for each scoring category is also shown within bars. Chi-square $^*P \leq 0.05$, $n = 16$ to 17 per group. (**C**) NASH liver expression of the 13 human NASH protease genes detected by RNA-seq was compared to healthy CD liver ($n = 4$ per group). (**D** and **E**) NASH animals at 9 weeks baseline (Reg. baseline, green, $n = 32$), and 1 week (Early Reg., $n = 15$) or 3 weeks (Late Reg., $n = 17$) after being switched to CD were applied naively to the FiND classifier (training on 48 CD and 48 CDAHFD for 9 weeks). FiND-predicted probability of having NASH is depicted in (D) and AUROC curves in (E), $^{***}P \leq 0.001$. (**F** and **G**) FiNR was retrained by including MS urine outputs from regressed animals into the FiND classifier and predicted early (F) and late (G) NASH regression from stable/progressing NASH.

scores likewise progressively decreased from week 1 (20% of early regression at F0 and 80% at F1, $n = 15$) to week 3 on CD (80% of late regression at F0 and 20% at F1, $n = 17$) compared to NASH animals (70 and 30% at F1 and F2, respectively, $n = 16$) (chi-square $^*P \leq 0.05$) (Fig. 4B). Although histology scores for fibrosis were decreased in regressed animal groups, collagen staining by PSR was unchanged compared to stable/progressing NASH mice (early regression PSR-positive area = 5.2 ± 0.2% or late regression PSR-positive area = 4.6 ± 0.4% versus CDAHFD 9 weeks PSR-positive area = 4.3 ± 0.3%, nonsignificant; fig. S7B). A possible explanation could be the absence of regression of perisinusoidal fibrosis (PSF) in this model (95% early regression, 100% late regression, and 100% CDAHFD 9 weeks with PSF score > 1, nonsignificant; fig. S7A), a feature not included in the classical NASH clinical research network (CRN) staging of fibrosis. NASH mice on CD for 1 week resulted in a significant decrease in mRNA expression of 7 of 13 NASH proteases when compared to the NASH 9 weeks baseline group ($^*P \leq 0.05$; Fig. 4C), which remained elevated compared to transcript levels from healthy mice (fig. S12).

On the basis of the improvement in liver phenotype by NAS and fibrosis histology scores and protease transcript abundance from diet change, we postulated that FiND classification would be able to track regression by assigning NASH probabilities that decrease over time. To test this, we administered GBTS-NASH to CDAHFD mice at 9 weeks and to CDAHFD mice switched to CD for 1 and 3 weeks (10 and 12 weeks total). As expected, urine samples from NASH animals before CD were undistinguishable from the previous NASH training cohort used to generate the FiND classifier, by both median NASH probabilities (0.83 versus 0.84, $P = 0.65$) (Fig. 4D) and AUROC analysis (AUROC = 0.53, 95% CI 0.40 to 0.66) (Fig. 4E). By contrast, NASH probabilities assigned to the early and late regression cohorts progressively decreased (Fig. 4D, early regression 0.75 ± 0.15, $n = 15$; late regression 0.59 ± 0.16, $n = 17$) relative to NASH samples ($^{***}P \leq 0.001$) and resulted in the ability to discriminate regression samples from fibrotic NASH with increasing AUROCs over time (early regression AUROC = 0.79, 95% CI 0.66 to 0.91; late regression AUROC = 0.84, 95% CI 0.70 to 0.95; Fig. 4E). Considering that FiND was trained on fibrotic NASH and healthy samples, we further asked whether detecting fibrotic NASH regression would be improved if classifier training set included urine samples from mice undergoing regression. To test this, we trained a separate fibrosis NASH regression (FiNR) classifier that included NASH, early regression, and late regression urine samples (Fig. 4, F and G). Under these conditions, FiNR predicted early regression (AUROC = 0.98, CI 95% 0.95 to 1.00) (Fig. 4F) and late regression from fibrotic NASH (AUROC = 0.98, CI 95% 0.92 to 1.00) (Fig. 4G) with higher accuracies compared to when regression samples were not part of the training set. Collectively, our results provided support that fibrotic NASH regression can be detected as early as 1 week after diet change, and classification accuracies improve when regression samples are included in the training set.

GBTS-NASH indicates early response to a hepatoprotective fibrotic NASH treatment combination

Fibrotic NASH is a metabolic disease and combinations of drugs with complementary mechanisms of action targeting metabolic

pathways are under investigation. A triple treatment combination (TRIPLE) composed of a farnesoid X nuclear receptor agonist (FXRa) to modulate bile acid, lipid, and glucose metabolism (*50*), an acetyl coenzyme A (CoA) carboxylase inhibitor (ACCi) to reduce fatty acid synthesis (*51*), and an apoptosis signal-regulating kinase 1 inhibitor (ASK1i) to block lipid-induced hepatocyte lipoapoptosis (*52*) was under evaluation as single agents and pairwise combinations as part of the phase 2 ATLAS trial (NCT03449446) (*53*). Dual treatment combinations ASK1i + FXRa, ASKi + ACCi, or ACCi + FXRa were also recently reported to exhibit hepatoprotection by reducing progression of inflammation and fibrosis in a rat model of NASH (*54*–*56*). We therefore designed a study to determine the ability of GBTS-NASH to indicate response to TRIPLE treatment. Our study included four cohorts of Male Wistar Han rats that were placed on CD for 12 weeks, CDAHFD for 6 weeks, CDAHFD for 12 weeks with daily administration of vehicle from week 6 to 12, or CDAHFD for 12 weeks with daily administration of TRIPLE treatment from week 6 to 12 (Fig. 5A). To confirm response to treatment, we compared plasma markers and histology of liver sections collected at weeks 6 and 12. Rats fed a CDAHFD for 6 weeks were characterized by elevated plasma ALT (287 ± 23 UI/liter, $^{****}P \leq 0.0001$) and CK 18 M30 antigen (158 ± 16 mUI/ml, $^{****}P \leq 0.0001$) (fig. S13, A to D); NAS scores (NAS ≥ 4, $n = 10$, chi-square $^{*}P \leq 0.05$) were increased in correlation with marked steatosis and lobular inflammation (CD68-positive area: 10.5 ± 0.6, $^{****}P \leq 0.0001$) with low to moderate fibrosis [40% F0, 40% F1, 10% F2, 10% F3, chi-square $^{*}P \leq 0.05$; PSR-positive area: 2.1 ± 0.3; α–smooth muscle actin (αSMA)–positive area 3.0 ± 0.4; hyaluronic acid (HA): 29.6 ± 3.4 ng/ml; procollagen III amino terminal propeptide (PIIINP): 24.1 ± 1.5 ng/ml, $^{***}P \leq 0.001$] (Fig. 5B and fig. S13A). After six additional weeks on CDAHFD, both NAS scores (median = 6) and fibrosis stage (100% ≥F3, chi-square $^{*}P \leq 0.05$; PSR-positive area: 8.9 ± 0.9) increased in severity compared to pretreatment liver sections ($^{****}P \leq 0.0001$; Fig. 5B and fig. S13B). By contrast, at the end of TRIPLE treatment at week 12, NAS scores (median = 4) and fibrosis stage (20% F0, 60% F1, 13% F2, 7% F3, chi-square $^{*}P \leq 0.05$; PSR-positive area: 2.4 ± 0.2) were statistically equivalent to liver section pretreatment and did not progress in disease severity. Reduction in NAS score and liver collagen deposition in TRIPLE-treated animals was associated with significant decrease in αSMA- and CD68-positive cells compared to vehicle-treated animals (CD68-positive area, 10.8 ± 0.5 versus 4.3 ± 0.2; αSMA-positive area: 6.8 ± 0.8 versus 0.6 ± 0.1, $^{****}P \leq 0.0001$; fig. S13, C and D). Although the PSR-positive area was similar between NASH at baseline and TRIPLE-treated cohort, αSMA staining was significantly lower in TRIPLE-treated rats (0.6 ± 0.1, $^{****}P \leq 0.0001$), almost to that of the healthy cohort (3.0 ± 0.4) (fig. S13, B and C), reflective of the rapid reversibility of active fibrogenesis before complete reduction

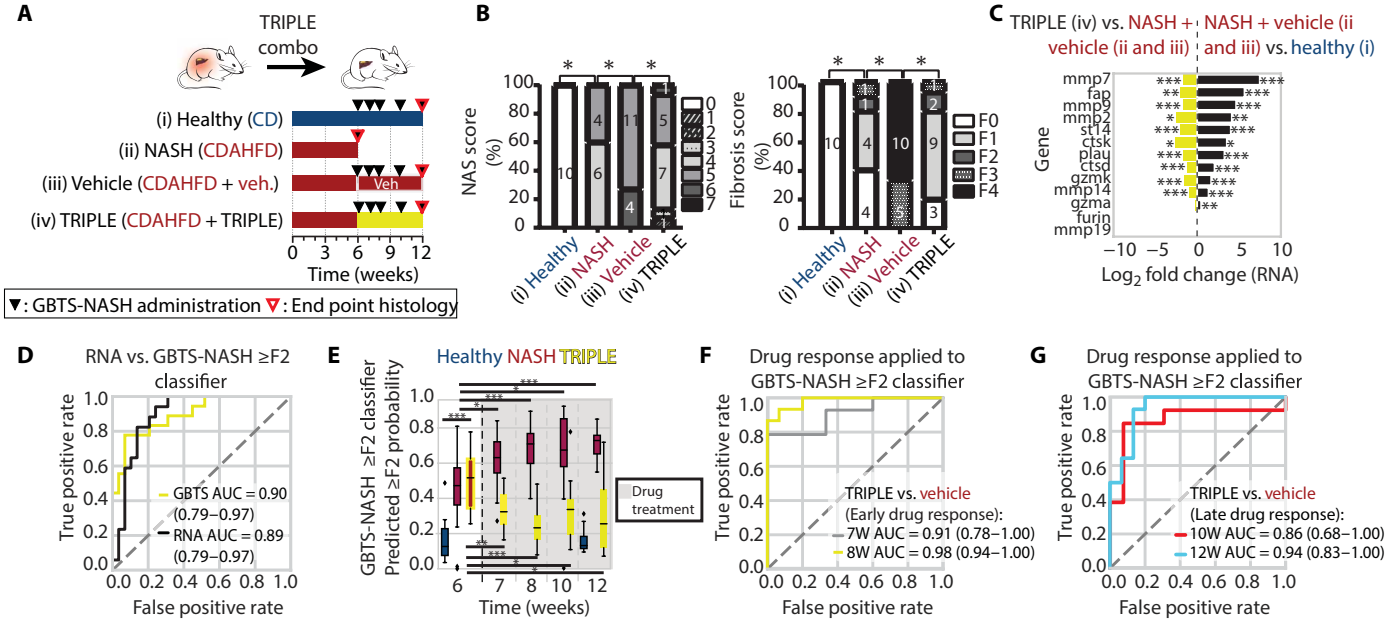

Fig. 5. GBTS-NASH sensors detect drug-induced fibrotic NASH hepatoprotection in rats. (**A**) We fed male Wistar Han rats either CD [(i), healthy, $n = 10$] or CDAHFD for up to 12 weeks ($n = 40$). From weeks 6 to 12 of CDAHFD feeding, a subset of NASH rats was administered vehicle [(iii), $n = 15$] or TRIPLE (ACC inhibitor at 10 mg/kg, FXR agonist at 30 mg/kg, and ASK1 inhibitor at 0.03% in the diet) for 6 weeks [(iv), TRIPLE-treated, $n = 15$]. A baseline CDAHFD 6-week animal group was also tested [(ii), NASH, $n = 10$] and used for histology scoring. Animals were administered GBTS-NASH (0.41 nmol per PEGylated peptide) at 6, 7, 8, 10, and 12 weeks. Bladders were voided at 0 to 120 min, and urine was collected and pooled over the following 120- to 180-min interval. (**B**) NAS and fibrosis histology score on end point histology liver samples. The percentage of animals with a specific assigned NAS or fibrosis score is displayed. The number of animals per group is shown within bars. Chi-square $^{*}P \leq 0.05$. (**C**) Log_2 fold change protease expression from RNA-seq analysis. (**D**) Logistic regression classifiers were trained on either fpkm (measured by RNA-seq) from 13 NASH protease gene (RNA ≥F2, black) or urine concentrations of 19 mass-barcoded reporters (measured by LC-MS/MS) (GBTS-NASH ≥F2 classifier, green) relative to histology fibrosis score with 100 rounds of randomized 80% training, 20% testing (80/20) cross-validation on end point NASH animals with available liver histology [NASH, (ii), $n = 10$; vehicle, (iii), $n = 15$; TRIPLE, (iv), $n = 15$]. (**E** to **G**) Baseline NASH 6 weeks (green) before administration of vehicle [(iii), $n = 15$] or TRIPLE [(iv), $n = 15$], and stable/progressing NASH at 7, 8, 10, and 12 weeks [veh., (iii), $n = 15$ for each], early [7 and 8 weeks of TRIPLE, (iv), $n = 15$], and late treatment response [10 and 12 weeks of TRIPLE, (iv), $n = 15$] MS urine outputs were applied naively to the GBTS-NASH ≥F2 classifier. The predicted probability of being ≥F2 is shown for each group in (E) and AUROC curves in (F) and (G). $^{*}P \leq 0.05$; $^{**}P \leq 0.01$; $^{***}P \leq 0.001$.

in established tissue collagen is seen. Also, lower plasma abundance of markers of inflammation and liver fibrosis was observed in TRIPLE-treated rats compared to vehicle-treated animals (ALT: 222 ± 10 versus 128 ± 10; CK18-M30: 180 ± 15 versus 81 ± 9 mUI/ml iv; HA: 54.8 ± 7.7 versus 25.7 ± 1.7 ng/ml iv; PIIINP: 19.4 ± 1.9 versus 10.1 ± 0.6 ng/ml iv, $^{***}P \leq 0.001$; fig. S13A).

To determine whether TRIPLE treatment affected liver protease expression, we performed RNA-seq on rat liver tissue. At 6 and 12 weeks on CDAHFD without TRIPLE treatment, liver expression of 11 and 12 proteases, respectively, from our 13-protease human NASH panel was significantly elevated compared to healthy rats ($^{*}P \leq 0.05$; Fig. 5C and fig. S14), confirming similar protease dysregulation in rat NASH. By the end of TRIPLE treatment, liver protease transcripts were significantly decreased compared to vehicle (Fig. 5C) but remained elevated compared to those in healthy rats ($^{*}P \leq 0.05$; fig. S14). We therefore asked whether changes in RNA transcript abundance could discriminate response to TRIPLE treatment. We trained a regularized logistic regression RNA-NASH ≥F2 classifier (100 rounds of cross-validation, 80/20 train/test split) to discriminate rats with ≥F2 fibrosis from F0-F1 (Fig. 5D) and found that it discriminated TRIPLE-treated animals from vehicle controls with an AUROC of 0.99 (95% CI 0.97 to 1.00) (fig. S15).

To test the ability of GBTS-NASH to indicate treatment response, we administered GBTS-NASH at weeks 6, 7, 8, 10, and 12 weeks (Fig. 5A) and analyzed urinary concentrations of our synthetic reporters. In contrast to mice, which do not develop fibrosis beyond F2 (Fig. 3C), rats on CDAHFD develop severe F3-F4 fibrosis (Fig. 5B). Therefore, we first trained a binary classifier to discriminate NASH ≥F2 versus F0-F1 using urine samples collected at 6 and 12 weeks from all cohorts of rats on the CDAHFD ($n = 10$ at 6-week baseline, $n = 15$ vehicle- or TRIPLE-treated animals). Samples from 7, 8, and 10 weeks were excluded from training as rats were not euthanized at those time points, and therefore, histological "ground truths" of disease severity were not available. Similar to the classification accuracy achieved in mice, our GBTS-NASH ≥F2 classifier in rats discriminated ≥F2 fibrosis from F0-F1 with high sensitivity and specificity (AUROC = 0.90, 95% CI 0.79 to 0.97, specificity = 82.8%, sensitivity = 83.3%). Moreover, our GBTS-NASH ≥F2 classifier achieved an accuracy that was statistically equivalent to RNA classification (Fig. 5D) and predicted ≥F2 probabilities that highly correlated to the output probabilities by RNA classification (Pearson's $r = 0.7$, $P = 5.16 \times 10^{-8}$; fig. S16). To assess treatment response, we reasoned that the ≥F2 threshold could be applied to determine on-treatment response given that rats on vehicle developed severe fibrosis (F3-F4) whereas most of the TRIPLE-treated cohort (12 of 15 rats) did not progress beyond F1 (Fig. 5B). Therefore, we applied our GBTS-NASH ≥F2 classifier without retraining to intermediary time points at weeks 7, 8, and 10 (1, 2, and 4 weeks total treatment, respectively). Under these conditions, the output NASH ≥F2 probabilities for TRIPLE combination treatment and vehicle controls, which were statistically equivalent at pretreatment (median 52% versus 47% respectively, $P = 0.56$), significantly diverged within 1 week on treatment (median 32% versus 63%, $^{***}P < 0.001$) (Fig. 5E), allowing discrimination of treatment response with an average test AUROC of 0.91 (95% CI 0.78 to 1.00) (Fig. 5F). At later time points, further separation of the median NASH ≥F2 probabilities between TRIPLE- and vehicle-treated cohorts was observed, maintaining the ability to distinguish treatment responses by AUROC analysis (Fig. 5, E and G). Collectively, these results highlighted the use of GBTS-NASH to discriminate NASH ≥F2 from F0-F1 disease and to indicate response to a hepatoprotective TRIPLE combination treatment as early as 1 week on treatment in rats.

DISCUSSION

Noninvasive methods to diagnose NASH, monitor changes in disease severity, and report on response to drug treatment earlier are needed. We showed by transcriptomic analysis of human liver biopsies that proteases are widely dysregulated in progressive NASH, and panels as few as 13 proteases discriminated ≥F2 from F0-F1 with high accuracy. This led us to develop a bespoke library of PEGylated peptides to detect NASH proteases by releasing cleaved reporters into urine for multiplexed quantification by mass spectrometry. In rodent models of NASH, we showed that binary classifiers trained on urine samples accurately distinguished NASH from NAFL, separated ≥F2 from F0-F1 fibrosis, and indicated disease regression and treatment response as early as 1 week after diet change or TRIPLE combination treatment.

Our discovery pipeline focused on transcriptomic analysis of human NAFLD liver biopsies to identify protease targets based on differential expression. Across two independent patient cohorts, we identified a panel of 13 proteases whose transcript abundances allowed binary classification of NASH ≥F2 from F0-F1, which is an important disease stage for clinical intervention and patient entry into NASH drug trials. This binary classification approach could potentially be extended to identify protease signatures that discriminate additional thresholds (for example, ≥F3) or developed into a multiclass staging classifier with a larger set of samples (for example, F1, F2, F3, and F4 simultaneously). Similar to other classifiers that rely on liver biopsy as the reference standard, the accuracy of RNA-based classification will likely be bounded by the accuracy of the liver biopsy itself, which, owing to sampling heterogeneity and interpretation errors, is accurate for staging fibrosis in 80 to 90% of patients (*57*). This implies that even if all other potential sources of error have been minimized, our classification results from the MGH/STM and NCL cohorts will likely not markedly deviate from our reported AUROCs (0.86 to 0.90). For these reasons, our discovery pipeline served primarily to nominate protease candidates for GBTS-NASH probe development.

Our panel of 13 NASH proteases informed the design of a 19-plex library of PEGylated peptides that was larger than the number of proteases to provide overlapping coverage and the ability to capture high-dimensional data. Although the same set of probes were used in all our animal experiments, the ability to capture high-dimensional data allowed us to develop separate classifiers to detect NASH, regression, and treatment response with high accuracies (AUROCs >0.9). Moreover, post hoc analysis revealed that not all probes were equally weighted and similar performance accuracy could be attained when as many as 10 probes were excluded from the classifier. Although the latter can be partially attributed to the lack of variation in isogenic models, these observations also highlight the potential to use a "superset" of probes for patient use. This may provide the ability to train separate classifiers across a range of intended use cases while allowing for the possibility to down-select to a smaller set of probes after validation.

Preclinical models of NASH are the cornerstone for testing pharmacological agents and diagnostics, but they have limitations. The CDAHFD model causes more severe inflammation and

fibrogenesis compared to other nutritional models of NASH but fails to reproduce the full spectrum of liver pathology that characterizes human NASH. For instance, mice fed CDAHFD are contrary to human fatty liver disease because they do not develop weight gain, dyslipidemia, or insulin resistance (*39*). Also, advanced fibrotic NASH and cirrhotic NASH (≥F3) do not develop in CDAHFD mice even after extended periods of feeding (20 weeks or more). Although rats fed CDAHFD replicate fibrosis range from F1 to F4, several key NASH histological features, such as hepatocyte ballooning, were not observed in FFPE-liver samples. Despite these limitations, we identified proteases that were similarly dysregulated between human and rodent NASH by transcriptomic homology analysis. As the majority of overlapping proteases (MMPs and cathepsins) are associated with inflammatory fibrosis (*58*, *59*), our test readily discriminated fibrotic NASH. By contrast, our test did not identify NAFL mice as being very different from healthy mice, as mice on HFD become obese and develop hepatic steatosis but with little inflammation and no fibrosis. Furthermore, our test was not confounded by comorbid chronic conditions associated with NASH such as obesity, diabetes, and CKD. These findings imply that our current test may be sensitive to inflammatory and fibrotic signals, potentially opening up the possibility for a future panel of probes to distinguish pathway-dependent activation signals.

Our results highlight potential strategies for clinical trial design. In both mice and rats, we observed that urinary classifiers trained to differentiate NASH from NAFL and healthy urine samples (for example, FiND and GBTS-NASH ≥F2 classifiers) could be applied to track NASH regression by diet change or TRIPLE treatment without retraining separate classifiers, albeit with reduced accuracy compared to classifiers directly trained on the regression samples (like FiNR). In both studies, the classifiers assigned NASH probabilities to urine samples that decreased as early as 1 week after diet change or TRIPLE combination treatment. These results indicate that it may be possible to design a clinical trial to train a single classifier to accurately differentiate NASH and stages of hepatic fibrosis, which then could be applied to track patient trajectory across multiple use cases (for example, regression or drug response). This classifier would, in principle, report changes in patient livers by monitoring how urine samples approached or deviated from a classifier-identified protease signature for a particular threshold, such as NASH ≥F2 as shown in our rat study. This approach would be similar to histological staging and could be further bolstered by future studies showing that protease dysregulation at particular stages of NASH fibrosis is conserved (for example, stable versus rapid progressors) regardless of dynamics (that is, progression or regression to the same stage) or drugs with different mechanisms of action. Our transcriptomic comparisons throughout this study show that the underlying protease biology is indeed similar, considering the high correlation of protease dysregulation between the MGH/STM and NCL datasets despite the cohorts representing different NASH populations (bariatric surgery versus natural history) with different clinical variables (BMI, age, ALT, and AST), as well as similar protease dysregulation across NASH regression and treatment response in both mouse and rat studies. An alternative strategy to a single classifier would be to design clinical trials for each intended use case. The advantage of dedicated classifiers that have been trained on the samples of interest is that their classification accuracy would likely exceed that of a single stage-specific classifier. This is supported by our animal studies where we observed that the ability to discriminate NASH regression by diet change improved when these samples were included in the training cohort.

We recognize potential limitations of our approach using preclinical NASH models lacking concurrent features of the metabolic syndrome with histological disease phenotype. An alternative model, the DIAMOND model (*60*), recapitulates both fibrotic NASH together with metabolic features (obesity, glucose intolerance, and adipose tissue inflammation) of the human disease and could be considered for further validation of GBTS-NASH specificity. Additional studies to directly assign our sensed proteolytic signal to liver cell-type population—for example, in liver-specific animal models depleted in macrophages or hepatic stellate cells—are also needed in future studies to provide insights into sensed biological pathways. In the human NAFLD staging system, fibrosis is defined in the context of disease progression and it is not yet established whether the current staging system adapts well to regression patterns. Thus, although diet-induced regression of NASH for 3 weeks reversed fibrosis stage to zero using standard histological staging, other markers of fibrosis (such as PSR-positive area) were unchanged compared to stable/progressing NASH animals. This discordance in the regressed animals could be due to unchanged PSF score, a histological parameter not included in the NASH CRN staging of fibrosis. In the future, deconvoluting multiplexed NASH protease signature using knockout animal models for specific inflammation or fibrosis-related proteases will clarify their relative contribution to the biological signal measured by GBTS-NASH.

Together, our studies provide support for the use of GBTS-NASH as a noninvasive urine test to diagnose, monitor, and assess pharmacodynamic responses by multiplexed quantification of liver protease activities. Additional future studies are warranted to evaluate the safety of GBTS-NASH in humans and qualify context of use and clinical efficacy in comparative end point studies.

MATERIALS AND METHODS

Study design

The goal of this study was to identify a liver protease signature specific to patients with NASH ≥F2 and determine whether intravenous administration of a multiplexed library of activity-based PEGylated peptides could be used to sensitively and specifically predict NASH and stage fibrosis in rodent models of NASH. A total of 355 NAFLD liver biopsies covering the full histological disease spectrum were included in this study. This discovery cohort was used for NanoString analysis to identify protease signature in NASH ≥F2. We cross-validated our results with a second independent cohort of 146 NAFLD liver biopsies using the same NanoString technology. Detailed phenotypic description and demographics are reported in table S1. Both cohorts were stratified according to histopathological disease grade and stage (NAFL, NASH-F0, -F1, -F2, -F3, and -F4). Logistic regression modeling was used to correlate gene expression with histological features. We further developed a library of 19 PEGylated peptides (GBTS-NASH) to detect NASH proteases by releasing cleaved reporters into urine for multiplexed quantification by mass spectrometry. We tested their performance in predicting NASH and staging fibrosis in rodent models of NASH progression and regression. Several logistic regression classifiers were used to correlate urinary cleaved reporters to disease phenotype or to fibrosis stage. The FiND classifier was trained and tested on 48 healthy (CD 9 weeks) animals and 48 NASH (CDAHFD 9 weeks)

to predict NASH progression. NAFL cohorts (HFD 16 weeks, $n = 15$) or additional cohorts of CD and CDAHFD mice (AM, PM, R1, R2, fed, fasted, vehicle, FA cohorts; $n = 14$ to 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 (CDAHFD 9 weeks + CD 1 week) or on 32 NASH baseline 9 weeks and 17 late regression animals (CDAHFD 9 weeks + CD 3 weeks) to predict NASH regression. Last, GBTS-NASH ≥F2 classifier was trained and tested on rat cohorts of NASH (CDAHFD 6 weeks, $n = 10$), NASH with vehicle (CDAHFD 6 weeks + vehicle 6 weeks, $n = 15$), and NASH with TRIPLE treatment (CDAHFD 6 weeks + TRIPLE 6 weeks, $n = 15$) to predict NASH ≥F2.

Human liver biopsies

Biopsies were collected from individuals visiting the Weight Loss Surgery and Obesity clinic at MGH and STM. All protocols, consent forms, and manual of operations were approved by the institutional review board for each site. Before surgery, patients were prospectively screened for NAFLD based on preset criteria based on abnormal ALT, AST blood biochemistry assays, or ultrasound features and offered a liver biopsy. Exclusion criteria included excessive alcohol consumption and liver diseases other than NAFLD. The subjects' demographic characteristics, medical history, medication use, clinical tests, and liver biopsy results were collected (table S1). A wedge (for 90 patients) or core needle biopsy (for 237 patients) was taken at the beginning of the surgery and flash-frozen (for 82 patients) or stored in RNA-later (for 245 patients). Because the weight loss surgery cohort had few patients with advanced fibrosis, we augmented the cohort with archived FFPE-biopsy samples (for 28 patients, all core needle biopsies) with NASH F2 to F4. By pathology read, the final cohort included 76 histologically normal liver samples, 90 samples with NAFLD, and 189 samples with NASH. Of the 144 NASH samples, there were 74 F0, 62 F1, 34 F2, 13 F3, and 6 F4 using the NASH CRN scoring system closely similar to the expected stage distribution from the suspected U.S. NASH population (*61*). FFPE and frozen biopsies from a natural history cohort of 146 NAFLD patient samples recruited at Newcastle upon Tyne Hospitals NHS, UK, and Hôpital Pitié Salpêtrière, France, were obtained from the European NAFLD Registry, including 33 NAFL and 113 NASH (5 F0, 29 F1, 24 F2, 46 F3, and 9 F4) (*34*). In the NASH group, longitudinal liver samples were collected twice for 21 patients, three times for 3 patients, and four times for 2 patients with a median time between biopsies of 5.9 years (min-max range of 1.2 to 14.6 year) for a total of 179 liver samples. All biopsies were read by the same blinded pathologist to improve scoring accuracy. The FIB-4 score was calculated and applied as previously described (*62*). Liver stiffness as assessed by FibroScan was available for 19 NAFL and 63 NASH: 25 F0-F1, 10 F2, 25 F3, and 3 F4. RNA extraction was performed on 5 to 20 μg of liver tissue using the miRNeasy Mini Kit (Qiagen) for flash-frozen samples and RNA-later liver or High Pure FFPET RNA Isolation Kit (Roche Life Science) for FFPE samples, according to the manufacturer's instructions. RNA quantity and quality were assessed using Agilent 2100 Bioanalyzer (Santa Clara). RNA integrity number (RIN) numbers, ranging from 1.5 to 9, were appropriate for use with the nCounter Technology (*63*).

Animal models

Five-week-old C57BL/6 (Taconic, Charles River Laboratories) or BTBR WT and BTBR *ob/ob* male mice (BTBR.Cg-*Lepob*/WiscJ; The Jackson Laboratory) were acclimated for a week at the Charles River Accelerator and Development Lab animal facility (Cambridge, MA). At 6 weeks of age, mice were fed a CDAHFD (from Research Diets, #A-06071302) for up to 20 weeks to develop a NASH phenotype with progressive fibrosis (*35*). To obtain a benign simple steatosis (NAFL) liver phenotype, C57BL/6 mice were fed with a 60% HFD (Research Diet, #D12492) for 8 and 16 weeks. CDAHFD and HFD have similar protein (18 versus 20% kcal), fat (62 versus 60% kcal), carbohydrate (21 versus 20% kcal), and energy density (5.21 kcal/g for both) content. Age-matched mice fed a standard CD (Charles River Diet 18% Vac Pac 5066) were used as controls. To induce NASH regression, C57BL/6 mice were fed a CDAHFD for 9 weeks and then switched back to a standard CD for an additional 1 or 3 weeks. A single dose of FA (Sigma-Aldrich) dissolved in 0.3 mM sodium bicarbonate (250 mg/kg body weight) was administered intraperitoneally to C57BL/6 mice fed a CD or CDAHFD for 9 weeks; experiments were carried out 2 weeks after FA administration. At the appropriate time of the experiment, mice were dosed via retro-orbital route with 100 μl of a solution of mannitol/sodium containing 0.125 nmol of each of 19 PEG-conjugated substrates (GBTS-NASH). After injection with GBTS-NASH, mice were placed on 96-well plates surrounded by custom-made restrainer to keep the mice contained for urine collection. Bladders were voided 60 min after GBTS-NASH administration, and all urine produced 60 to 120 after GBTS-NASH injection was collected and pooled together (between 50 and 200 μl of urine on average).

Male Wistar Han rats (10 weeks old, Charles River Laboratories) were fed either a control diet (5CR4) or CDAHFD (Research Diets, A16092003) for 12 weeks. From weeks 6 to 12 of CDAHFD feeding, a subset of CDAHFD rats were administered vehicle or a TRIPLE drug combination from GILEAD Sciences Inc., which consisted of an ACC inhibitor (ACCi; GS-834536) dosed at 10 mg/kg qd, an FXR agonist (FXRag, CILO) dosed at 30 mg/kg qd, and an ASK1 inhibitor (ASK1i, GS-4793) dosed in chow at 0.03% (*55*–*57*). At the appropriate time of the experiment, rats were dosed intravenously with 0.9 ml of a mannitol/sodium containing 0.41 nmol of each of 19 PEGylated substrates (GBTS-NASH). After injection with GBTS-NASH, rats were placed in metabolic cages to allow facile urine recovery. Bladders were voided 120 min after GBTS-NASH administration, and all urine produced 120 to 180 after GBTS-NASH injection was collected and pooled together (between 500 and 1000 μl of urine on average).

Peptide synthesis and conjugations

All peptides were synthesized by CPC Scientific Inc. For fluorogenic assays, peptides were flanked with the fluorophore-quencher pair 5-FAM-CPQ2. For in vivo urine experiments, peptides were barcoded with Glu-Fib–derived and heavy isotope–labeled peptides. PEGylated peptides were pooled prepared before in vivo experiments and stored at 4°C in phosphate-buffered saline. Typically, 0.125 nmol per PEGylated peptide in a total volume of 100 μl of sterile saline was injected per mouse or 0.41 nmol in 0.9 ml per rat.

Statistical analysis

Analyses were performed using Python 3.6.9 (scikit-learn 0.21.2, statsmodels 0.11.1, numpy 1.18.5, pandas 1.0.4) or GraphPad Prism. Significance testing was performed using either Student's t test or one-way ANOVA for normally distributed data, or Mann-Whitney U test for data that were not normally distributed. Normality was

checked using the Shapiro-Wilk test ($\alpha = 0.05$). Multiple testing correction was performed using the Bonferroni method. t-SNE was calculated on fluorescence change in 159 fluorescence resonance energy transfer–based protease substrates when incubated with NASH proteases. Logistic regression was used for gene or MS urine signal [using mean-normalized peak area ratio (PAR) values]. Classification performance was estimated through cross-validation with randomized 80% training and 20% validation splits, as the classifier is trained on 80% of the data and then blindly applied to the remaining 20%. This process was repeated 100 times with randomized training/validation splits. The classification performance is reported as the mean across the validation sets. AUCs were compared using the bootstrap method with 10,000 permutations. Diagnostic cutoffs were chosen to balance specificity and sensitivity using a separate cross-validation loop to avoid overfitting.

SUPPLEMENTARY MATERIALS

www.science.org/doi/10.1126/scitranslmed.abe8939
Supplementary Materials and methods
Figs. S1 to S16
Tables S1 to S3
Data file S1
Reference (*64*)

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. Z. Younossi, Q. M. Anstee, M. Marietti, T. Hardy, L. Henry, M. Eslam, J. George, E. Bugianesi, Global burden of NAFLD and NASH: Trends, predictions, risk factors and prevention. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 11–20 (2018).
2. Q. M. Anstee, H. L. Reeves, E. Kotsiliti, O. Govaere, M. Heikenwalder, From NASH to HCC: Current concepts and future challenges. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 411–428 (2019).
3. C. Fierbinteanu-Braticevici, M. Purcarea, Non-biopsy methods to determine hepatic fibrosis. *J. Med. Life* **2**, 401–406 (2009).
4. P. Bedossa, F. Carrat, Liver biopsy: The best, not the gold standard. *J. Hepatol.* **50**, 1–3 (2009).
5. A. Regev, M. Berho, L. J. Jeffers, C. Milikowski, E. G. Molina, N. T. Pyrsopoulos, Z. Z. Feng, K. R. Reddy, E. R. Schiff, Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am. J. Gastroenterol.* **97**, 2614–2618 (2002).
6. M. E. Rinella, Z. Lominadze, R. Loomba, M. Charlton, B. A. Neuschwander-Tetri, S. H. Caldwell, K. Kowdley, S. A. Harrison, Practice patterns in NAFLD and NASH: Real life differs from published guidelines. *Therap. Adv. Gastroenterol.* **9**, 4–12 (2016).
7. M. S. Siddiqui, S. A. Harrison, M. F. Abdelmalek, Q. M. Anstee, P. Bedossa, L. Castera, L. Dimick-Santos, S. L. Friedman, K. Greene, D. E. Kleiner, S. Megnien, B. A. Neuschwander-Tetri, V. Ratziu, E. Schabel, V. Miller, A. J. Sanyal; Liver Forum Case Definitions Working Group, Case definitions for inclusion and analysis of endpoints in clinical trials for nonalcoholic steatohepatitis through the lens of regulatory science. *Hepatology* **67**, 2001–2012 (2018).
8. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools). Silver Spring (MD): Food and Drug Administration (US); 2016 (2016).
9. C. Caussy, M. H. Alquiraish, P. Nguyen, C. Hernandez, S. Cepin, L. E. Fortney, V. Ajmera, R. Bettencourt, S. Collier, J. Hooker, E. Sy, E. Rizo, L. Richards, C. B. Sirlin, R. Loomba, Optimal threshold of controlled attenuation parameter with MRI-PDFF as the gold standard for the detection of hepatic steatosis. *Hepatology* **67**, 1348–1359 (2018).
10. P. Kennedy, M. Wagner, L. Castéra, C. W. Hong, C. L. Johnson, C. B. Sirlin, B. Taouli, Quantitative elastography methods in liver disease: Current evidence and future directions. *Radiology* **286**, 738–763 (2018).
11. A. E. Feldstein, A. Wieckowska, A. R. Lopez, Y. C. Liu, N. N. Zein, A. J. McCullough, Cytokeratin-18 fragment levels as noninvasive biomarkers for nonalcoholic steatohepatitis: A multicenter validation study. *Hepatology* **50**, 1072–1078 (2009).
12. M. Boyle, D. Tiniakos, J. M. Schattenberg, V. Ratziu, E. Bugianessi, S. Petta, C. P. Oliveira, O. Govaere, R. Younes, S. McPherson, P. Bedossa, M. J. Nielsen, M. Karsdal, D. Leeming, S. Kendrick, Q. M. Anstee, Performance of the PRO-C3 collagen neo-epitope biomarker in non-alcoholic fatty liver disease. *JHEP Rep.* **1**, 188–198 (2019).
13. E. Vilar-Gomez, N. Chalasani, Non-invasive assessment of non-alcoholic fatty liver disease: Clinical prediction rules and blood-based biomarkers. *J. Hepatol.* **68**, 305–315 (2018).
14. H. K. Drescher, S. Weiskirchen, R. Weiskirchen, Current status in testing for Nonalcoholic Fatty Liver Disease (NAFLD) and Nonalcoholic Steatohepatitis (NASH). *Cell* **8**, 845 (2019).
15. J. Kryczka, J. Boncela, Proteases revisited: Roles and therapeutic implications in fibrosis. *Mediators Inflamm.* **2017**, 1–14 (2017).
16. T. Houben, Y. Oligschlaeger, T. Hendrikx, A. V. Bitorina, S. M. A. Walenbergh, P. J. van Gorp, M. J. J. Gijbels, S. Friedrichs, J. Plat, F. G. Schaap, D. Lütjohann, M. H. Hofker, R. Shiri-Sverdlov, Cathepsin D regulates lipid metabolism in murine steatohepatitis. *Sci. Rep.* **7**, 3494 (2017).
17. K. Uchimura, M. Hayata, T. Mizumoto, Y. Miyasato, Y. Kakizoe, J. Morinaga, T. Onoue, R. Yamazoe, M. Ueda, M. Adachi, T. Miyoshi, N. Shiraishi, W. Ogawa, K. Fukuda, T. Kondo, T. Matsumura, E. Araki, K. Tomita, K. Kitamura, The serine protease prostasin regulates hepatic insulin sensitivity by modulating TLR4 signalling. *Nat. Commun.* **5**, 3428 (2014).
18. M. A. Sánchez-Garrido, K. M. Habegger, C. Clemmensen, C. Holleman, T. D. Müller, D. Perez-Tilve, P. Li, A. S. Agrawal, B. Finan, D. J. Drucker, M. H. Tschöp, R. D. DiMarchi, A. Kharitonenkov, Fibroblast activation protein (FAP) as a novel metabolic target. *Mol. Metab.* **5**, 1015–1024 (2016).
19. E. J. Toonen, A. M. Mirea, C. J. Tack, R. Stienstra, D. B. Ballak, J. A. van Diepen, A. Hijmans, T. Chavakis, W. H. Dokter, C. T. Pham, M. G. Netea, C. A. Dinarello, L. A. Joosten, Activation of proteinase 3 contributes to Nonalcoholic Fatty Liver Disease and insulin resistance. *Mol. Med.* **22**, 202–214 (2016).
20. S. M. Walenbergh, T. Houben, S. S. Rensen, V. Bieghs, T. Hendrikx, P. J. van Gorp, Y. Oligschlaeger, M. L. Jeurissen, M. J. Gijbels, W. A. Buurman, A. C. Vreugdenhil, J. W. Greve, J. Plat, M. H. Hofker, S. Kalhan, J. Pihlajamäki, P. Lindsey, G. H. Koek, R. Shiri-Sverdlov, Plasma cathepsin D correlates with histological classifications of fatty liver disease in adults and responds to intervention. *Sci. Rep.* **6**, 38278 (2016).
21. W. Ando, H. Yokomori, N. Tsutsui, E. Yamanouchi, Y. Suzuki, M. Oda, Y. Inagaki, K. Otori, I. Okazaki, Serum matrix metalloproteinase-1 level represents disease activity as opposed to fibrosis in patients with histologically proven nonalcoholic steatohepatitis. *Clin. Mol. Hepatol.* **24**, 61–76 (2018).
22. K. M. Irvine, L. F. Wockner, I. Hoffmann, L. U. Horsfall, K. J. Fagan, V. Bijin, B. Lee, A. D. Clouston, G. Lampe, J. E. Connolly, E. E. Powell, Multiplex serum protein analysis identifies novel biomarkers of advanced fibrosis in patients with chronic liver disease with the potential to improve diagnostic accuracy of established biomarkers. *PLOS ONE* **11**, e0167001 (2016).
23. F. D'Amico, M. Consolo, A. Amoroso, E. Skarmoutsou, B. Mauceri, F. Stivala, G. Malaponte, G. Bertino, S. Neri, M. C. Mazzarino, Liver immunolocalization and plasma levels of MMP-9 in non-alcoholic steatohepatitis (NASH) and hepatitis C infection. *Acta Histochem.* **112**, 474–481 (2010).
24. H. Toyoda, T. Kumada, S. Kiriyama, M. Tanikawa, Y. Hisanaga, A. Kanamori, T. Tada, Y. Murakami, Higher hepatic gene expression and serum levels of matrix metalloproteinase-2 are associated with steatohepatitis in non-alcoholic fatty liver diseases. *Biomarkers* **18**, 82–87 (2013).
25. D. Ljumovic, I. Diamantis, A. K. Alegakis, E. A. Kouroumalis, Differential expression of matrix metalloproteinases in viral and non-viral chronic liver diseases. *Clin. Chim. Acta* **349**, 203–211 (2004).
26. Q. D. Mac, D. V. Mathews, J. A. Kahla, C. M. Stoffers, O. M. Delmas, B. A. Holt, A. B. Adams, G. A. Kwong, Non-invasive early detection of acute transplant rejection via nanosensors of granzyme B activity. *Nat. Biomed. Eng.* **3**, 281–291 (2019).
27. J. D. Kirkpatrick, A. D. Warren, A. P. Soleimany, P. M. K. Westcott, J. C. Voog, C. Martin-Alonso, H. E. Fleming, T. Tammela, T. Jacks, S. N. Bhatia, Urinary detection of lung cancer in mice via noninvasive pulmonary protease profiling. *Sci. Transl. Med.* **12**, eaaw0262 (2020).
28. G. A. Kwong, J. S. Dudani, E. Carrodeguas, E. V. Mazumdar, S. M. Zekavat, S. N. Bhatia, Mathematical framework for activity-based cancer biomarkers. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12627–12632 (2015).
29. G. A. Kwong, G. von Maltzahn, G. Murugappan, O. Abudayyeh, S. Mo, I. A. Papayannopoulos, D. Y. Sverdlov, S. B. Liu, A. D. Warren, Y. Popov, D. Schuppan, S. N. Bhatia, Mass-encoded synthetic biomarkers for multiplexed urinary monitoring of disease. *Nat. Biotechnol.* **31**, 63–70 (2013).
30. M. Ryaboshapkina, M. Hammar, Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Sci. Rep.* **7**, 12361 (2017).
31. P. Angulo, D. E. Kleiner, S. Dam-Larsen, L. A. Adams, E. S. Bjornsson, P. Charatcharoenwitthaya, P. R. Mills, J. C. Keach, H. D. Lafferty, A. Stahler, S. Haflidadottir, F. Bendtsen, Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients With nonalcoholic fatty liver disease. *Gastroenterology* **149**, 389–397.e10 (2015).
32. T. Wong, R. J. Wong, R. G. Gish, Diagnostic and treatment implications of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. *Gastroenterol. Hepatol.* **15**, 83–89 (2019).
33. O. Govaere, S. Cockell, D. Tiniakos, R. Queen, R. Younes, M. Vacca, L. Alexander, F. Ravaioli, J. Palmer, S. Petta, J. Boursier, C. Rosso, K. Johnson, K. Wonders, C. P. Day, M. Ekstedt, M. Orešič, R. Darlay, H. J. Cordell, F. Marra, A. Vidal-Puig, P. Bedossa, J. M. Schattenberg, K. Clément, M. Allison, E. Bugianesi, V. Ratziu, A. K. Daly, Q. M. Anstee, Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci. Transl. Med.* **12**, eaba4448 (2020).

34. K. B. Bang, Y. K. Cho, Comorbidities and metabolic derangement of NAFLD. *J. Lifestyle Med.* **5**, 7–13 (2015).
35. V. Pandyarajan, R. G. Gish, N. Alkhouri, M. Noureddin, Screening for nonalcoholic fatty liver disease in the primary care clinic. *Gastroenterol. Hepatol.* **15**, 357–365 (2019).
36. P. J. Eddowes, M. Sasso, M. Allison, E. Tsochatzis, Q. M. Anstee, D. Sheridan, I. N. Guha, J. F. Cobbold, J. J. Deeks, V. Paradis, P. Bedossa, P. N. Newsome, accuracy of fibroscan controlled attenuation parameter and liver stiffness measurement in assessing steatosis and fibrosis in patients with nonalcoholic fatty liver disease. *Gastroenterology* **156**, 1717–1730 (2019).
37. M. Swierczewska, K. C. Lee, S. Lee, What is the future of PEGylated therapies? *Expert Opin. Emerg. Drugs* **20**, 531–536 (2015).
38. J. S. Dudani, M. Ibrahim, J. Kirkpatrick, A. D. Warren, S. N. Bhatia, Classification of prostate cancer using a protease activity nanosensor library. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8954–8959 (2018).
39. M. Matsumoto, N. Hada, Y. Sakamaki, A. Uno, T. Shiga, C. Tanaka, T. Ito, A. Katsume, M. Sudoh, An improved mouse model that rapidly develops fibrosis in non-alcoholic steatohepatitis. *Int. J. Exp. Pathol.* **94**, 93–103 (2013).
40. E. M. Brunt, D. E. Kleiner, L. A. Wilson, P. Belt, B. A. Neuschwander-Tetri, N. C. R. Network, Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: Distinct clinicopathologic meanings. *Hepatology* **53**, 810–820 (2011).
41. R. N. Helsley, V. Varadharajan, A. L. Brown, A. D. Gromovsky, R. C. Schugar, I. Ramachandiran, K. Fung, M. N. Kabbany, R. Banerjee, C. K. Neumann, C. Finney, P. Pathak, D. Orabi, L. J. Osborn, W. Massey, R. Zhang, A. Kadam, B. E. Sansbury, C. Pan, J. Sacks, R. G. Lee, R. M. Crooke, M. J. Graham, M. E. Lemieux, V. Gogonea, J. P. Kirwan, D. S. Allende, M. Civelek, P. L. Fox, L. L. Rudel, A. J. Lusis, M. Spite, J. M. Brown, Obesity-linked suppression of membrane-bound. *eLife* **8**, e49882 (2019).
42. C. D. Byrne, G. Targher, NAFLD as a driver of chronic kidney disease. *J. Hepatol.* **72**, 785–801 (2020).
43. L. J. Stallons, R. M. Whitaker, R. G. Schnellmann, Suppressed mitochondrial biogenesis in folic acid-induced acute kidney injury and early fibrosis. *Toxicol. Lett.* **224**, 326–332 (2014).
44. M. Rabe, F. Schaefer, Non-transgenic mouse models of kidney diseases. *Nephron* **133**, 53–61 (2016).
45. K. B. Russ, T. M. Stevens, A. K. Singal, Acute kidney injury in patients with cirrhosis. *J. Clin. Transl. Hepatol.* **3**, 195–204 (2015).
46. K. L. Hudkins, W. Pichaiwong, T. Wietecha, J. Kowalewska, M. C. Banas, M. W. Spencer, A. Mühlfeld, M. Koelling, J. W. Pippin, S. J. Shankland, B. Askari, M. E. Rabaglia, M. P. Keller, A. D. Attie, C. E. Alpers, BTBR Ob/Ob mutant mice model progressive diabetic nephropathy. *J. Am. Soc. Nephrol.* **21**, 1533–1542 (2010).
47. A. Ericsson, P. Tonelius, M. Lal, A. Sabirsh, G. Böttcher, L. William-Olsson, M. Strömstedt, C. Johansson, G. Hyberg, S. Tapani, A. C. Jönsson-Rylander, R. Unwin, The effects of dual PPARα/γagonism compared with ACE inhibition in the BTBRob/ob mouse model of diabetes and diabetic nephropathy. *Physiol. Rep.* **5**, (2017).
48. D. E. Kleiner, H. R. Makhlouf, Histology of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis in adults and children. *Clin. Liver Dis.* **20**, 293–312 (2016).
49. M. A. T. Han, O. Altayar, S. Hamdeh, V. Takyar, Y. Rotman, O. Etzion, E. Lefebvre, R. Safadi, V. Ratziu, L. J. Prokop, M. H. Murad, M. Noureddin, Rates of and factors associated with placebo response in trials of pharmacotherapies for nonalcoholic steatohepatitis: Systematic review and meta-analysis. *Clin. Gastroenterol. Hepatol.* **17**, 616–629.e26 (2019).
50. K. Patel, S. A. Harrison, M. Elkhashab, J. F. Trotter, R. Herring, S. E. Rojter, Z. Kayali, V. W. Wong, S. Greenbloom, S. Jayakumar, M. L. Shiffman, B. Freilich, E. J. Lawitz, E. J. Gane, E. Harting, J. Xu, A. N. Billin, C. Chung, C. S. Djedjos, G. M. Subramanian, R. P. Myers, M. S. Middleton, M. Rinella, M. Noureddin, Cilofexor, a nonsteroidal FXR agonist, in patients with noncirrhotic NASH: A phase 2 randomized controlled trial. *Hepatology* **72**, 58–71 (2020).
51. N. Alkhouri, E. Lawitz, M. Noureddin, R. DeFronzo, G. I. Shulman, GS-0976 (Firsocostat): An investigational liver-directed acetyl-CoA carboxylase (ACC) inhibitor for the treatment of non-alcoholic steatohepatitis (NASH). *Expert Opin. Investig. Drugs* **29**, 135–141 (2020).
52. R. Loomba, E. Lawitz, P. S. Mantry, S. Jayakumar, S. H. Caldwell, H. Arnold, A. M. Diehl, C. S. Djedjos, L. Han, R. P. Myers, G. M. Subramanian, J. G. McHutchison, Z. D. Goodman, N. H. Afdhal, M. R. Charlton; GS-US-384-1497 Investigators, The ASK1 inhibitor selonsertib in patients with nonalcoholic steatohepatitis: A randomized, phase 2 trial. *Hepatology* **67**, 549–559 (2017).
53. P. Dibba, A. A. Li, B. J. Perumpail, N. John, S. Sallam, N. D. Shah, W. Kwong, G. Cholankeril, D. Kim, A. Ahmed, Emerging therapeutic targets and experimental drugs for the treatment of NAFLD. *Diseases* **6**, 83 (2018).
54. J. Bates, H. David, Z. Anna, R. B. Grant, T. L. John, L. Henry, L. Kathy, R. B. Saritha Kusam, N. David, M. Igor, D. G. Breckenridge, Combination of ASK1 and ACC inhibitors increases efficacy in rodent models of NASH. *Hepatology* **66**, 1149–1185 (2017).
55. J. Bates, K. Liu, D. Hollenback, A. Zagorska, G. Budas, J. T. Liles, H. Liu, S. Kusam, R. Brockett, D. Newstrom, I. Mikaelian, T. Wang, A. S. Ray, D. G. Breckenridge, Combination of an FXR agonist and an acc inhibitor increases antifibrotic efficacy in rodent models of NASH. *J. Hepatol.* **68**, S395–S396 (2018).
56. J. Bates, S. Nandita, R. Ricardo, L. David, W. Kari, H. David, L. Kathy, Z. Anna, B. Grant, L. John, B. Robert, N. David, M. Igor, L. Li, X. Ren, B. David, Hepatic metabolomics and plasma MicroRNA analysis of combinations of an ASK1 inhibitor, an ACC inhibitor, and an FXR agonist in a rat choline-deficient high fat diet model reveal reductions in oxidative stress, inflammation and fibrosis. *Hepatology* **68**, S773A (2018).
57. N. H. Afdhal, M. Curry, Technology evaluation: A critical step in the clinical utilization of novel diagnostic tests for liver fibrosis. *J. Hepatol.* **46**, 543–545 (2007).
58. E. Roeb, Matrix metalloproteinases and liver fibrosis (translational aspects). *Matrix Biol.* **68-69**, 463–473 (2018).
59. A. Moles, N. Tarrats, J. C. Fernández-Checa, M. Marí, Cathepsins B and D drive hepatic stellate cell proliferation and promote their fibrogenic potential. *Hepatology* **49**, 1297–1307 (2009).
60. A. Asgharpour, S. C. Cazanave, T. Pacana, M. Seneshaw, R. Vincent, B. A. Banini, D. P. Kumar, K. Daita, H. K. Min, F. Mirshahi, P. Bedossa, X. Sun, Y. Hoshida, S. V. Koduru, D. Contaifer Jr., U. O. Warncke, D. S. Wijesinghe, A. J. Sanyal, A diet-induced animal model of non-alcoholic fatty liver disease and hepatocellular cancer. *J. Hepatol.* **65**, 579–588 (2016).
61. S. Singh, A. M. Allen, Z. Wang, L. J. Prokop, M. H. Murad, R. Loomba, Fibrosis progression in nonalcoholic fatty liver vs nonalcoholic steatohepatitis: A systematic review and meta-analysis of paired-biopsy studies. *Clin. Gastroenterol. Hepatol.* **13**, 643–654. e641–649; quiz e639–640 (2015).
62. A. G. Shah, A. Lydecker, K. Murray, B. N. Tetri, M. J. Contos, A. J. Sanyal, N. C. R. Network, Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clin. Gastroenterol. Hepatol.* **7**, 1104–1112 (2009).
63. M. H. Veldman-Jones, R. Brant, C. Rooney, C. Geh, H. Emery, C. G. Harbron, M. Wappett, A. Sharpe, M. Dymond, J. C. Barrett, E. A. Harrington, G. Marshall, Evaluating robustness and sensitivity of the nanostring technologies ncounter platform to enable multiplexed gene expression analysis of clinical samples. *Cancer Res.* **75**, 2587–2593 (2015).
64. P. Bedossa; FLIP Pathology Consortium, Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. *Hepatology* **60**, 565–575 (2014).

Acknowledgments: We thank G. M. Subramanian and R. P. Myers for supporting the collaborative effort with Gilead Inc.; GILEAD employees K. Walker, K. Liu, and D. Lopez for support with the rat CDAHFD model; P. Bedossa (Liverpat, Paris, France) for technical help with histology data analysis; P. Qiu (Georgia Tech, Atlanta, GA) for critical help in computational data analysis; A. Mancino from Syneos Health-Inventiv Health (Princeton, NJ) for help in mass spectrometry analysis; M. Oudin (Tufts University, MA) for helpful discussions; A. Serer for help in GBTS-NASH preparation; and M. Rowe for help in graph editing. **Funding:** This study was supported in part by Small Business Innovation Research grant 4R44DK116635-02 from the NIH. G.A.K. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund. S.B. is an HHMI investigator. R.T.C. was supported by the MGH Research Scholars Program. Q.M.A., A.K.D., and O.G. are supported by the EPoS (Elucidating Pathways of Steatohepatitis) consortium funded by the Horizon 2020 Framework Program of the European Union under grant agreement 634413, the LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis) consortium funded by the Innovative Medicines Initiative (IMI2) Program of the European Union under grant agreement 777377, and the Newcastle NIHR Biomedical Research Centre. Q.M.A. and V.R. are European NAFLD Registry investigators. **Author contributions**: S.C.C., A.D.W., E.K.H., S.B., and G.A.K. designed research. S.C.C., A.D.W., F.T., S.S., and M.C. performed research. A.Z., J.B., A.N.B., J.T.L., G.R.B., D.G.B., D.T., V.R., A.K.D., O.G., Q.M.A., L.G., J.L., R.T.C., and K.E.C. contributed to reagents/analytical tools. S.C.C., A.D.W., M.P. and F.T., A.Z., N.G., W.W., S.B., and G.A.K. analyzed data. S.C.C., M.P., F.T., N.G., and G.A.K. wrote the paper. All authors revised the manuscript for intellectual content and approved the final draft. **Competing interests:** S.C.C., M.P., F.T., M.C., S.I., and W.W. are all current employees and shareholders of Glympse Bio Inc. E.K.H., A.D.W., S.C.C., S.B., and G.A.K are listed as inventors on patent applications pertaining to the results of the paper. S.B. is a director at Vertex; is a cofounder and consultant at Glympse Bio, Satellite Bio, and Impilo Therapeutics; is a consultant for Cristal, Maverick, and Moderna; and receives sponsored research funds from Johnson & Johnson. G.A.K. is a cofounder of Glympse Bio and consults for Glympse Bio and Satellite Bio. This study could affect his personal financial status. The terms of this arrangement have been reviewed and approved by Georgia Tech in accordance with its conflict-of-interest policies. A.D.W., S.S., and E.K.H. were employed by Glympse Bio at the time of data generation and have equity in Glympse Bio Inc. A.D.W. is also a consultant at Exact Sciences Corp and an advisor to Ovation.io Inc. A.Z., J.B., A.N.B., J.T.L., G.R.B., and D.G.B. are all current employees and shareholders of GILEAD Science Inc. Q.M.A. reports Research Grant Funding: Abbvie, Allergan/Tobira, AstraZeneca, GlaxoSmithKline, Glympse Bio, Novartis Pharma AG, Pfizer Ltd., and Vertex. Active Research Collaborations (including research collaborations supported through the EU IMI2 LITMUS Consortium): Abbvie, Antaros Medical, Allergan/Tobira*, AstraZeneca, BMS, Boehringer Ingelheim International GMBH, Echosens,

Ellegaard Gottingen Minipigs AS, Eli Lilly & Company Ltd., Exalenz Bioscience Ltd., Genfit SA, Glympse Bio, GlaxoSmithKline, HistoIndex, Intercept Pharma Europe Ltd., iXscient Ltd., Nordic Bioscience, Novartis Pharma AG, Novo Nordisk A/S, One Way Liver Genomics SL, Perspectum Diagnostics, Pfizer Ltd., Resoundant, Sanofi-Aventis Deutschland GMBH, SomaLogic Inc., and Takeda Pharmaceuticals International SA. Consultancy: 89Bio, Abbott Laboratories, Acuitas Medical, Allergan/Tobira, Altimmune, AstraZeneca, Axcella, Blade, BMS, BNN Cardio, Celgene, Cirius, CymaBay, EcoR1, E3Bio, Eli Lilly & Company Ltd., Galmed, Genentech, Genfit SA, Gilead, Grunthal, HistoIndex, Indalo, Imperial Innovations, Intercept Pharma Europe Ltd., Inventiva, IQVIA, Janssen, Madrigal, MedImmune, Metacrine, NewGene, NGMBio, North Sea Therapeutics, Novartis, Novo Nordisk A/S, PathAI, Pfizer Ltd., Poxel, ProSciento, Raptor Pharma, Servier, Terns, and Viking Therapeutics. Speaker: Abbott Laboratories, Allergan/Tobira, BMS, Clinical Care Options, Falk, Fishawack, Genfit SA, Gilead, Integritas Communications, Kenes, and MedScape. Royalties: Elsevier Ltd. D.T. received consultation fees from Intercept Pharmaceuticals Inc., Allergan plc, Cirius Therapeutics Inc., Alimentiv Inc., and Clinnovate Health UK Ltd. and an educational grant from Histoindex Pte. **Data and materials availability:** All data associated with this study are present in the paper or the Supplementary Materials. Raw data are available in data file S1. Restrictions over access to raw data may apply due to material transfer agreements (MTAs) established between Glympse Bio, Inc. and Newcastle University and Gilead Sciences Inc. NanoString data generated in this study from liver biopsies of MGH/STM patients have been deposited into the Gene Expression Omnibus (GEO) database (GSE163211).

Submitted 22 November 2020
Resubmitted 28 March 2021
Accepted 9 August 2021
Published 20 October 2021
10.1126/scitranslmed.abe8939

Citation: S. C. Cazanave, A. D. Warren, M. Pacula, F. Touti, A. Zagorska, N. Gural, E. K. Huang, S. Sherman, M. Cheema, S. Ibarra, J. Bates, A. N. Billin, J. T. Liles, G. R. Budas, D. G. Breckenridge, D. Tiniakos, V. Ratziu, A. K. Daly, O. Govaere, Q. M. Anstee, L. Gelrud, J. Luther, R. T. Chung, K. E. Corey, W. Winckler, S. Bhatia, G. A. Kwong, Peptide-based urinary monitoring of fibrotic nonalcoholic steatohepatitis by mass-barcoded activity-based sensors. *Sci. Transl. Med.* **13**, eabe8939 (2021).

Science Translational Medicine

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

Sophie C. CazanaveAndrew D. WarrenMaciej PaculaFayçal ToutiAnna ZagorskaNil GuralEric K. HuangSarah ShermanMehar CheemaSabrina IbarraJamie BatesAndrew N. BillinJohn T. LilesGrant R. BudasDavid G. BreckenridgeDina TiniakosVlad RatziuAnn K. DalyOlivier GovaereQuentin M. AnsteeLouis GelrudJay LutherRaymond T. ChungKathleen E. CoreyWendy WincklerSangeeta BhatiaGabriel A. Kwong

Sci. Transl. Med., 13 (616), eabe8939.

A urinary readout of NASH fibrosis

Biopsy-free assessment of nonalcoholic steatohepatitis (NASH) fibrosis is currently lacking. Cazanave *et al*. showed that the gene expression of 13 liver proteases is associated with the presence of NASH in human RNA-seq data. They accordingly developed a set of 19 peptides that release identifiable “barcodes” in the presence of these proteases and which are readily detectable in urine. Parenteral administration of the peptides to a diet-induced rat model of NASH allowed discrimination of F2 stage fibrosis from earlier stages. A reduction in liver fibrosis after 1 week of treatment was also detected in the animals demonstrating that this technique allows preclinical NASH assessment without heavily invasive techniques.

Science Translational Medicine (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Translational Medicine* is a registered trademark of AAAS.