UCLA
UCLA Previously Published Works

Title
Metabolomic-based noninvasive serum test to diagnose nonalcoholic steatohepatitis: Results from discovery and validation cohorts

Permalink
https://escholarship.org/uc/item/8421g0p6

Journal
Hepatology Communications, 2(7)

ISSN
2471-254X

Authors
Mayo, Rebeca
Crespo, Javier
Martínez-Arranz, Ibon
et al.

Publication Date
2018-07-01

DOI
10.1002/hep4.1188

Peer reviewed

Metabolomic-Based Noninvasive Serum Test to Diagnose Nonalcoholic Steatohepatitis: Results From Discovery and Validation Cohorts

Rebeca Mayo,[1*] Javier Crespo,[2*] Ibon Martínez-Arranz,[1*] Jesus M Banales,[3] Mayte Arias,[2] Itziar Mincholé,[1]

Rocío Aller de la Fuente,[4] Raúl Jimenez-Agüero,[3] Cristina Alonso,[1] Daniel A. de Luis,[5] Libor Vitek,[6] Jan Stritesky,[6]

Joan Caballería,[7] Manuel Romero-Gómez,[8] Antonio Martín-Duce,[9] Jose Maria Mugüerza Huguet,[10]

José Ignacio Busteros-Moraza,[11] Michael O. Idowu,[12] Azucena Castro,[1] M. Luz Martínez-Chantar,[13] Pablo Ortiz,[1]

Radan Bruha,[6] Shelly C. Lu,[14] Pierre Bedossa,[15] Mazen Noureddin,[14] Arun J. Sanyal,[16] and José M. Mato[13]

Nonalcoholic fatty liver disease (NAFLD) is the most common type of chronic liver disease worldwide and includes a broad spectrum of histologic phenotypes, ranging from simple hepatic steatosis or nonalcoholic fatty liver (NAFL) to nonalcoholic steatohepatitis (NASH). While liver biopsy is the reference gold standard for NAFLD diagnosis and staging, it has limitations due to its sampling variability, invasive nature, and high cost. Thus, there is a need for noninvasive biomarkers that are robust, reliable, and cost effective. In this study, we measured 540 lipids and amino acids in serum samples from biopsy-proven subjects with normal liver (NL), NAFL, and NASH. Using logistic regression analysis, we identified two panels of triglycerides that could first discriminate between NAFLD and NL and second between NASH and NAFL. These noninvasive tests were compared to blinded histology as a reference standard. We performed these tests in an original cohort of 467 patients with NAFLD (90 NL, 246 NAFL, and 131 NASH) that was subsequently validated in a separate cohort of 192 patients (7 NL, 109 NAFL, 76 NASH). The diagnostic performances of the validated tests showed an area under the receiver operating characteristic curve, sensitivity, and specificity of 0.88 ± 0.05, 0.94, and 0.57, respectively, for the discrimination between NAFLD and NL and 0.79 ± 0.04, 0.70, and 0.81, respectively, for the discrimination between NASH and NAFL. When the analysis was performed excluding patients with glucose levels >136 mg/dL, the area under the receiver operating characteristic curve for the discrimination between NASH and NAFL increased to 0.81 ± 0.04 with sensitivity and specificity of 0.73 and 0.80, respectively. *Conclusion:* The assessed noninvasive lipidomic serum tests distinguish between NAFLD and NL and between NASH and NAFL with high accuracy. (*Hepatology Communications* 2018;2:807-820)

Introduction

Nonalcoholic fatty liver disease (NAFLD) represents a spectrum of liver diseases that ranges from simple triglyceride (TG) accumulation (nonalcoholic fatty liver [NAFL]) to fat accumulation with inflammation and cellular injury (nonalcoholic steatohepatitis [NASH]) that ultimately may progress to fibrosis, cirrhosis, and hepatocellular carcinoma.[1,2] During the last 3 decades, the incidence of NAFLD has substantially increased. It is now the most common chronic disease of the liver, with a prevalence between 10% and 40% in Western countries. Approximately 10%-30% of patients with NAFLD progress to

NASH.[3,4] Distinguishing between NASH and NAFL and ascertaining the stage of fibrosis are critical for distinguishing between a generally benign condition and one with increased morbidity and mortality.[5] Currently, liver biopsy remains the gold standard for NAFLD diagnosis and staging. However, it is an invasive procedure and is limited by sampling error, high cost, procedure-related complications, and observer variability, even when performed by expert pathologists.[6,7] Therefore, there is an unmet need for noninvasive biomarkers that are robust, reliable, and cost effective that

will allow practitioners not only to diagnose and stage NAFLD but also to monitor NAFLD progression. Such biomarkers could also be a useful addition to the current design of NASH clinical trials.

Magnetic resonance imaging proton density fat fraction (MRI-PDFF) and magnetic resonance elastography (MRE) have emerged as accurate tools for quantifying steatosis[8-10] and fibrosis[11-13]; however, they have limitations in detecting inflammation, ballooning, and cellular injury. The latter are key components in NASH diagnosis and part of the histologic

ARTICLE INFORMATION:

From the [1]OWL Metabolomics, Bizkaia Technology Park, Derio, Spain; [2]Department of Digestive Disease, Marqués de Valdecilla University Hospital, Cantabaria University, Research Institute Marqués de Valdecilla (IDIVAL), Santander, Spain; [3]Department of Liver and Gastrointestinal Diseases, Biodonostia Research Institute, Donostia University Hospital, University of the Basque Country (UPV-EHU), CIBERehd, IKERBASQUE, Donostia, Spain; [4]Department of Digestive Disease, Clinic University Hospital, University of Valladolid, Valladolid, Spain; [5]Center of Investigation of Endocrinology and Clinical Nutrition, Medicine School and Department of Endocrinology and Nutrition, Clinic University Hospital, University of Valladolid, Valladolid, Spain; [6]Faculty General Hospital and the First Faculty of Medicine, Charles University, Prague, Czech Republic; [7]Liver Unit, Hospital Clinic, CIBERehd, IDIBAPS, Barcelona, Spain; [8]Unit for the Clinical Management of Digestive Diseases and CIBERehd, Valme University Hospital, Seville, Spain; [9]Faculty of Medicine and Health Science; [10]Surgery Department; and [11]Department of Pathology, University Hospital Príncipe de Asturias, Alcalá University, Madrid, Spain; [12]Division of Surgical Pathology, Department of Pathology, Virginia Commonwealth University Medical Center, Richmond, VA, United States; [13]CIC bioGUNE, CIBERehd, Bizkaia Technology Park, Derio, Bizkaia, Spain; [14]Division of Digestive and Liver Diseases, Cedars-Sinai Medical Center, Los Angeles, CA, United States; [15]Department of Pathology, University Hospitals Paris Nord Val de Seine, Beaujon, Hauts-de-Seine, Clichy, France; [16]Division of Gastroenterology and Hepatology, Virginia Commonwealth University Medical Center, Richmond, VA, United States.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

José M. Mato, Ph.D.
CIC bioGUNE, Parque Tecnológico de Bizkaia
48160 Derio, Spain

E-mail: director@cicbiogune.es
Tel: +34 946 572 517

criteria to monitor NASH progression in ongoing clinical trials.[14] Therefore, there is an active search for noninvasive NAFLD biomarkers measured mainly in blood. These biomarkers include metabolites (lipids, carbohydrates), peptides, proteins, and nucleic acids (DNA, messenger RNA, microRNA).[15-18] Although numerous possible biomarkers have been proposed, few if any have been shown to have good accuracy and/or have been validated against a large cohort of patients with biopsy-proven NAFLD.

Metabolomics provides a powerful technology for discovery of noninvasive biomarkers to improve diagnosis of complex diseases. Here, we applied liquid chromatography–mass spectrometry (LC-MS) metabolomics to identify serum biomarkers that differentiate between normal liver (NL) and NAFLD and between NASH and NAFL in a cohort of 467 patients with biopsy-proven NAFLD. These biomarkers were subsequently validated using serum samples from a separate cohort of 192 patients with biopsy-proven NAFLD.

Patients and Methods

PATIENT COHORTS

A total of 467 biopsied adult patients (90 with NL, 246 with diagnosis of NAFL, and 131 with diagnosis of NASH) seen at 11 participating hospitals formed the discovery cohort. The full description of this cohort has been described by Barr et al.[19] Patient characteristics are summarized in Table 1. For the validation group, an independent cohort of 239 biopsied adult patients were recruited in five hospitals (Marqués de Valdecilla University Hospital, Santander, Spain;

Donostia University Hospital, Donosti, Spain; Virgen de Valme University Hospital, Seville, Spain; University Clinic Hospital, Valladolid, Spain; and Faculty General Hospital and the First Faculty of Medicine, Charles University, Prague, Czech Republic). Principal component analysis (PCA) of the serum metabolomics data demonstrated that patients from both the discovery and validation cohorts clustered together (Fig. 1A). Likewise, PCA of the serum metabolomics data showed that patients of the validation cohort clustered together independently of the hospital of origin (Fig. 1B) and sex (Fig. 1C), as shown in the discovery cohort.[20] The inclusion criteria for both the discovery and validation cohorts were (1) age 18-70 years; (2) no known acute or chronic disease except for obesity or type 2 diabetes based on medical history, physical examination, and standard laboratory tests; and (3) alcohol consumption was less than 20 g/day for women and 30 g/day for men. Exclusion criteria included other causes of liver disease, such as viral hepatitis.

For all subjects, diagnoses were established histologically through liver biopsy specimens. Histologic slides were processed as usual, employing hematoxylin and eosin and Masson's trichrome stains. For the discovery group, the histologic diagnosis of NAFLD was established by a single liver pathologist in each participating hospital, according to the criteria defined by Kleiner et al.[21] and Brunt et al.[22] Following assessment, patients were classified by the pathologists into the following three histologic groups: (1) NL, (2) NAFL (hepatic steatosis alone), and (3) NASH (presence as determined by the pathologist). None of the patients had cirrhosis. For the validation study, initially 239 patients were recruited. One liver pathologist (M.O.I.) independently evaluated 178 biopsies; when there was disagreement with the original diagnosis given by

TABLE 1. CLINICOPATHOLOGIC CHARACTERISTICS OF THE DISCOVERY AND VALIDATION COHORTS

Histology	Discovery Cohort			Validation Cohort		
	NL	NAFL	NASH	NL	NAFL	NASH
n	90	246	131	7	109	76
Age (years)	44.8 ± 13.1	42.7 ± 12.2	45.1 ± 10.8	43.1 ± 11.0	45.0 ± 11.3	48.1 ± 12.4
Sex (Female %)	88.1*,†	70.4‡	74.6‡	71.4	51.4	56.6
BMI (kg/m²)	39.4 ± 13.3†	42.9 ± 11.4‡	45.9 ± 11.2	39.8 ± 13.2	40.2 ± 9.6	43.2 ± 8.6
AST (U/L)	21.2 ± 12.6*,†	35.2 ± 29.0	31.2 ± 33.0	25.6 ± 14.2	31.4 ± 18.6	38.1 ± 35.3
ALT (U/L)	19.0 ± 7.4 *,†,‡	46.7 ± 35.2	40.6 ± 30.8‡	38.4 ± 37.7	49.4 ± 40.0	52.1 ± 47.5
Total fasting cholesterol (mg/dL)	186.2 ± 32.5	203.4 ± 42.6	204.9 ± 43.1	170.4 ± 27.1	198.0 ± 41.6	206.2 ± 41.7
Fasting triglycerides (mg/dL)	120.3 ± 54.3	155.9 ± 98.4	141.2 ± 88.2	93.0 ± 49.6	167.6 ± 89.9	165.6 ± 83.8
Fasting glucose (mg/dL)	102.1 ± 34.7	109.5 ± 29.7	111.0 ± 38.1	91.4 ± 15.5	111.8 ± 32.5	119.7 ± 42.5

All diagnoses were established histologically in liver biopsy specimens. Data are expressed as mean ± SD. Significant differences between groups ($P < 0.05$) are represented by *NL–NAFLD, †NL–NASH, and ‡discovery-validation cohorts.
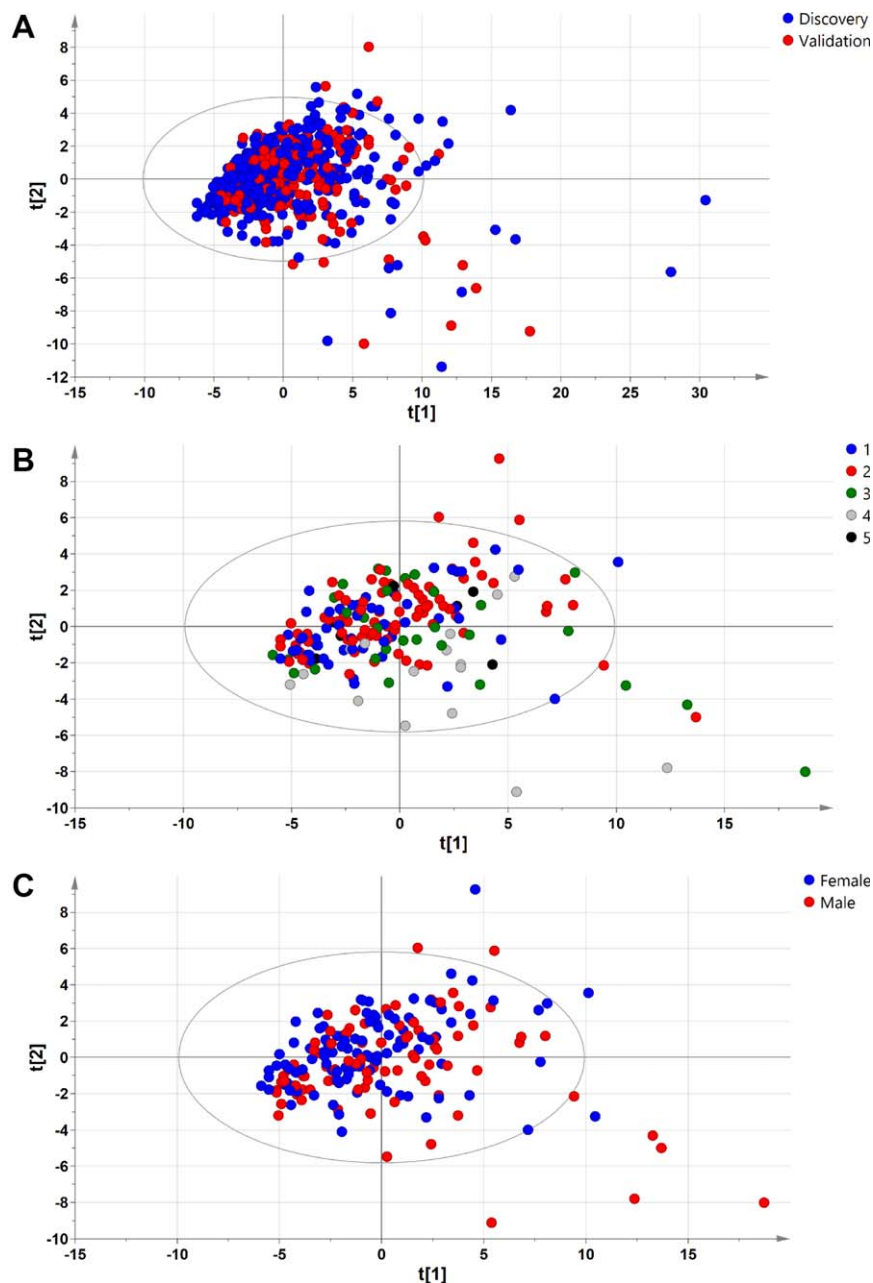
FIG. 1. PCA of samples included in the discovery and validation cohorts. (A) PCA analysis showed that patients clustered together in the discovery and validation cohorts. (B,C) PCA analysis also demonstrated that patients clustered together independently of the (B) hospital of origin and (C) sex in the validation cohort. Hospital of origin: 1, Marqués de Valdecilla University Hospital, Santander, Spain; 2, Biodonostia Research Institute-Donostia University Hospital, Donostia, Spain; 3, Clinic University Hospital, University of Valladolid, Valladolid, Spain; 4, Faculty General Hospital and the First Faculty of Medicine, Charles University, Prague, Czech Republic; 5, Virgen de Valme University Hospital, Seville, Spain.

the liver pathologist from each participating hospital, a third experienced NAFLD pathologist (P.B.) independently evaluated the biopsy and his diagnosis was considered. Biopsies that showed total disagreement among the three pathologists were excluded from the analysis. The histopathologic definition of NASH was determined by the joint presence of steatosis, lobular inflammation, and hepatocellular ballooning on liver specimens, independently of the total NAFLD activity score (NAS).[23] Biopsies with only steatosis, steatosis

with inflammation but without ballooning, or steatosis with ballooning but without inflammation were classified as NAFL. Biopsies with NAS = 0 were classified as NL. We excluded patients if they had a histopathologic reading of NAS = 0 and fibrosis >0, NAS >0 but steatosis = 0, and samples with fibrosis = 4. Following these criteria, 192 patients were selected for further analysis (7 NL, 109 NAFL, 76 NASH). Clinical, biochemical, and histologic characteristics obtained from the validation cohort are summarized in

Table 1. The protocol was performed in accordance with the Helsinki Declaration and was approved by the institutional review board at each of the participating hospitals. All patients gave written informed consent.

BLOOD SAMPLE DETAILS

Blood specimens were collected under fasting conditions, and serum was prepared by incubating patient venous blood in serum separator tubes before centrifugation (2,500g, 15 minutes). Supernatants were aliquoted into cryovials and stored at −80°C until metabolomics analysis. Blood collection was carried out on the day of liver biopsy; height and weight were recorded at this time.

CHEMICALS

LC/MS-grade solvents were purchased from Sigma Aldrich (St. Louis, MO) and Fisher Scientific (Pittsburgh, PA). Reference metabolite standard compounds were obtained from Sigma Aldrich, Avanti Polar Lipids (Alabaster, AL), and Larodan Fine Chemicals (Malmö, Sweden).

LIPIDOMIC PROFILING

Lipidomic profiles of the serum samples were obtained as described.[19] Serum extracts were mixed with sodium chloride (50 mM) and chloroform/methanol (2:1) in 1.5 mL microtubes on ice. The extraction solvent was spiked with a TG compound, TG(13:0/13:0/13:0), not detected in unspiked human serum extracts. TGs are glycerides in which the glycerol is esterified with three fatty acid groups. TG(13:0/13:0/13:0), in particular, consists of one chain of tridecanoic acid at the C-1 position, one chain of tridecanoic acid at the C-2 position and one chain of tridecanoic acid acid at the C-3 position of the glycerol group. After brief vortex mixing, the samples were incubated for 1 hour at −20 °C. Once centrifuged at 16,000g for 15 minutes, the organic phase was collected and the solvent removed. The dried extracts were then reconstituted in acetonitrile/isopropanol (1:1), centrifuged (16,000g for 5 minutes), and transferred to vials for ultra-performance LC coupled to MS time of flight (UPLC-MS-TOF) analysis.

Chromatography was performed on a 2.1-mm internal diameter × 100 mm ACQUITY 1.7 μm C18 BEH column (Waters Corp., Milford, MA) using an ACQUITY UPLC system (Waters Corp.). The column was maintained at 60°C and eluted with a 10-minute linear gradient. The mobile phase, at a flow rate of 400 μL/minute, initially consisted of 60% solvent A (water with acetonitrile [2:3] + 10 mM ammonium formate) and 40% solvent B (acetonitrile with isopropanol [1:9] + 10 mM ammonium formate), increasing to 100% B over 10 minutes. After 5 minutes, the mobile phase was reset to the initial composition in readiness for the subsequent injection, which followed a 45-second system recycle time. The sample volume injected onto the column was 3 μL. The eluent was introduced into the mass spectrometer (SYNAPT G2 system or Xevo G2 QTof (Waters Corp.) by electrospray ionization, with capillary and cone voltages set in the positive ion mode to 3,200 and 30 V, respectively. The nebulization gas was set to 1,000 L/hour and 500°C. The cone gas was fixed at 30 L/hour, and the source temperature was maintained at 120°C. Centroid data were acquired over the mass range of 50-1,200 Da, using an accumulation time of 0.2 seconds per spectrum. All spectra were mass corrected in real time by reference to leucine enkephalin, infused at 10 μL/minute through an independent reference electrospray, sampled every 10 seconds. An appropriate test mixture of standard compounds were analyzed before and after the entire set of randomized sample injections in order to examine the retention time stability, mass accuracy, and sensitivity of the system throughout the course of the run, which lasted a maximum of 48 hours per batch of samples injected. In addition, two different types of quality control (QC) samples were used to assess the data quality. The QC samples were reference serum samples that were evenly distributed over the batches and extracted and analyzed at the same time as the clinical samples. The first QC (PromoCell Inc., Germany) was used for both internal standard intrabatch and interbatch correction and was considered the QC calibration sample, whereas the second QC (Sigma Aldrich) was used to evaluate the overall reproducibility (QC validation sample). Data were processed using the TargetLynx application manager for MassLynx 4.1 software (Waters Corp.). A set of predefined retention time mass-to-charge ratio pairs, corresponding to metabolites included in the analysis, were fed into the program. Associated extracted ion chromatograms (mass tolerance window, 0.05 Da) were then peak detected and noise reduced in both the LC and MS domains such that only true metabolite-related features were processed by the software. Normalization factors were

calculated for each metabolite as described by Martínez-Arranz et al.[24]

STATISTICAL ANALYSIS

Logistic regression analysis[25] was applied to achieve a predictive signature capable of first discriminating between NAFLD and NL and second between NASH and NAFL, introducing BMI as a continuous variable. A forward stepwise method was used as a variable selection criterion where the process started with no variables in the model and variables were added one at a time as long as these inclusions were worthy. Once a variable was added, the model was evaluated to ensure its discriminatory capability. The process was repeated until no more variables improved the model. Variables with missing values were not included in the analysis as they would provide useless information in this classification method. Receiver operating characteristic (ROC) curve analysis was used to assess the discriminatory power. Overall diagnostic accuracy for a given two-class comparison was done by the area under the ROC curve (AUROC) with its associated standard error. Sensitivity, specificity, and positive and negative predictive values (PPV and NPV, respectively) were estimated. All calculations were performed using R v.3.2[26] with caret, caTools, and ROC R (ROCR) packages to produce ROC curves and AUROC estimates and the MASS package to generate the logistic regression analysis model.

Results

IDENTIFICATION OF BIOMARKERS AND TEST DEVELOPMENT

We have previously demonstrated that serum lipidomic profiling could be used to differentiate between individuals with NL and patients with NAFLD.[19] Additionally, we also demonstrated that serum lipidomics could differentiate between patients with NASH and NAFL and that this differentiation improved if subjects were categorized by their BMI.[19] This BMI-dependent serum lipidomic test measures the probability of a certain sample to belong to one of these three groups: NL, NAFL, and NASH. In that previous work, a total of 467 biopsy-proven samples, classified as NL, NAFL, or NASH, were analyzed using three different metabolic UPLC-MS platforms that measured 540 different lipids and amino acids.[19] Taking

this work as a reference point, the next steps were conducted to reduce the complexity of the test by means of reducing the number of metabolites that needed to be measured as well as the required analytical platforms. Hepatic TG accumulation is closely related to NAFLD,[27] and there is a close correlation between the content of specific molecular species of TG in the liver and their relative content in serum.[20] Accordingly, we hypothesized that the transition from NL to NAFLD and the progression from NAFL to NASH may be described by changes in specific molecular species of TG in the liver and that these changes would be reflected in serum. As serum TGs were one of the studied lipid families that showed larger differences among the groups studied in our previous work,[19] here we focused on a group of 69 serum TGs (Supporting Table S1) that could be measured using a single UPLC-MS platform.

Different mathematical models (including random forest, support vector machine, and logistic linear regression analysis) were applied to the data to distinguish between the different stages of the disease (NL, NAFL, and NASH). An ideal biomarker of NAFLD is one that would distinguish significantly in a dose-dependent manner among the three diagnostic categories of NL, NAFL, and NASH. Out of the 69 selected TGs that optimally fulfilled this criterion, two panels totaling 28 TGs could be identified: one, integrated by 11 TGs, that significantly differentiated between NL and NAFLD but could not discriminate NASH from NAFL and a second set of 20 TGs that significantly distinguished between NASH and NAFL but could not separate NL from NASH (Fig. 2). Logistic regression analysis was applied to this set of TGs to distinguish between NL and NAFLD and between NASH and NAFL. The objective of logistic regression analysis is to find the best fitting model to explain the connection between the dichotomous variable and a set of independent conditions. In our case, those independent conditions were the lipidomic features of the panel of selected TGs and the BMI, which was introduced in the algorithms as a continuous variable.

DIAGNOSTIC PERFORMANCE IN THE DISCOVERY COHORT

To evaluate the diagnostic performance of both tests, we assessed the AUROC curves (Fig. 3). When used to distinguish between NAFLD and NL, the AUROC was 0.90 ± 0.02 with sensitivity, specificity, PPV, and NPV of 0.98, 0.78, 0.89, and 0.88,
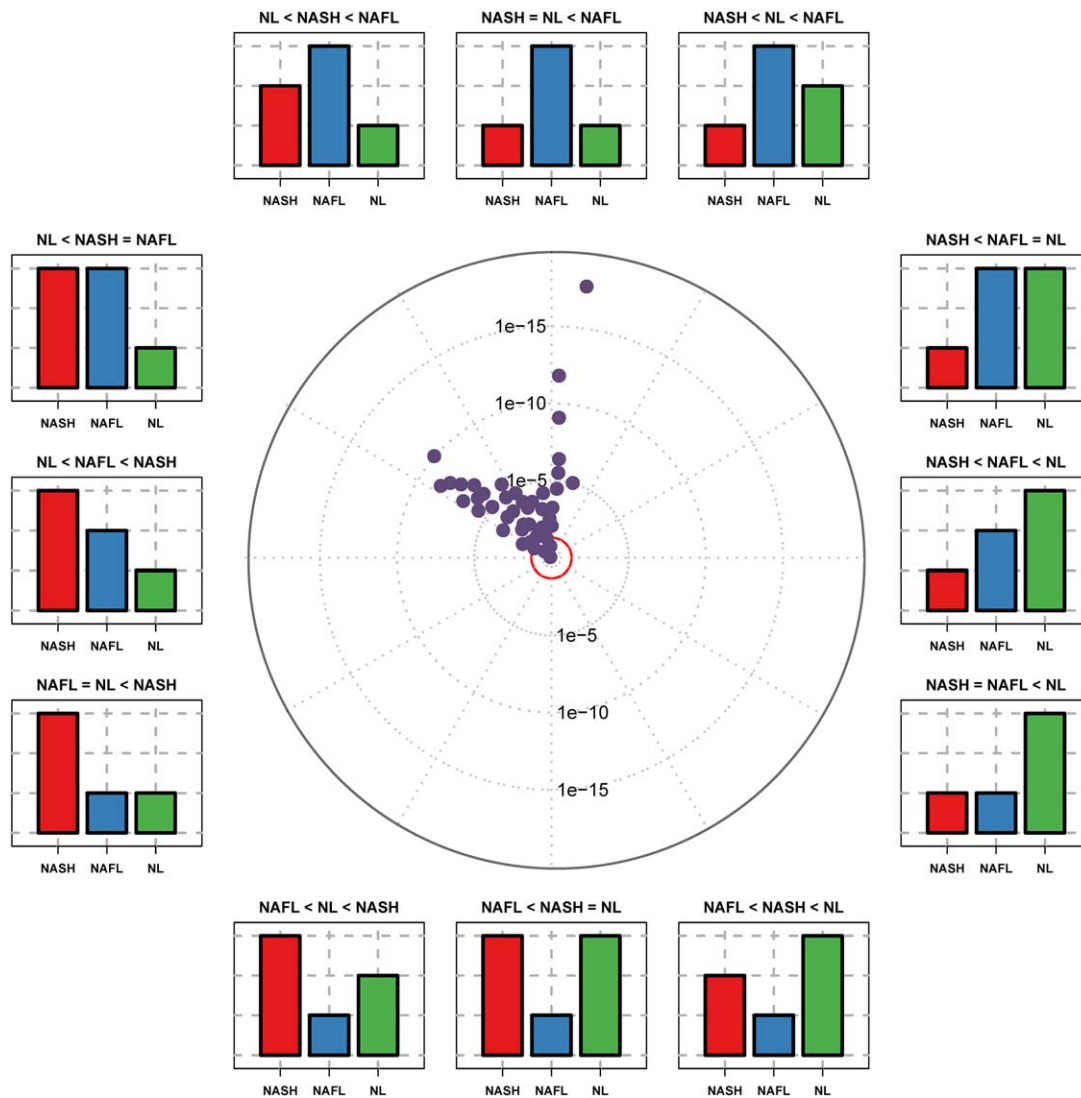
FIG. 2. Triglyceride distribution in the discrimination among patients with NL, NAFL and NASH: ANOVA plot representation. The ANOVA plot is a scatter plot of the negative log10-transformed P values from the ANOVA test. This ANOVA plot is a variation of the volcano plot and represents the effect of the TGs studied in three categories (NL, NAFL, and NASH). The distance of a TG from the center indicates its significance in the test. The position in the circle indicates the effect of each TG in the three diagnostic categories represented by the bar plots. TGs outside the inner red circle are $P < 0.05$. Abbreviation: ANOVA, analysis of variance.

respectively (Fig. 3A,C). The AUROC value for the discrimination between NASH and NAFL was 0.95 $\pm$ 0.01, with sensitivity, specificity, PPV, and NPV of 0.83, 0.94, 0.89 and 0.90, respectively (Fig. 3B,D).

DIAGNOSTIC PERFORMANCE IN THE VALIDATION COHORT

For validation of these TG biomarkers of NAFLD, we used serum samples from a new cohort of 192 adult patients who underwent liver biopsy at five

participating hospitals and who were assigned to three histologic groups using the histologic criteria described in Patients and Methods. Both the discovery and validation cohorts were similar with respect to total fasting cholesterol, TG, and glucose (Table 1). After measuring the content of the TG biomarkers of NAFLD in the serum samples of the validation cohort, we analyzed their diagnostic performance using the logistic regression algorithms described above. The relationship between the NAS and the diagnostic of NASH using the biopsy histologic data or the serum
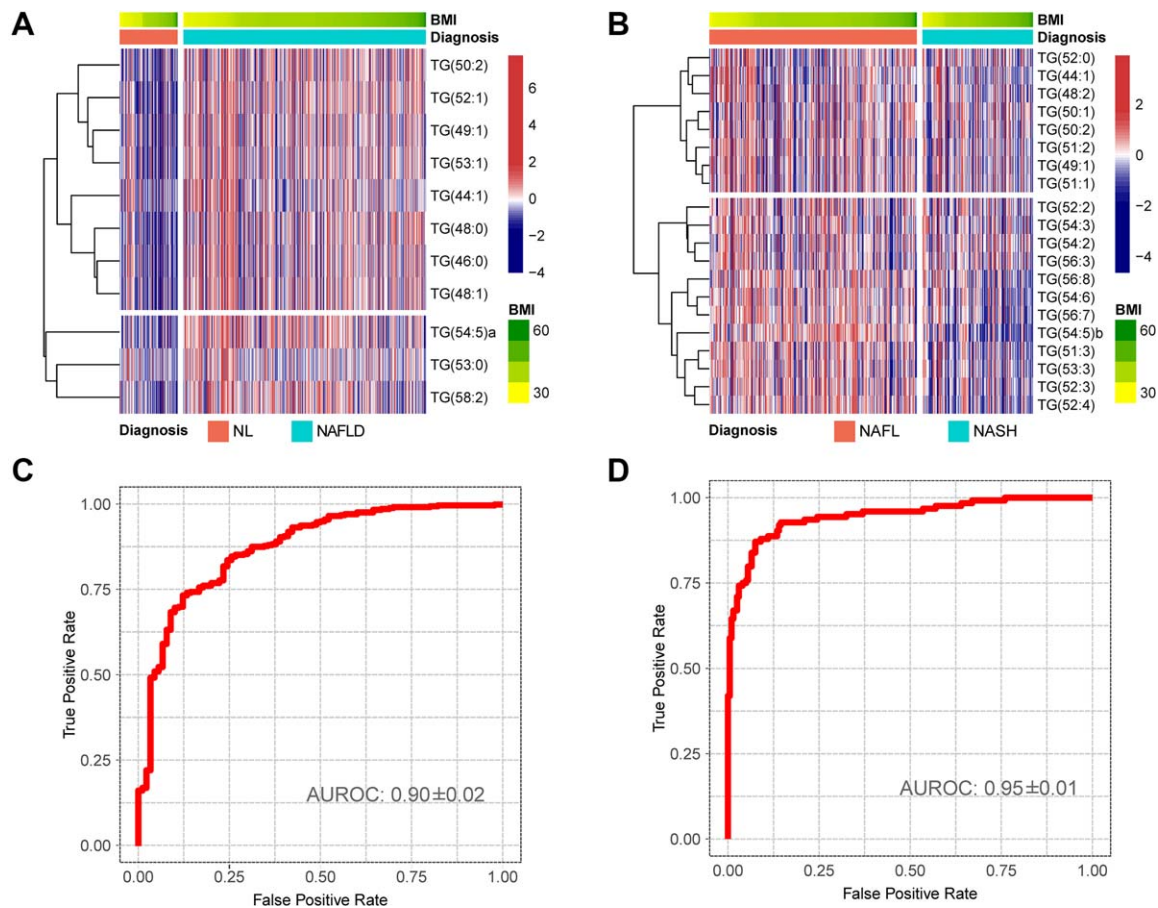
FIG. 3. NAFLD serum metabolic profile for the discrimination between NL and NAFLD and between NAFL and NASH in the discovery cohort. Obesity-dependent metabolic profile of the 28 TGs that participate in the subsequent mathematical models for the discrimination among patients with NL, NAFL and NASH. (A,B) Heatmap representation of the serum metabolic profile obtained from 467 subjects with NL (90), NAFL (246), and NASH (131) included in the discovery cohort. Each data point corresponds to the relative ion abundance of a given TG (vertical axis) in an individual patient's serum with respect to that found in the rest of the study population, 0 being the mean value. For each heatmap, log-transformed ion abundance ratios for (A) NAFLD versus NL and (B) NASH versus NAFL are depicted, as represented by the scales, where blue represents a decrease in TG abundance in NAFLD versus NL and NASH versus NAFL and red represents an increase in TG abundance. (C,D) AUROC curves calculated (C) NAFLD versus NL and (D) NASH versus NAFL.

TG biomarkers showed a similar profile, although 15% of cases with NAS = 0-2 were categorized as NASH and a higher proportion of the NAS = 4 biopsies were classified as NASH using the lipidomic test (Fig. 4). The relationship of the NAS and the histologic diagnostic of NASH was nearly identical to that published.[21,22] Cases with NAS = 4 were divided by pathologists between NAFL and NASH, as reported.[21,22] Taking this into account, we defined NASH borderline cases as those with a maximum score for one of the features of the NAS and a score of at least 1 for one of the other two features. Thus, for instance, biopsies with NAS = 4, steatosis = 3, and

inflammation = 1 or ballooning = 1 were considered borderline. Similarly, cases with NAS = 5, steatosis = 2, and inflammation = 3 and no ballooning were also considered borderline. Following this criteria, a total of 38 patients were considered as NASH borderline. The AUROC value for the diagnosis of NAFLD (discrimination between NAFLD and NL) was 0.88 ± 0.05 with accuracy, sensitivity, specificity, PPV, and NPV of 0.92, 0.94, 0.57, 0.98, and 0.25, respectively (Fig. 5A,C). Of the 185 patients with NAFLD analyzed, only 12 were incorrectly classified as NL, the majority of them (8 cases) were patients with NAFL with minimal steatosis (score 1) and NAS = 1-2. The AUROC
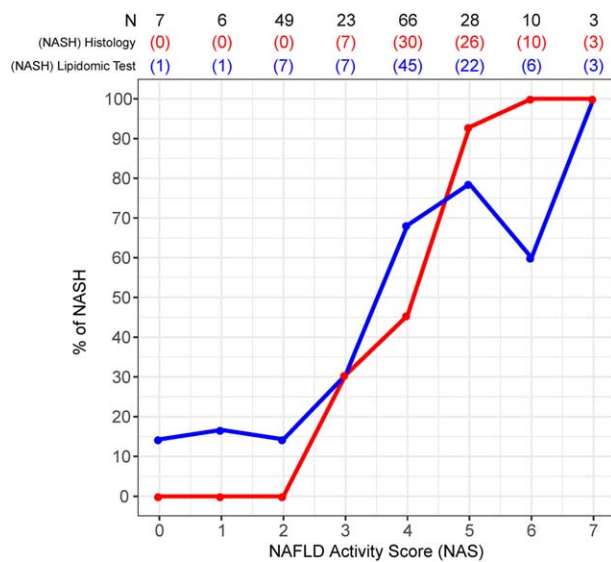
FIG. 4. Relationship between the NAS and the diagnosis of NASH using different approaches. For each NAS, the fraction of cases diagnosed as NASH is represented. The total number of observations for each score (N) and the number of patients diagnosed as NASH are shown at the top of the graph. Blue, histologic diagnosis as described in Patients and Methods; red, diagnosis of NASH using the TG biomarkers.

value for the diagnosis of NASH (discrimination between NASH and NAFL) was 0.79 ± 0.04, with accuracy, sensitivity, specificity, PPV, and NPV of 0.77, 0.70, 0.81, 0.81, and 0.69, respectively (Fig. 5B,D). NASH borderline biopsies were not included in the analysis as those were divided between both diagnostic categories. Finally, if borderline biopsies were included, the AUROC for the discrimination between NAFL and NASH decreased to 0.68 ± 0.04 (not shown).

ERROR RATES IN THE DETERMINATION OF NASH

Next, we evaluated the impact of the stage of each of the histologic features of the NAS (steatosis, inflammation, and ballooning) on the accuracy of the TG biomarkers to diagnose NASH. In terms of steatosis, the diagnostic error rate in patients with NAFL and steatosis score 1 and 2 was around 15% and for steatosis score 3 it was 60%, the majority of them (38 of 40) corresponding to borderline cases (Supporting Fig. S1A). The diagnostic error rate in patients with NASH inversely correlated with the steatosis score (5 of 10 patients [50%] with score 1, 10 of 33 patients

[30%] with score 2, and 8 of 30 patients [27%] with score 3) (Supporting Fig. S1A). With respect to inflammation, the diagnostic error rate in patients with NAFL was around 30% for scores 0 and 1, while 1 of 2 patients with score 2 was misclassified as NASH. The degree of inflammation in patients with NASH inversely correlated with the diagnostic error rate (21 of 63 patients with score 1, 2 of 9 patients with score 2, and 0 of 1 patient with score 3 were incorrectly classified as NAFL) (Supporting Fig. S1B). In terms of ballooning, the diagnostic error rate in patients with NAFL was around 35% in cases with ballooning scores 0 and 1; the degree of ballooning in patients with NASH inversely correlated with the diagnostic error rate (21 of 58 patients with score 1 and 2 of 15 patients with score 2 were incorrectly classified as NAFL) (Supporting Fig. S1C). To identify how the degree of fibrosis would affect the performance of the diagnosis of NASH, we estimated the diagnostic error rate for the different values of fibrosis. The average diagnostic error rate in patients with NAFL was 30% when the score of fibrosis was 0-1 and 3 and was 67% when the score was 2 (2 of 3 patients were incorrectly classified) (Supporting Fig. S1D). Around 46% of patients with NASH were incorrectly classified as NAFL when the score of fibrosis was 0; this decreased to around 30% for scores 1-3 (Supporting Fig. S1D). Finally, the effect of BMI and glucose concentration was also evaluated due to the well-known relationship among these two variables and NASH.[28,29] The relationship between the BMI and the diagnostic error rate of NASH was constant and around 30% in the range of 25-60 kg/m^2 (Supporting Fig. S1E). The average diagnostic error rate for glucose was around 30% for glucose levels <136 mg/dL, increasing to 60%-70% for larger values (Supporting Fig. S1F). Accordingly, when the analysis was performed excluding patients with glucose levels >136 mg/dL, the AUROC for the discrimination between NASH and NAFL increased to 0.81 ± 0.04, with sensitivity, specificity, PPV, and NPV of 0.73, 0.80, 0.78, and 0.75, respectively (Supporting Fig. S2).

Discussion

Liver biopsy remains the diagnostic test of choice for NASH, but it is an invasive procedure and is limited by sampling error, high cost, procedure-related complications, and observer variability, even when performed by expert pathologists.[6,7] In this study, from a
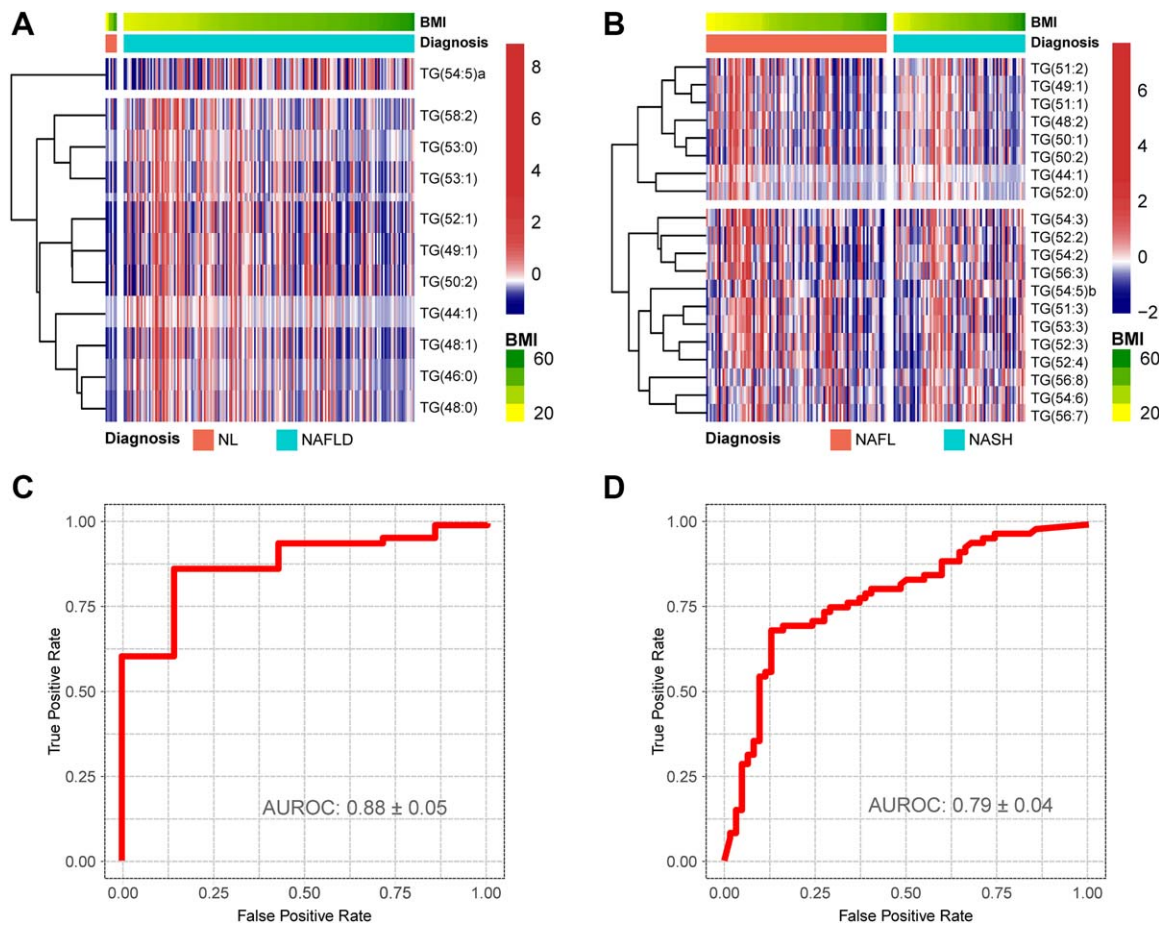
FIG. 5. NAFLD serum metabolic profile for the discrimination between NL and NAFLD and between NAFL and NASH in the validation cohort. Obesity-dependent metabolic profile of the 28 TGs that participate in the discrimination among NL, NAFL and NASH for the patients included in the validation cohort. (A,B) Heatmap representation of the serum metabolic profile obtained from 192 patients with NL (7), NAFL (109), and NASH (76) included in the validation cohort. Each data point corresponds to the relative ion abundance of a given TG (vertical axis) in an individual patient's serum with respect to that found in the rest of the study population, 0 being the mean value. For each heatmap, log-transformed ion abundance ratios (A) NAFLD versus NL and (B) NASH versus NAFL are depicted, as represented by the scales, where blue represents a decrease in TG abundance in NAFLD versus NL and NASH versus NAFL and red represents an increase in TG abundance. (C,D) AUROC curves calculated (C) NAFLD versus NL and (D) NASH versus NAFL.

large cohort of patients with NAFLD with different disease activity, we performed a detailed metabolomics analysis from serum samples that showed good accuracy for distinguishing between NL and NAFLD and NASH from NAFL. This lipidomic test provides promising easy-to-perform serum biomarkers to differentiate disease severity and potentially to monitor disease progression or improvement.

Many advances have appeared in relation to the noninvasive diagnosis of NAFLD over the last decade. MRI-PDFF and MRE have been demonstrated to be accurate tools for the detection of steatosis[8-10] and fibrosis.[11-13] MRE performance to distinguish

various fibrosis stages has been shown to have an AUROC between 0.82 and 0.89[31] but has been limited to distinguish NAFL from NASH (AUROC of 0.70).[30] In addition, MRI-PDFF and MRE are expensive and require special software. Furthermore, these tests showed limited promise for discriminating NAFL from NASH.

Many serum biomarkers have been reported to differentiate between NAFLD and NL or to distinguish NASH from NAFL, but none of them have been established to be effective in a large cohort of patients with biopsy-proven NAFLD[15-19] or have been validated using an independent blind cohort. Cytokine

keratin 18 initially gained interest and was suggested in practice guidelines as a tool to distinguish NAFLD from NASH. However, recent studies have commented on its lack of reproducibility and reliability.[31,32] The NAFLD-fibrosis score is calculated based on age, BMI, hyperglycemia or diabetes, aspartate aminotransferase (AST), alanine aminotransferase (ALT), platelets, and albumin (www.nafldscore.com), while the FIB-4 index[33] is based on age, AST, ALT, and platelets. These biomarkers along with the enhanced liver fibrosis score[34] and others are increasingly used but usually targeted to test advanced fibrosis. In a recent study of 600 patients from a European consortium in which biomarkers of NAFLD were studied, the authors tested only a few biomarkers, mainly to diagnose advanced fibrosis. These markers included SteatoTest, ActiTest, FibroTest, and FIB-4. They found that the AUROC of these biomarkers ranged from 0.80 to 0.88.[35] Some of these biomarkers have been suggested by the recent European Association for the Study of the Liver guidelines to assess the severity of fibrosis and rule out advanced stages.[23] However, there were no recommendations on any biomarkers to diagnose NASH; liver biopsy remained the only recommendation for this in the recent guidelines. The objective of this study was the identification of a panel of TGs that, combined with the BMI and a classification algorithm, would accurately differentiate between NL and NAFLD and distinguish NAFL from NASH, in a cohort of 467 patients with liver biopsies. We used MS lipidomics and then validated these panels of biomarkers and the classification algorithm in an independent blind cohort of patients with biopsy-proven NAFLD. Our serum biomarkers test is the first of its kind to distinguish between NL and NAFLD and NAFL from NASH with high accuracy in a large blinded validation cohort.

UPLC-MS is routinely used in analytical chemistry laboratories for toxicity studies and diagnosis of inherited metabolic disorders. The implementation and growth in clinical laboratories will depend on the emergence of new diagnosis tests as the one discussed here.[36] MS lipidomics is a high-throughput technology that can measure simultaneously hundreds of different lipid classes and molecular species in a small volume of serum.[37-40] We first demonstrated this concept by showing that a panel of 540 lipids could differentiate NAFLD from NL and NASH from NAFL.[19] We also observed that categorizing patients by their BMI improved the power of the classification algorithm.[41] Here, we have drastically reduced the number of metabolites that were necessary to diagnose NASH from a metabolomic profile of roughly 540 lipids and amino acids[19] to a panel of 28 TGs. TGs were chosen because NAFLD is characterized by an abnormal regulation of liver TG homeostasis[42,43] and the observation that changes in the molecular species of liver TGs are mirrored in serum.[44]

The performance of the lipidomic test that distinguishes between NL and NAFLD showed an AUROC of 0.90 ± 0.02 in the discovery cohort, a value that was maintained when applied to the validation cohort (AUROC of 0.88 ± 0.05) with a good sensitivity of 0.94, indicating a low number of false-negative cases.[45] The performance of the lipidomic test that distinguishes between NASH and NAFL showed an AUROC of 0.95 ± 0.01 in the discovery cohort. After validation, the test maintained a suitable capacity to distinguish NASH from NAFL (AUROC of 0.79 ± 0.04) with a good specificity of 0.81, indicating that it may perform well for NASH diagnosis because of low false-positive errors.[45] This test divided patients classified as NASH borderline between the two diagnostic categories (NAFL and NASH), which agrees with previous studies.[21,22] Analysis of the impact of the stage of each of the histologic features of the NAS system on the accuracy of the diagnostic test revealed that, in general, the error rate of NASH diagnosis ranged between 20% and 40% for the different values of the features of the NAS system and between 30% and 60% for the different score values of fibrosis. The effect of BMI on the error rate of NASH diagnosis was constant and around 30% in the range of 25-60 kg/m^2. This result is in line with the fact that the BMI was introduced in the algorithms as a continuous variable, modulating the effect of this variable in the diagnostic performance of the tests. However, the effect of glucose level was not taken into account in the algorithms, and we observed an increase in the error rate for glucose levels >136 mg/dL that corresponded with decompensated diabetes. When the analysis was performed excluding patients with glucose levels >136 mg/dL, the AUROC for the discrimination between NASH and NAFL increased to 0.81 ± 0.04, with sensitivity, specificity, PPV, and NPV of 0.73, 0.80, 0.78, and 0.75, respectively.

The differences found in both cohorts regarding sex, ALT, and AST are not considered of great importance for the diagnosis of NASH as we did not find any evidence of their influence in the study of the discovery cohort.[19] Regarding sex, although the proportion of female individuals was lower in the validation cohort,

we did not see differences in the diagnosis related to sex as can be observed in Fig. 1C where all the samples clustered together independently of sex. PCA analysis of the serum metabolomics data of both cohorts also indicates that they are comparable. With regards to age, although several studies have been published highlighting the relationship between the prevalence of NASH and age,[46,47] the average age was similar in the different stages of the disease for both cohorts.

Mechanistically, the reason by which this group of 28 specific serum TGs served to differentiate between NL and NAFLD and between NASH and NAFL is not clear at first glance. This panel included saturated (TG[46:0], TG[48:0], TG[53:0]) and unsaturated (TG[44:1], TG[48:1], TG[49:1], TG[52:1], TG [53:1], TG[50:2], TG[54:5], TG[58:2], for instance) TGs of different fatty acid lengths, which were always increased in NAFLD compared to NL but were lower in NASH than in NAFL. This may reflect an impaired synthesis and release of very low-density lipoprotein in NASH[48] and agrees with the observation that NASH does not require a particular amount of TG.[49]

Our study has both strengths and limitations. A limitation is the cross-sectional nature of the analysis in both the original and validation cohorts. Further studies are needed to assess changes in our tests longitudinally. Our test does not assess the degree of fibrosis; however, noninvasive tests for fibrosis, such as enhanced liver fibrosis, FIB-4, and MRE, are already established, while we are introducing a new test to distinguish NAFL from NASH, which is an unmet need. The test does not take into account the potential role of noncompensated diabetes, ethnicity, age, patatin-like phospholipase domain-containing 3 (PNPLA3) genotype or the possible differences in the lipidomic profile that could be found between different NAFLD phenotypes, for instance, metabolically healthy obese subjects and metabolically unhealthy obese patients. Therefore, additional efforts should be made to enrich our cohort with patients of different conditions, ethnicities, and phenotypes. On the other hand, the strength of our study includes the large sample size with a validation cohort, multicenter samples, detailed and standardized metabolomics and histologic analysis, confirmation of pathology reading by multiple pathologists (all experts in NAFLD), diagnosis with a central reader assessing discrepancy between readers, and finally the reproducibility of this accurate test in the original and validation cohorts. Addressing these and other limitations of this model will be important next

steps toward perfect discrimination between NAFL and NASH in serum. In conclusion, two lipidomic tests have been validated that provide an easy-to-use tool to differentiate disease severity and potentially to monitor disease progression.

REFERENCES

1) Page JM, Harrison SA. NASH and HCC. Clin Liver Dis 2009; 13:631-647.

2) **Starley BQ, Calcagno CJ**, Harrison SA. Nonalcoholic fatty liver disease and hepatocellular carcinoma: a weighty connection. Hepatology 2010;51:1820-1832.

3) Bellentani S, Scaglioni F, Marino M, Bedogni G. Epidemiology of non-alcoholic fatty liver disease. Dig Dis 2010;28:155-161.

4) Petta S, Muratore C, Craxì A. Non-alcoholic fatty liver disease pathogenesis: the present and the future. Dig Liver Dis 2009;41: 615-625.

5) Mishra A, Younossi ZM. Epidemiology and natural history of non-alcoholic fatty liver disease. J Clin Exp Hepatol 2012;2:135-144.

6) **El-Badry AM**, **Breitenstein S**, Jochum W, Washington K, Paradis V, Rubbia-Brandt L, et al. Assessment of hepatic steatosis by expert pathologists: the end of a gold standard. Ann Surg 2009;250:691-697.

7) **Ratziu V**, **Charlotte F**, Heurtier A, Gombert S, Giral P, Bruckert E, et al.; LIDO Study Group. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology 2005;128:1898-1906.

8) Idilman IS, Aniktar H, Idilman R, Kabacam G, Savas B, Elhan A, et al. Hepatic steatosis: quantification by proton density fat fraction with MR imaging versus liver biopsy. Radiology 2013; 267:767-775.

9) Nasr P, Forsgren MF, Ignatova S, Dahlström N, Cedersund G, Leinhard OD, et al. Using a 3% proton density fat fraction as a cut-off value increases sensitivity of detection of hepatic steatosis, based on results from histopathology analysis. Gastroenterology 2017;153:53-55.e7.

10) Noureddin M, Lam J, Peterson MR, Middleton M, Hamilton G, Le TA, et al. Utility of magnetic resonance imaging versus histology for quantifying changes in liver fat in nonalcoholic fatty liver disease trials. Hepatology 2013;58:1930-1940.

11) Dulai PS, Sirlin CB, Loomba R. MRI and MRE for noninvasive quantitative assessment of hepatic steatosis and fibrosis in NAFLD and NASH: clinical trials to clinical practice. J Hepatol 2016;65:1006-1016.

12) Venkatesh SK, Yin M, Ehman RL. Magnetic resonance elastography of liver: technique, analysis, and clinical applications. J Magn Reson Imaging 2013;37:544-555.

13) **Cui J**, **Heba E**, Hernandez C, Haufe W, Hooker J, Andre MP, et al. Magnetic resonance elastography is superior to acoustic

radiation force impulse for the diagnosis of fibrosis in patients with biopsy-proven nonalcoholic fatty liver disease: a prospective study. Hepatology 2016;63:453-461.

14) Hannah WN, Torres DM, Harrison SA. Nonalcoholic steatohepatitis and endpoints in clinical trials. Gastroenterol Hepatol (N Y) 2016;12:756-763.

15) Pirola CJ, Fernández Gianotti T, Castaño GO, Mallardi P, San Martino J, Mora Gonzalez Lopez Ledesma M, et al. Circulating microRNA signature in non-alcoholic fatty liver disease: from serum non-coding RNAs to liver histology and disease pathogenesis. Gut 2015;64:800-812.

16) Afonso MB, Rodrigues PM, Simao AL, Castro RE. Circulating microRNAs as potential biomarkers in non-alcoholic fatty liver disease and hepatocellular carcinoma. J Clin Med 2016;5.pii.E30.

17) Shen J, Chan HL, Wong GL, Choi PC, Chan AW, Chan HY, et al. Non-invasive diagnosis of non-alcoholic steatohepatitis by combined serum biomarkers. J Hepatol 2012;56:1363-1370.

18) Yang M, Xu D, Liu Y, Guo X, Li W, Guo C, et al. Combined serum biomarkers in non-invasive diagnosis of non-alcoholic steatohepatitis. PLoS One 2015;10:e0131664.

19) Barr J, Caballería J, Martínez-Arranz I, Domínguez-Díez A, Alonso C, Muntané J, et al. Obesity-dependent metabolic signatures associated with nonalcoholic fatty liver disease progression. J Proteome Res 2012;11:2521-2532.

20) Alonso C, Fernández-Ramos D, Varela-Rey M, Martínez-Arranz I, Navasa N, Van Liempd SM, et al. Metabolomic identification of subtypes of nonalcoholic steatohepatitis. Gastroenterology 2017;152:1449-1461.e7.

21) Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al.; Nonalcoholic Steatohepatitis Clinical Research Network. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005; 41:1313-1321.

22) Brunt EM, Kleiner DE, Wilson LA, Belt P, Neuschwander-Tetri BA, et al.; NASH Clinical Research Network (CRN). Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct clinicopathologic meanings. Hepatology 2011;53:810-820.

23) European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. J Hepatol 2016;64:1388-1402.

24) Martínez-Arranz I, Mayo R, Pérez-Cormenzana M, Mincholé I, Salazar L, Alonso C, et al. Enhancing metabolomics research through data mining. J Proteomics 2015;127:275-288.

25) Dobson AJ, Barnett A. An Introduction to Generalized Linear Models. 3rd ed. Boca Raton, FL: CRC Press; 2008.

26) R Development CoreTeam. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Available at: https://www.r-project.org; 2017.

27) Ogawa Y, Imajo K, Yoneda M, Nakajima A. Pathophysiology of NAsh/NAFLD associated with high levels of serum triglycerides. [in Japanese] Nihon Rinsho 2013;71:1623-1629.

28) Bhatt HB, Smith RJ. Fatty liver disease in diabetes mellitus. Hepatobiliary Surg Nutr 2015;4:101-108.

29) Williams KH, Shackel NA, Gorrell MD, Mclennan SV, Twigg SM. Diabetes and nonalcoholic fatty liver disease: Endocr Rev 2013;34:84-129.

30) Park CC, Nguyen P, Hernandez C, Bettencourt R, Ramirez K, Fortney L, et al. Magnetic resonance elastography vs transient elastography in detection of fibrosis and noninvasive measurement of steatosis in patients with biopsy-proven nonalcoholic fatty liver disease. Gastroenterology 2017;152:598-607.e2.

31) Chalasani N, Younossi Z, Lavine JE, Diehl AM, Brunt EM, Cusi K, et al.; American Association for the Study of Liver Diseases; American College of Gastroenterology; American Gastroenterological Association. The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. Hepatology 2012;55: 2005-2023.

32) Pimentel CF, Jiang ZG, Otsubo T, Feldbrügge L, Challies TL, Nasser I, et al. Poor inter-test reliability between CK18 kits as a biomarker of NASH. Dig Dis Sci 2016;61:905-912.

33) Vallet-Pichard A, Mallet V, Nalpas B, Verkarre V, Nalpas A, Dhalluin-Venier V, et al. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. Comparison with liver biopsy and fibrotest. Hepatology 2007;46:32-36.

34) Lichtinghagen R, Pietsch D, Bantel H, Manns MP, Brand K, Bahr MJ. The enhanced liver fibrosis (ELF) score: normal values, influence factors and proposed cut-off values. J Hepatol 2013;59:236-242.

35) Munteanu M, Tiniakos D, Anstee Q, Charlotte F, Marchesini G, Bugianesi E, et al.; FLIP Consortium and the FibroFrance Group. Diagnostic performance of FibroTest, SteatoTest and ActiTest in patients with NAFLD using the SAF score as histological reference. Aliment Pharmacol Ther 2016;44:877-889.

36) Trivedi DK, Hollywood KA, Goodacre R. Metabolomics for the masses: the future of metabolomics in a personalized world. New Horiz Transl Med 2017;3:294-305.

37) Pietiläinen KH, Sysi-Aho M, Rissanen A, Seppänen-Laakso T, Yki-Järvinen H, Kaprio J, et al. Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects--a monozygotic twin study. PLoS One 2007;2:e218.

38) Oberbach A, Blüher M, Wirth H, Till H, Kovacs P, Kullnick Y, et al. Combined proteomic and metabolomic profiling of serum reveals association of the complement system with obesity and identifies novel markers of body fat mass changes. J Proteome Res 2011;10:4769-4788.

39) Kim JY, Park JY, Kim OY, Ham BM, Kim H-J, Kwon DY, et al. Metabolic profiling of plasma in overweight/obese and lean men using ultra performance liquid chromatography and Q-TOF mass spectrometry (UPLC-Q-TOF MS). J Proteome Res 2010;9:4368-4375.

40) Ritchie SA, Heath D, Yamazaki Y, Grimmalt B, Kavianpour A, Krenitsky K, et al. Reduction of novel circulating long-chain fatty acids in colorectal cancer patients is independent of tumor burden and correlates with age. BMC Gastroenterol 2010;10:140.

41) Alonso C, Mato JM. Nonalcoholic fatty liver disease. In: Wilson ID, Theodoridis G, Nicholls, A, eds. Global Metabolic Profiling: Clinical Applications. London, United Kingdom: Future Science Ltd; 2014:110-122.

42) Kawano Y, Cohen DE. Mechanisms of hepatic triglyceride accumulation in non-alcoholic fatty liver disease. J Gastroenterol 2013;48:434-441.

43) Cortez-Pinto H, Chatham J, Chacko VP, Arnold C, Rashid A, Diehl AM. Alterations in liver ATP homeostasis in human nonalcoholic steatohepatitis: a pilot study. JAMA 1999;282:1659-1664.

44) Wu X, Tong Y, Shankar K, Baumgardner JN, Kang J, Badeaux J, et al. Lipid fatty acid profile analyses in liver and serum in rats with nonalcoholic steatohepatitis using improved gas chromatography-mass spectrometry methodology. J Agric Food Chem 2011;59:747-754.

45) Maxim LD, Niebo R, Utell MJ. Screening tests: A review with examples. Inhal Toxicol 2014;26:811-828.

46) Noureddin M, Yates KP, Vaughn IA, Neuschwander-Tetri BA, Sanyal AJ, Mccullough A, et al.; NASH CRN. Clinical and histological determinants of nonalcoholic steatohepatitis and advanced fibrosis in elderly patients. Hepatology 2013;58:1644-1654.

47) Gan L, Chitturi S, Farrell GC. Mechanisms and implications of age-related changes in the liver: nonalcoholic fatty liver disease in the elderly. Curr Gerontol Geriatr Res 2011;2011:831536.

48) Fujita K, Nozaki Y, Wada K, Yoneda M, Fujimoto Y, Fujitake M, et al. Dysfunctional very-low-density lipoprotein synthesis and release is a key factor in nonalcoholic steatohepatitis pathogenesis. Hepatology 2009;50:772-780.

49) **Yeh MM**, **Brunt EM**. Pathology of nonalcoholic fatty liver disease. Am J Clin Pathol 2007;128:837-847.

Author names in bold designate shared co-first authorship.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep4.1188/full.