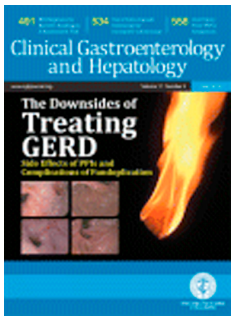
Accepted Manuscript



Non-invasive Detection of Non-alcoholic Steatohepatitis Using Clinical Markers and Circulating Levels of Lipids and Metabolites

You Zhou, Matej Orešič, Marja Leivonen, Peddinti Gopalacharyulu, Jenni Hyysalo, Johanna Arola, An Verrijken, Sven Francque, Luc Van Gaal, Tuulia Hyötyläinen, Hannele Yki-Järvinen

PII: S1542-3565(16)30315-9
DOI: 10.1016/j.cgh.2016.05.046
Reference: YJCGH 54802

To appear in: *Clinical Gastroenterology and Hepatology*
Accepted Date: 16 May 2016

Please cite this article as: Zhou Y, Orešič M, Leivonen M, Gopalacharyulu P, Hyysalo J, Arola J, Verrijken A, Francque S, Van Gaal L, Hyötyläinen T, Yki-Järvinen H, Non-invasive Detection of Non-alcoholic Steatohepatitis Using Clinical Markers and Circulating Levels of Lipids and Metabolites, *Clinical Gastroenterology and Hepatology* (2016), doi: 10.1016/j.cgh.2016.05.046.

Title: Non-invasive Detection of Non-alcoholic Steatohepatitis Using Clinical Markers and Circulating Levels of Lipids and Metabolites

You Zhou[1,2], Matej Orešič[3], Marja Leivonen[4], Peddinti Gopalacharyulu[5], Jenni Hyysalo[1,6], Johanna Arola[7], An Verrijken[8], Sven Francque[9], Luc Van Gaal[8], Tuulia Hyötyläinen[3], Hannele Yki-Järvinen[1,6]

[1]*Minerva Foundation Institute for Medical Research, Helsinki, Finland*

[2]*Systems Immunity University Research Institute and Division of Infection and Immunity, School of Medicine, Cardiff University, Cardiff, United Kingdom*

[3]*Steno Diabetes Center, Gentofte, Denmark*

[4]*Department of Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland*

[5]*Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland*

[6]*Department of Medicine, University of Helsinki, Helsinki, Finland*

[7]*Department of Pathology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland*

[8]*Department of Endocrinology, Diabetology and Metabolism, Antwerp University Hospital, University of Antwerp, Antwerp, Belgium*

[9]*Department of Gastroenterology and Hepatology, Antwerp University Hospital, University of Antwerp, Antwerp, Belgium*

Conflict of interest: None **DISCLOSURES:** None.

Financial support This study was supported by research grants from the Academy of Finland (HY), EU/EFPIA Innovative Medicines Initiative Joint Undertaking (EMIF grant no 115372, HY, MO, TH), EU H2020 EPoS 634413 (HY, MO, TH), the Sigrid Juselius (HY), EVO (HY) and the Novo Nordisk Foundation (HY).

CORRESPONDING AUTHOR:

Hannele Yki-Järvinen, MD, Haartmaninkatu 8, 00290 Helsinki, FINLAND

E-mail: hannele.yki-jarvinen@helsinki.fi

Phone: +358 50 427 1664

ELECTRONIC WORD COUNT: 3989 words.

FIGURES: 3

TABLES: 3

AUTHOR CONTRIBUTIONS: YZ – study concept and design; acquisition of data; analysis and interpretation of data; drafting of the manuscript; statistical analysis. YZ – study concept and design; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content; statistical analysis. MO, ML, PG, JH, JA, AV, LVG, TH – acquisition of data. HY – study concept and design; analysis and interpretation of data; drafting of the manuscript; obtained funding; study supervision.

Background & Aims: Use of targeted mass-spectrometry (MS)-based methods is increasing in clinical chemistry laboratories. We investigate whether MS-based profiling of plasma improves non-invasive risk estimates of non-alcoholic steatohepatitis (NASH) compared to routinely available clinical parameters and *PNPLA3* genotype at rs738409.

Methods: We used MS-based analytical platforms to measure levels of lipids and metabolites in blood samples from 318 subjects who underwent a liver biopsy because of suspected NASH. The subjects were randomly divided into estimation (n=223) and validation (n=95) groups to build and validate the model. Gibbs sampling and stepwise logistic regression, which fulfilled the Bayesian information criterion, were used for variable selection and modeling.

Results: Features of the metabolic syndrome and the variant in *PNPLA3* encoding I148M were significantly more common among subjects with than without NASH. We developed a model to identify subjects with NASH based on clinical data and *PNPLA3* genotype (NASH Clin Score), which included aspartate aminotransferase (AST), fasting insulin, and *PNPLA3* genotype. This model identified subjects with NASH with an area under the receiver operating characteristic (AUROC) of 0.792 (95% CI, 0.726–0.859). We then used backward stepwise logistic regression analyses of variables from the NASH Clin Score and MS-based factors associated with NASH to develop the NASH ClinLipMet Score. This included glutamate, isoleucine, glycine, lysophosphatidylcholine 16:0, phosphoethanolamine 40:6, AST, and fasting insulin, along with *PNPLA3* genotype. It identified patients with NASH with an AUROC of 0.866 (95% CI, 0.820–0.913). The NASH ClinLipMet score identified patients with NASH with significantly higher accuracy than the NASH Clin Score or MS-based profiling alone.

Conclusion: A score based on MS (glutamate, isoleucine, glycine, lysophosphatidylcholine 16:0, phosphoethanolamine 40:6) and knowledge of AST, fasting insulin and *PNPLA3* genotype is significantly better than a score based on clinical or metabolic profiles alone in determining risk of NASH.

KEY WORDS: liver; non-alcoholic fatty liver disease; diagnosis; prediction; triglycerides

Abbreviations

ALT	alanine aminotransferase
AST	aspartate aminotransferase
AUROC	area under the receiver operating characteristic curve
BIC	bayesian index criterion
BMI	body mass index
DNL	*de novo* lipogenesis
fP	fasting plasma
fS	fasting serum
GGT	gamma glutamyl transpeptidase
GC	gas chromatography
Glu	glutamate
Gly	glycine
HbA_{1c}	glycosylated hemoglobin 1c
HDL	high-density lipoprotein
HOMA-IR	Homeostasis model assessment of insulin resistance
Ile	isoleucine
IR	insulin resistance
LC	lipid cluster
Leu	leucine
LDL	low-density lipoprotein
LysoPC	lysophosphatidylcholines
MetS	metabolic syndrome
MS	mass spectrometry
MUFA	monounsaturated fatty acid

NAFL	non-alcoholic fatty liver
NAFLD	non-alcoholic fatty liver disease
NASH	non-alcoholic steatohepatitis
NPV	negative predictive values
PC	phosphatidylcholine
PE	phosphatidylethanolamine
PG	phosphatidylglycerol
PNPLA3	patatin-like phospholipase domain-containing protein 3
PPV	positive predictive values
Ser	serine
SFA	saturated fatty acid
SM	sphingomyelin
TG	triacylglycerol
TOFMS	time-of-flight mass spectrometry
Tyr	tyrosine
UPLC	ultra-performance liquid chromatography
Val	valine

1

2

Introduction

The spectrum of non-alcoholic fatty liver disease (NAFLD) ranges from the nonalcoholic fatty liver (NAFL) to non-alcoholic steatohepatitis (NASH), which increases the risk of cirrhosis and mortality from liver disease.[1] Only a fraction of patients with NAFLD progress to NASH, of which the diagnosis requires a liver biopsy. There is thus a need to develop non-invasive tools to identify patients, who might be at risk of having NASH.

Factors such as age, gender, liver enzymes, components of the metabolic syndrome (MetS) as well as circulating markers of inflammation, fibrosis, apoptosis and extracellular matrix components have been shown to associate with NASH.[1] Genetic factors, especially the I148M variant in PNPLA3 also confers susceptibility to NASH.[1]

Ultra-performance liquid chromatography mass spectrometry (UPLC-MS) based techniques are rapidly entering clinical chemistry laboratories and replacing many conventional techniques.[2] Few data are available regarding such approaches to identify new markers for non-invasive estimation of the risk of NASH. In 24 obese patients with NASH, 11 with steatosis and 25 lean controls, who did not undergo a liver biopsy, plasma concentrations of glutamate (Glu), isoleucine (Ile), leucine (Leu), tyrosine (Tyr) and valine (Val) were significantly increased but diagnostic performance compared to routinely available markers was not examined.[3] Barr et al. characterized lipids and metabolites in serum of 467 Caucasian patients and found between 9 and 237 metabolites to be markers of NASH depending on the degree of obesity.[4]

The human liver lipidome differs markedly between subjects with NAFLD associated with insulin resistance (‘IR NAFLD’) compared to those without, and between subjects with NAFLD and the PNPLA3 I148M genotype (‘PNPLA3 NAFLD’) compared to those lacking the gene variant.[5] The liver lipidome is markedly enriched with saturated and monounsaturated triglycerides and free fatty

acids in 'IR NAFLD', and with polyunsaturated triglycerides in 'PNPLA3 NAFLD'.[5] These differences also influence the circulating lipidome in NAFLD in subjects whose liver fat content has been measured using proton magnetic resonance spectroscopy but who have not undergone a liver biopsy.[6] There are, however, no studies in patients who have undergone a liver biopsy which would have determined whether knowledge of the PNPLA3 genotype influences biomarkers of NASH as compared to non-NASH (NAFL or normal liver histology) subjects. Furthermore, no study has analysed whether MS -based markers significantly improve predictive performance of scores based on routinely available physical and biochemical parameters. In the present study, we developed scores based on i) routinely available clinical parameters and PNPLA3 genotype, ii) UPLC-MS analyses alone and iii) all available information for estimation of the risk of NASH. The diagnostic performance of the three models was then compared.

Materials and methods

Study subjects

Metabolic studies were conducted at the University of Helsinki (Finland) and Antwerp University Hospital (Belgium). A total of 318 subjects were recruited amongst those referred to the Department of Gastroenterology (Finland, n=54) because of chronically elevated serum transaminase concentrations and amongst those referred for bariatric surgery in Belgium (n=193) and Finland (n=71). Subjects were eligible if they met the following criteria: (a) age 18 to 75 years; (b) no known acute or chronic disease except for obesity or type 2 diabetes on the basis of medical history, physical examination and standard laboratory tests (blood counts, serum creatinine, thyroid-stimulating hormone, electrolyte concentrations) and electrocardiogram; (b) alcohol consumption less than 20 g per day. Hepatitis B surface antigen, transferrin saturation, and antibodies against hepatitis A and C and anti-smooth muscle-, anti-nuclear- and anti-mitochondrial antibodies were measured in all patients referred to the gastroenterologist because of chronically elevated liver function tests using routine methods of local laboratories. Patients were excluded if they used

thiazolidinediones or were pregnant. The study protocol was approved by the ethics committees of the Helsinki and the Antwerp University Hospitals. Each participant provided written informed consent.

Metabolic study

All subjects were invited to a clinical visit one week prior to surgery for metabolic characterization after an overnight fast. After anthropometric measurements (body weight, height and waist circumference), an intravenous cannula was inserted in an antecubital vein for withdrawal of blood for measurement of HbA_{1c}, serum insulin and adiponectin, plasma glucose, LDL- and HDL-cholesterol, triglyceride, total blood counts, albumin, AST, ALT, ALP, GGT and albumin concentrations and for genotyping of *PNPLA3* at rs738409 as described.[9] Blood sampling was performed prior to intake of any medications.

Histological assessment

Immediately at the beginning of the surgery, wedge biopsies of the liver were obtained. The biopsies from Belgium were sent to Finland, where they were assessed simultaneously with the Finnish samples by an experienced liver pathologist in a blinded fashion according to the criteria proposed by Brunt et al.[7] Liver fat was quantified as the percentage of hepatocytes with macrovesicular steatosis.

MS-based profiling

Lipidomic analysis. An unthawed plasma sample was used from all subjects and extracted for lipidomic and metabolomic analysis (*vide infra*). An established platform based on acquity ultra-performance liquid chromatography quadrupole time of flight mass spectrometry (UPLC-QTOFMS) was used to analyze the plasma samples. The data were processed by using MZmine 2

software[8] and the lipid identification was based on an internal spectral library (Supplementary Methods).

Metabolomic analysis. Polar metabolites are analyzed using comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry (GC×GC-TOFMS, Supplementary Methods).

Other analytical procedures and measurements

Body weight, waist circumference, blood pressure, and fasting concentrations of plasma glucose, serum free insulin, lipids (HDL and LDL cholesterol and TGs), liver enzymes (AST, ALT and GGT) concentrations were measured as previously described.[9] The MetS was defined and PNPLA3 at rs739409 was genotyped as described.[9]

Statistical analyses

Assessment of abundances of TG species. After log2 transformation, the average abundances of TG molecules were compared between the NASH and non-NASH groups by student's t-tests. Multiple comparisons were corrected by using Benjamini-Hochberg's method.[10] (Supplementary Methods)

Cluster analysis of lipidomics data. Lipids were grouped by using Bayesian model-based clustering as previously described.[11]

Diagnostic model. The biopsy subjects were randomly divided into estimation (n=223) and validation (n=95) groups to build and validate the model, respectively (vide infra). All study subjects (n=318) were used as the second validation group. Shapiro-Wilk's test was used to test the normality of the distribution. The unpaired t-test or Wilcoxon's rank sum test was used to compare the differences between the estimation and validation groups. Normally distributed data are shown

as means ± SEM and non-normally distributed data as median followed by the 25th and 75th quartiles. Multiple hypotheses testing was performed by using Benjamini-Hochberg's false discovery rate method to calculate q-values.[10] In the data with a large number of variables measured by UPLC, Gibbs sampling algorithm was used for simulation.[12] After sampling for 10000 times, the variables were selected amongst the models based on Bayesian information criterion (BIC). Logistic regression including all the selected variables was used to build the scores. Variables in the scores were finally assessed by backward stepwise regression to identify the optimal NASH score. The area under the ROC-curve (AUROC) was used to describe the diagnostic accuracy of the scores. The optimal cut-off point was calculated using the Youden index. The sensitivity, specificity, positive predictive values (PPV) and negative predictive values (NPV) for relevant cut-offs were calculated as described.[6] The AUROCs were compared using the generalized U-statistics.[13] One-way analysis of variance was used to compare three groups. Tukey's Honestly Significant Differences test was used for post hoc analyses. A two-sided p-value of less than 0.05 was considered as statistically significant. The statistical analyses were performed by using R version 3.0.1 (http://www.r-project.org/).

Results

Characteristics of the study groups

Comparison of NASH and non-NASH groups (Table 1)

Characteristics of the NASH and non-NASH groups are shown in Table 1. Liver fat content and all liver enzymes were significantly higher in the NASH than the non-NASH group (Table 1). Features of the MetS (hyperglycemia, hypertriglyceridemia, hypertension and low HDL cholesterol) were significantly more common in the NASH than the non-NASH group. The NASH group had a significantly increased prevalence of the PNPLA I148M variant compared to the non-NASH group ($p<0.001$). These significances remained significant after adjusting for age (data not shown).

Comparison of estimation and validation groups (Supplementary Table 1)

The estimation and validation groups were comparable with respect to clinical and biochemical features such as age, gender, BMI, components of the MetS, prevalence of NASH, liver fat, liver function tests, as well as PNPLA3 genotype (Supplementary Table 1). Within the estimation and validation cohorts (Supplementary Tables 2 and 3), the NASH group displayed similar abnormalities compared to the non-NASH group as was observed in the entire group (Table 1).

Development of a model to predict NASH

We first developed a model based on clinical parameters and the PNPLA3 genotype alone, then models based on MS-based profiling and finally a model using all data.

Model based on clinical parameters and I148M variant in PNPLA3 ('NASH Clin Score')

To build the '*NASH Clin Score'*, we used variables differing significantly between NASH and non-NASH groups in univariate analysis in the estimation group (Supplementary Table 2). The model was developed using multivariate logistic regression analysis based on clinical variables and PNPLA3 genotype. The final model included the same predictors as the 'NASH score', which was recently developed in a group of 296 Finnish patients and validated in a cohort of 380 Italians[12] i.e. fasting insulin, AST and PNPLA3 genotype. *'The NASH Clin Score'* was as follows:

$$-3.05 + 0.562 \times \textit{PNPLA3 genotype (CC=1/GC=2/GG=3)} - 0.0092 \times \textit{fS-insulin (mU/L)} + 0.0023 \times \textit{AST (IU/L)} + 0.0019 \times \textit{(fS-insulin}\times\textit{AST)}$$

The AUROC for the '*NASH Clin Score'* in the entire group (n=318) was 0.778 (95%CI: 0.709, 0.846).

Lipidomics and metabolomics data

Using the two MS-based analytical platforms, a total of 597 molecular lipids and metabolites were measured and 168 identified. Total fS-TG measured enzymatically was closely correlated with the sum of plasma TGs identified by UPLC-MS (r=0.92, $p < 0.001$).

Cluster analysis of lipidomics data

We analyzed the global lipidome by clustering the data into a subset of clusters using Bayesian model-based clustering. The lipidomic platform data were decomposed into 8 lipid clusters (LCs), which to a large extent adhered to different lipid functional or structural groups. Data on each cluster and representative lipids are shown in Table 2. In the NASH as compared to the non-NASH group, significant differences were found in 3 lipid clusters (LC3, LC4, LC6) (Supplementary Fig. 1). We found NASH to be significantly associated with increased concentrations of saturated and monounsaturated TGs (LC4). In contrast, concentrations of sphingomyelins (LC3) and lysophosphatidylcholines (lysoPC) (LC6) were significantly lower in the NASH than the non-NASH group (Supplementary Fig. 1, Table 2).

Absolute and relative concentrations of TGs

Absolute concentrations of circulating TGs between the NASH and non-NASH groups are compared in a heatmap (Fig. 1, panel on the left). In the NASH as compared to the non-NASH groups, the absolute concentrations of especially saturated and monounsaturated TGs such as TG(46:0), TG(48:0), TG(50:0), TG(46:1) and TG(51:1) were significantly increased.

The relative distribution of TGs (the concentration of an individual TG divided by total TGs measured by UPLC-MS) between the NASH and non-NASH groups is shown in Fig. 1 (panel on the right). The relative concentrations of saturated and monounsaturated TGs were increased in the NASH as compared to the non-NASH group (Fig. 1, on the right). Consistent with an increase in

TGs containing saturated and monounsaturated TGs, the fold-changes (NASH/non-NASH) of absolute (r = -0.75, $p < 0.0001$; Fig. 2, panel on the left) and relative (r = -0.75, $p < 0.0001$; Fig. 2, panel on the right) concentrations of TGs were inversely correlated with the number of double bonds.

Lipidomics-based model

Using the lipidomics data from the estimation cohort, we derived a logistic regression model (see methods) for NASH. The final score based on lipidomics data *('NASH Lip Score'*) included 3 selected molecular lipids, TG(48:0), PE(40:6) and LysoPC(16:0), and was as follows:

$$2.531 + 2.334 \times \log_{10}(TG(48:0))\ (\mu M) + 1.555 \times \log_{10}(PE(40:6))\ (\mu M) - 4.081 \times \log_{10}(LysoPC(16:0))\ (\mu M)$$

In the estimation group, the AUROC was 0.767 (95%CI: 0.687, 0.847). For the validation group (n=95; n=318), the AUROC was 0.809 (95% CI: 0.714, 0.905) and in the entire dataset 0.779 (95% CI: 0.717-0.841). The NPV, PPV, sensitivity and specificity of the entire dataset are shown in Table 3.

Metabolomics-based model

We also derived a logistic regression model for NASH using the metabolomics data from the estimation cohort. The score based on metabolomics *('NASH Met Score'*) included 5 selected molecular metabolites, Glu, Ile, Tyr, glycine (Gly) and serine (Ser). The *'NASH Met Score'* was as follows:

$$-10.701 + 1.852 \times \log_{10}(Glu)\ (\mu M) + 6.461 \times \log_{10}(Ile)\ (\mu M) + 3.556 \times \log_{10}(Tyr)\ (\mu M) - 3.908 \times \log_{10}(Gly)\ (\mu M) - 2.822 \times \log_{10}(Ser)\ (\mu M)$$

The model had an AUROC of 0.729 (95% CI: 0.649-0.808) in the estimation cohort. In the validation group (n=95), the AUROC was 0.710 (95% CI: 0.604 -0.816). In the entire group, the AUROC was 0.719 (95% CI: 0.655-0.782) (Table 3).

Model based on all data

By applying backward stepwise logistic regression analyses of the variables from all aforementioned models, we developed the *'NASH ClinLipMet Score'* (NASH score based on clinical variables, PNPLA3 genotype, lipidomics and metabolomics data, corrected for the number of variable included in the model, which was as follows:

$$-8.167 + 0.954 \times \textit{PNPLA3 genotype (CC=1/GC=2/GG=3)} + 0.0451 \times \textit{AST (IU/L)} + 0.0667 \times \textit{fS-insulin (mU/L)} - 3.151 \times \log_{10}(\textit{LysoPC(16:0)) (µM)} + 2.617 \times \log_{10}(\textit{PE(40:6)) (µM)} + 2.357 \times \log_{10}(\textit{Glu) (µM)} + 7.813 \times \log_{10}(\textit{Ile) (µM)} - 6.102 \times \log_{10}(\textit{Gly) (µM)}$$

The AUROC was 0.882 (95%CI: 0.827, 0.938) in the estimation and 0.856 (95% CI: 0.774, 0.938) in the validation cohort. In the entire group, the AUROC was 0.866 (95%CI: 0.820, 0.913). The sensitivity was 85.5% and specificity 72.1% (Table 3).

Diagnostic performances in the subgroups

Because bariatric patients might differ from non-bariatric patients, we excluded 54 patients not undergoing bariatric surgery and measured the performance of all scores in the specific group with bariatric patients. The AUROCs of the *'NASH Clin Score'*, *'NASH Lip Score'* and *'NASH Met Score'* in the bariatric*surgery* patients were 0.774 ((95%CI: 0.696, 0.852), 0.789 (95%CI: 0.720, 0.858) and 0.738 (95%CI: 0.672, 0.804). The *'NASH ClinLipMet Score'* had an AUROC of 0.865 (95%CI: 0.812, 0.918). The AUROCs of the *'NASH ClinLipMet Score'* did not differ significantly between the bariatric surgery group and entire cohort (p = 0.961).

To assess potential confounding effect of lipid lowering medications, we excluded 58 patients who received lipid medications and reanalyzed the diagnostic performance of all scores. In subjects not using lipid-lowering medications, the AUROCs of the *'NASH Clin Score'*, *'NASH Lip Score'*, *'NASH Met Score'* and *'NASH ClinLipMet Score'* were 0.799 ((95%CI: 0.725, 0.873), 0.816 (95%CI: 0.754, 0.878), 0.731 (95%CI: 0.660, 0.801) and 0.889 (95%CI: 0.844, 0.934) . The AUROC of the *'NASH ClinLipMet Score'* was not significantly different between the entire cohort and the group not using statins (p = 0.496).

Comparison of AUROCs

The AUROC of the *'NASH ClinLipMet Score'* was significantly higher than that of the *'NASH Lip Score'* ($p < 0.05$), the *'NASH Met Score'* ($p < 0.001$) and the *'NASH Clin Score'* ($p < 0.01$) ($p<0.001$) (Fig. 3). Performance of each score is summarized in Table 3.

Comparison of NASH to NAFL

We also determined whether the MS-based markers in the NASH group were specific to NASH or also observed between NAFL as compared to Non-NAFLD subjects. The clinical characteristics of NASH, NAFL and Non-NAFLD groups are shown in Table 1.

Comparison of lipid concentrations between the three groups are shown as Supplementary Fig. 1. Of the three lipids entering the final lipidomics model, TG48:0 (shown as TG(14:0/16:0/18:0)+TG(16:0/16:0/16:0) in Supplementary Fig. 2) differed significantly between NASH, and NAFL and NAFL and Non-NAFLD. PE40:6 and LysoPC16:0 differed significantly between NASH and NAFL but not between NAFL and controls (Supplementary Fig. 2).

Discussion

This is the first study to develop an MS-based model and score for NASH and compare its diagnostic performance to scores based on routinely available data and on PNPLA3 genotype at rs738409. We identified a set of lipids and metabolites that significantly associated with NASH in a liver biopsy cohort of 318 subjects. We performed Gibbs sampling and backward stepwise logistic regression to select variables that fulfilled BIC. A model which included AST, the PNPLA3 genotype, fasting insulin, LysoPC(16:0), PE(40:6), Glu, Ile and Gly best predicted NASH (the *'NASH ClinLipMet Score'*). The AUROC of this score was 0.86, which was significantly higher than that of the *'NASH Lip score'*, *'NASH Met Score'* and the *'NASH Clin Score'*,[6] 'NASH Liver Fat Score'[6, 9] and the 'NAFLD lipid triplet score'.[14] These data show that MS-based profiling combined with clinical variables may help in the development of non-invasive diagnosis of NASH.

The NASH as compared to the non-NASH group had an absolute and relative excess of saturated and monounsaturated TGs in their circulating lipidomics profile (Fig. 1) TGs containing saturated fatty acids (SFAs) and monounsaturated fatty acid (MUFAs) were previously shown to be overproduced in a study involving 9 subjects by the splanchnic area.[15] *De novo* lipogenesis produces exclusively SFAs.[16] Stable isotope studies tracing the origin of intrahepatocellular TGs suggest that *de novo* lipogenesis (DNL) is prominent and perhaps the only abnormal pathway in patients with NAFLD.[17] Hence, circulating TGs containing SFAs and MUFAs might reflect increased DNL. Individual TGs did not, however, remain significant independent predictors of NASH in the final model including both clinical and MS-profiling based parameters (*NASH ClinLipMet*). This is most likely because of multicolinearity i.e. saturated and monounsaturated TG were were closely correlated with features of IR such as fS-insulin and thus more markers of IR and steatosis than NASH.

Circulating LysoPC16:0 deficiency was associated with NASH. This metabolite as well as other lysoPCs and PCs (Supplementary Fig. 2), which were also deficient in NASH, are mostly found in the HDL lipoprotein fraction, which was low in the NASH as compared to other groups.[18] LysoPC16:0 was recently found to be the most deficient metabolite when comparing 180 metabolites between 20 insulin resistant and 20 insulin-sensitive morbidly obese subjects with NAFLD.[19] In 14 subjects who underwent a liver biopsy, lysoPC16:0 levels were higher in insulin resistant subjects with a trend toward higher inflammation in their liver.[19] Low lysoPC16:0 concentrations were also observed in preadipocyte cultures from 10 metabolically unhealthy as compared to 10 metabolically healthy obese subjects.[20]

The metabolite data are consistent with those reported in several small studies. Branch chain amino acids (BCAA) and essential amino acids are increased in obese/insulin-resistant subjects.[21] BCAA also promote IR induced by high fat feeding. Increases in BCAA are accompanied by increases in C3 and C5 acylcarnitines, which are BCAA metabolites in the liver and in skeletal muscle.[22] The increase in the BCAA Ile, and in Glu, which is the first step of BCAA catabolism, could therefore be attributed to the obesity/IR, which is associated with NASH. Increases in Glu have previously been found in studies that included 24[3] and 16[23] patients with NASH. Very recently, a genome-scale metabolomics model was constructed to interpret liver transcriptome data in NASH patients. Altered Glu metabolism was predicted to be the single most abnormal site of metabolism in NASH.[24] The second most common abnormality was predicted to be Ser deficiency, which is known to characterize patients with NASH and insulin resistant as compared to insulin sensitive subjects.[3, 25] Consistently, Ser deficiency also characterized the patients with NASH in the present study. Gly is formed from Ser in a reaction catalyzed by SHMT1, an enzyme leading to Ser formation that was predicted to be down-regulated in NASH.[24] Thus, the observed changes in amino acid concentrations in the NASH as compared to the non-NASH group reflect previously described pathophysiologic changes in humans and in experimental animals.

Use of the predictive equation developed in this study requires set-up of an assay specifically measuring each component. This is feasible given that an increasing number of analytical methods in clinical chemistry laboratories use targeted MS-based methods.[26] Once established and automated, such method requires a minute amount of plasma and is less time-consuming and cumbersome for clinicians than assessment of the different components of especially scores that necessitate inclusion of parameters from physical examination such as waist circumference.[27] Regarding the cost and reproducibility of the "omics" technology, it is important to establish whether screening using the *NASH ClinLipMet Score* is cost-effective. This can not be performed based on the present study, which is a first step and shows that it is possible to improve the diagnostic accuracy of predictive score for NASH using MS-based analytical platforms in morbidly obese patients with a high prevalence of NASH.

Limitations of the present study should be considered when interpreting the results. The score was derived from a cohort including a large number of obese patients, which may hamper its application to the general population. Although the histologic criteria for NASH are similar irrespective of obesity, it is important to validate the *NASH ClinLipMet Score* in a cohort which is not morbidly obese. Treatment with lipid-lowering drugs may influence their plasma lipid levels and play as a potential confounder. However, the performance of the *NASH Clin Score'*, *'NASH Lip Score'*, *'NASH Met Score'* and the *'NASH ClinLipMet Score'* was not influenced by use of lipid lowering medications, which thus suggest that the scores are robust and that use of lipid lowering medications does not limit usefulness of these scores. The study was cross-sectional and thus the term 'predictor' merely denotes a factor that is associated with risk of NASH. Scores should ideally be validated in a longitudinal study, but such a study is challenging as it is ethically unacceptable to obtain repeated liver biopsies from individuals with no indication for such a procedure. Although the *'NASH ClinLipMet Score'* had the highest AUROC of 0.86 in diagnosing NASH amongst the formulae tested with sensitivities and specificities of 80.6% and 75.3%, the diagnosis of NASH will

be missed in 19.4% of those with NASH and 24.7% will be incorrectly diagnosed as having the disease. The score developed in Finnish and Belgian Caucasian subjects may not be valid in other ethnic groups.

In conclusion, use of MS-based methods helps in improving non-invasive diagnosis of NASH compared to scores relying on routinely available clinical data and PNPLA3 genotype at rs738409. In the present study the findings of increases of the saturated TG 48:0, Glu, Ile, and decreases in lysoPC 16:0, Ser and Gly in a relatively large cohort of patients with NASH are consistent with known pathophysiology of NASH.

Acknowledgements

We thank Anne Salo, Aila Karioja-Kallio and Pentti Pölönen for their excellent technical assistance.

FIGURE LEGENDS

Fig. 1. Comparison of concentrations of TGs between NASH and non-NASH groups.
The color code denotes the log2 of the ratio between means of the groups for an individual TG (Left: absolute concentrations of TG. Right: relative concentrations of TG). The y-axes denote the number of carbons, and the x-axes the number of double bonds. Blue represents a decrease in NASH as compared to 'non-NASH'. The significances (*$p < 0.05$, **$p < 0.005$, ***$p < 0.0005$) for the comparisons are marked.

Fig. 2. Relationships between fold changes of mean concentrations of individual TGs (NASH/non-NASH) and numbers of double bonds contained in each TG. Each dot represents a TG molecule.
The y-axis denotes the log2 of of the ratio of concentrations of TGs between NASH and non-NASH patients, and the x-axis the number of double bonds in TGs. Left: log2 of fold changes of absolute TG abundances plotted against the number of double bonds. Right: log2 of fold changes of relative TG abundances plotted against the number of double bonds.

Fig. 3. ROC-curves of the three scores to predict NASH in the entire biopsy cohort. The AUROCs are compared in Table 3. Please see text for definition of scores.

References

1. Yki-Jarvinen H. Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome. The lancet.Diabetes & endocrinology 2014;2:901-910.

2. Pitt JJ. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. The Clinical biochemist.Reviews / Australian Association of Clinical Biochemists 2009;30:19-34.

3. Kalhan SC, Guo L, Edmison J, et al. Plasma metabolomic profile in nonalcoholic fatty liver disease. Metabolism: clinical and experimental 2011;60:404-413.

4. Barr J, Caballeria J, Martinez-Arranz I, et al. Obesity-dependent metabolic signatures associated with nonalcoholic fatty liver disease progression. Journal of proteome research 2012;11:2521-2532.

5. Luukkonen PK, Zhou Y, Sadevirta S, et al. Ceramides Dissociate Steatosis and Insulin Resistance in the Human Liver in Non-Alcoholic Fatty Liver Disease. Journal of hepatology 2016;64:1167-1175.

6. Hyysalo J, Mannisto VT, Zhou Y, et al. A population-based study on the prevalence of NASH using scores validated against liver histology. Journal of hepatology 2014;60:839-846.

7. Brunt EM, Janney CG, Di Bisceglie AM, et al. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. The American Journal of Gastroenterology 1999;94:2467-2474.

8. Pluskal T, Castillo S, Villar-Briones A, et al. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC bioinformatics 2010;11:395-2105-11-395.

9. Kotronen A, Peltonen M, Hakkarainen A, et al. Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and genetic factors. Gastroenterology 2009;137:865-872.

10. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Statistics in medicine 1990;9:811-818.

11. Zhou Y, Llaurado G, Oresic M, et al. Circulating triacylglycerol signatures and insulin sensitivity in NAFLD associated with the E167K variant in TM6SF2. Journal of hepatology 2015;62:657-663.

12. George EI and Robert EM. Variable selection via gibbs sampling. Journal of the American Statistical Association 1993;88:881-889.

13. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-845.

14. Oresic M, Hyotylainen T, Kotronen A, et al. Prediction of non-alcoholic fatty-liver disease and liver fat content by serum molecular lipids. Diabetologia 2013;56:2266-2274.

15. Westerbacka J, Kotronen A, Fielding BA, et al. Splanchnic balance of free fatty acids, endocannabinoids, and lipids in subjects with nonalcoholic fatty liver disease. Gastroenterology 2010;139:1961-1971.e1.

16. Aarsland A, Wolfe RR. Hepatic secretion of VLDL fatty acids during stimulated lipogenesis in men. Journal of lipid research 1998;39:1280-1286.

17. Lambert JE, Ramos-Roman MA, Browning JD, et al. Increased de novo lipogenesis is a distinct characteristic of individuals with nonalcoholic fatty liver disease. Gastroenterology 2014;146:726-735.

18. Kotronen A, Velagapudi VR, Yetukuri L, et al. Serum saturated fatty acids containing triacylglycerols are better markers of insulin resistance than total serum triacylglycerol concentrations. Diabetologia 2009;52:684-690.

19. Lehmann R, Franken H, Dammeier S, et al. Circulating lysophosphatidylcholines are markers of a metabolically benign nonalcoholic fatty liver. Diabetes care 2013;36:2331-2338.

20. Bohm A, Halama A, Meile T, et al. Metabolic signatures of cultured human adipocytes from metabolically healthy versus unhealthy obese individuals. PloS one 2014;9:e93148.

21. Felig P, Wahren J, Hendler R, et al. Splanchnic glucose and amino acid metabolism in obesity. The Journal of clinical investigation 1974;53:582-590.

22. Newgard CB. Interplay between lipids and branched-chain amino acids in development of insulin resistance. Cell metabolism 2012;15:606-614.

23. Li H, Wang L, Yan X, et al. A proton nuclear magnetic resonance metabonomics approach for biomarker discovery in nonalcoholic fatty liver disease. Journal of proteome research 2011;10:2797-2806.

24. Mardinoglu A, Agren R, Kampf C, et al. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. Nature communications 2014;5:3083.

25. Thalacker-Mercer AE, Ingram KH, Guo F, et al. BMI, RQ, diabetes, and sex affect the relationships between amino acids and clamp measures of insulin action in humans. Diabetes 2014;63:791-800.

26. Kenar E, Franken H, Forcisi S, et al. Automated label-free quantification of metabolites from liquid chromatography-mass spectrometry data. Molecular & cellular proteomics 2014;13:348-359.

27. Grandison GA, Angulo P. Can NASH be diagnosed, graded, and staged noninvasively? Clinics in liver disease 2012;16:567-585.

Table 1. Clinical characteristics of the study subjects according to liver fat (%) and a proposal by Brunt et al.

Total	**Non-NASH (n=249)**	**Non-NAFLD (n=132)**	**NAFL (n=117)**	**NASH (n=69)**
Liver fat/Steatosis (%)	5 (0 - 15)	0 (0 - 5)	15 (10 - 30)	40 (30 - 60)***†††
Microscopic steatosis %	10 (0 - 30)	0 (0 - 10)	30 (11 - 40)	40 (20 - 40)***†††
Grade (0/1/2/3)	249/0/0/0	132/0/0/0	117/0/0/0	0/57/11/1***
Ballooning	0	0	0	12***
Inflammation	0	0	0	69***
Fibrosis stage (0/1/2/3/4)	210/31/6/0/2	117/10/3/0/2	93/21/3/0/0	6/47/10/5/1***
Fibrosis stage > 0 (%)	15.7	11.4	20.5	91.3***
Age (years)	45.4 ± 0.8	45.1 ± 1.1	45.9 ± 1.1	49.4 ± 1.3
Gender (n, % women)	160 (64.2)	96 (72.7)	64 (54.7)	37 (53.6)
BMI (kg/m^2)	39.6 ± 0.6	39.0 ± 0.8	40.3 ± 0.8	41.1 ± 1.0
Waist circumference (cm)	118 ± 1	114 ± 1	122 ± 1	126 ± 2***†††
fS-Insulin (mU/l)	12.7 (8.0 – 18.1)	11.0 (7.5 – 15.9)	14.3 (9.5 – 18.9)	20.6 (14.3 – 28.7)***†††##
fP-Glucose (mmol/l)	4.9 (4.4 – 5.7)	4.8 (4.2 – 5.6)	5.1 (4.6 – 5.8)	5.7 (5.0 - 6.4)***†††###
HbA_{1C} (%)	5.6 (5.4 – 5.9)	5.5 (5.3 - 5.8)	5.7 (5.5 - 6.0)	6.0 (5.7 - 6.6)***†††##
HOMA-IR (mmol/l x mU/l)	2.8 (1.9 – 4.0)	2.4 (1.5-3.5)	3.2 (2.2-4.4)	5.0 (3.1-8.5)***†††###
fP-Triglycerides (mmol/l)	1.35 (1.03 – 1.92)	1.28 (0.97 - 1.78)	1.50 (1.04 - 1.95)	1.76 (1.26 – 2.54)***†††#
fP-HDL cholesterol (mmol/l)	1.25 (1.03 - 1.51)	1.27 (1.09 - 1.51)	1.19 (0.96 - 1.46)	1.07 (0.94 - 1.27)***†††
fP-LDL cholesterol (mmol/l)	2.90 ± 0.06	2.83 ± 0.08	2.92 ± 0.09	2.99 ± 0.12
P-AST (IU/l)	28 (24 - 38)	26 (22 - 33)	32 (27 - 42)	42 (29 - 58)***†††###
P-ALT (IU/l)	37 (30 - 51)	34 (27 - 42)	46 (33 - 62)	54 (40 - 89)***†††#
P-GGT (U/l)	28 (21 - 48)	27 (21 - 49)	30 (22 - 48)	47 (29 - 73)***†††
P-Albumin (g/l)	42.4 ± 0.3	42.1 ± 0.5	42.7 ± 0.5	41.7 ± 0.7
B-Platelets ($x10^9/l$)	259 (219 - 311)	264 (219 - 310)	258 (222 - 310)	240 (197 - 302)

PNPLA3 (CC/CG/GG) (n)	133/91/10	79/44/2	54/47/8	25/30/11**
Use of lipid-lowering drugs (%)	17.7	12.1	23.9	20.3
Type 2 Diabetes (n)	40	14	26	30****

Data are in n (%), means ± SEM or median (25th-75th percentile), as appropriate. *$P \leq 0.05$, **$P \leq 0.01$,***$P \leq 0.001$ for comparison with the 'Non-NASH'. †$p < 0.05$, ††$p < 0.01$, †††$p < 0.001$ for one-way ANOVA. #$p < 0.05$, ##$p < 0.01$, ###$p < 0.001$ for Tukey HSD test compared with the 'NAFL'. NASH: non-alcoholic steatohepatitis; NAFLD: non-alcoholic fatty liver; NAFLD: non-alcoholic fatty liver disease; BMI: body mass index; HOMA-IR: Homeostasis model assessment of insulin resistance was calculated by formula: fS-Insulin (mU/L) x fP-Glucose (mmol/L) / 22.5; HDL: High-density lipoprotein; LDL: low-density lipoprotein; AST: aspartate aminotransferase; ALT: alanine aminotransferase; GGT: gamma glutamyl transpeptidase; PNPLA3: patatin-like phospholipase domain-containing protein 3.

Table 2. Composition of circulating lipid clusters

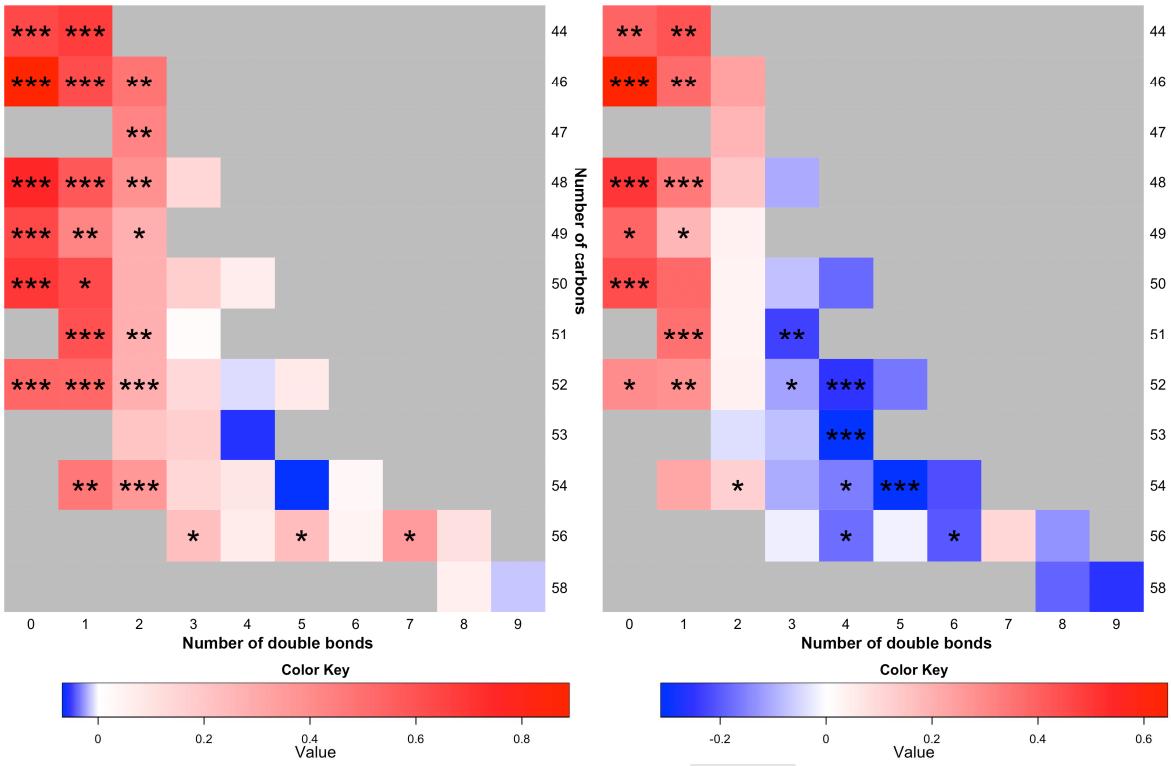
Cluster name	Size	Representative members
LC1	18	TG(16:0/18:2/18:1); TG(18:1/16:1/18:2)+TG(18:2/18:2/16:0); TG(18:1/18:2/18:1); TG(18:1/18:1/18:1); TG(54:5);
LC2	24	PC(34:2); PC(36:2); PC(34:1); PC(36:3); PC(38:3);
LC3	32	SM(d18:1/24:1); SM(d18:1/16:0); SM(d18:1/22:0); SM(d18:1/24:0); SM(d18:1/18:0); SM(d18:1/20:0); SM(d18:1/23:0); SM(d18:0/16:0); SM(d18:0/20:4);
LC4	23	TG(14:0/16:0/18:0)+TG(16:0/16:0/16:0); TG(16:0/16:0/18:0); TG(14:0/16:0/16:0)+TG(16:0/18:0/12:0); TG(44:0); TG(16:0/18:0/18:0); TG(44:1); TG(54:1)
LC5	15	PC(38:6); PC(40:6); PC(36:5); PE(40:6); PS(38:1); PS(36:1); PC(38:5); PE(38:5); PE(40:6)
LC6	18	LysoPC(16:0); LysoPC(18:2); LysoPC(18:0); LysoPC(18:1); LysoPC(18:3); LysoPC(20:3);
LC7	14	PC(38:7); PC(40:7); PE(38:4); PE(40:7); PE(40:6); PE(40:4);
LC8	16	PC(34:1e)+PE(37:1e); PC(33:2)+PE(36:2); PC(31:1)+PE(34:1); PC(33:1)+PE(36:1); PC(33:2)+PE(36:2)

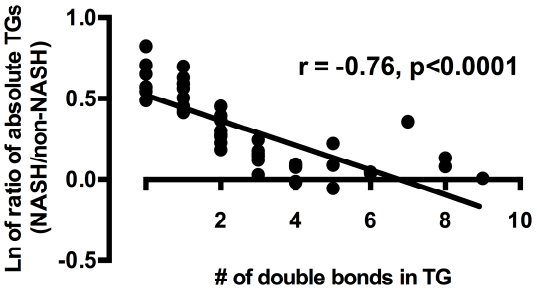
LC: lipid cluster; TG: triacylglycerol; PC: phosphatidylcholine; SM: sphingomyelin; LysoPC: lysophosphatidylcholines; PE: phosphatidylethanolamine.

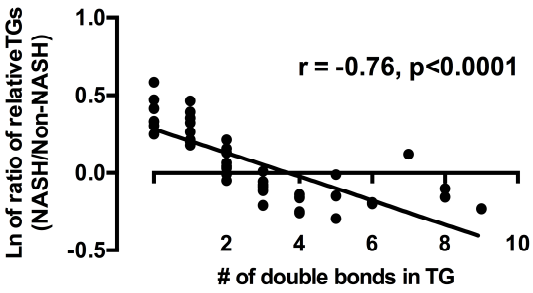
Table 3. Comparison of the performances of the scores for diagnosing NASH in 318 biopsy patients (see main text for statistical comparisons of AUROCs)

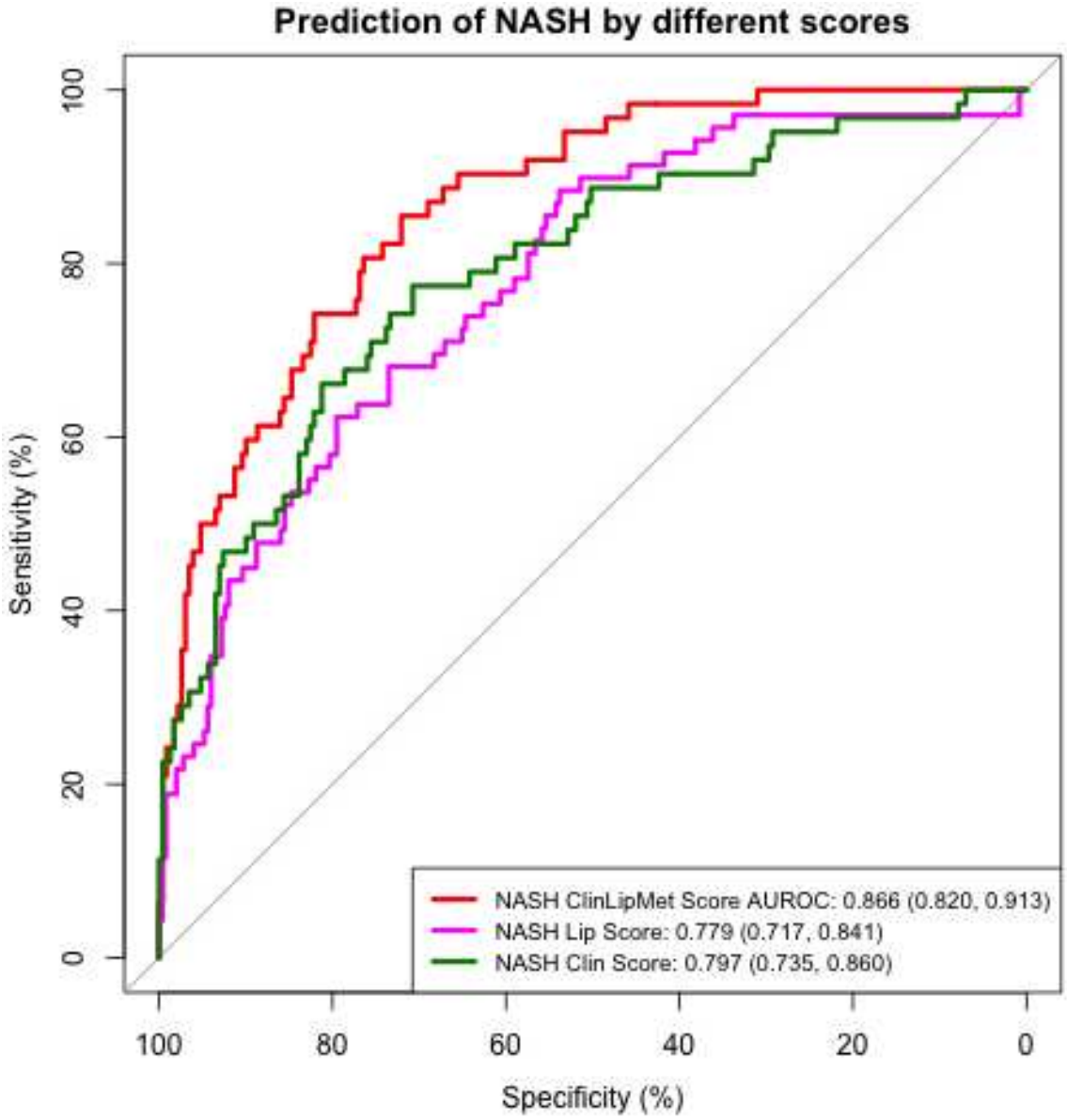
Test Scores	AUROC (95% CI)	Cut-off	Sens.%	Spec.%	PPV (%)	NPV (%)
NASH ClinLipMet Score	0.866 (0.820, 0.913)	0.134	85.5	72.1	45.3	94.8
NASH Lip Score	0.779 (0.717, 0.841)	0.148	88.4	53.8	34.7	94.4
NASH Met Score	0.719 (0.655, 0.782)	0.203	65.2	69.1	36.9	87.8
NASH Clin Score = NASH score (Hyysalo et al., 2013)	0.792 (0.726, 0.859)	-1.354	77.4	70.7	41.7	92.0

NASH: non-alcoholic steatohepatitis; NAFLD: non-alcoholic fatty liver disease; PPV: Positive predictive values. NPV: Negative predictive values.



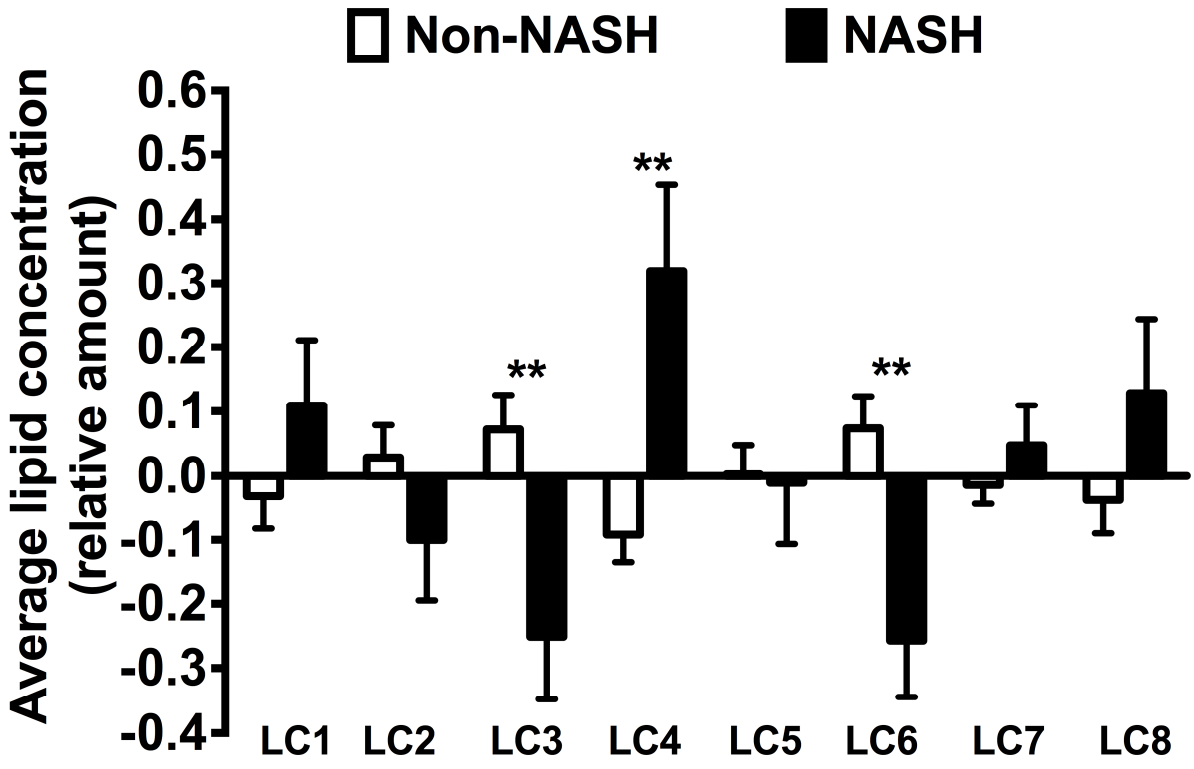


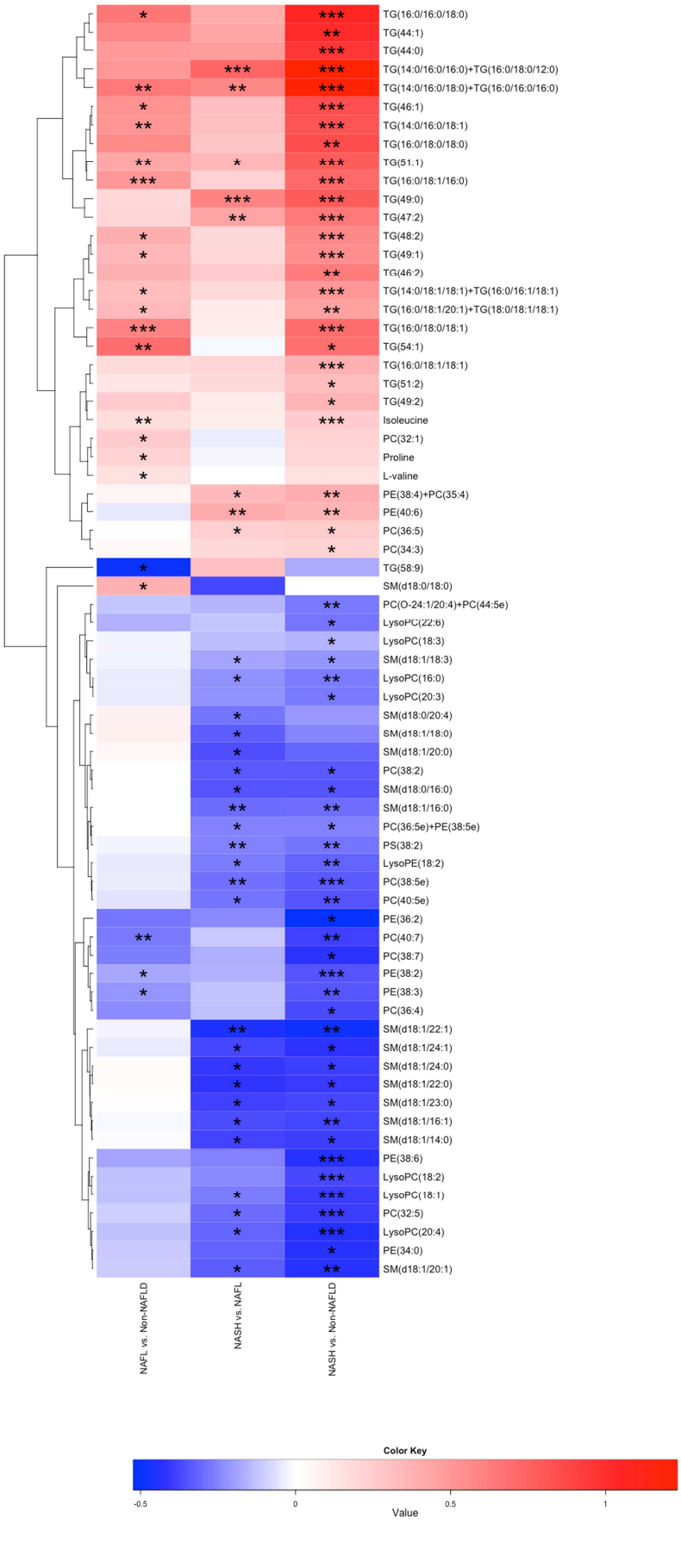



Prediction of NASH by different scores
Sensitivity (%)
100
80
60
40
20
0
Specificity (%)
100
80
60
40
20
0
NASH ClinLipMet Score AUROC: 0.866 (0.820, 0.913)
NASH Lip Score: 0.779 (0.717, 0.841)
NASH Clin Score: 0.797 (0.735, 0.860)

Supplementary Fig. 1. Mean lipid concentrations within each cluster between NASH (non-alcoholic steatohepatitis) and 'non-NASH' groups. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ for differences between the groups.

Supplementary Fig. 2. Comparison of concentrations of lipids between NASH, NAFL and Non-NAFLD groups. The color code denotes the log2 of the ratio between means of the groups for an individual lipid. The y-axe represents names of lipids, and the x-axe the groups for comparison (NAFL vs. Non-NAFLD, NASH vs. NAFL and NASH vs. Non-NAFLD). Tukey's Honestly Significant Differences post-hoc test was used to compare two groups after one-way analysis of variance. Blue represents a decrease while red shows an increase between groups. The brighter the red color, the greater increase of absolute concentration of the individual lipid between groups. The brighter the blue color, the greater decrease. The significances (*$p < 0.05$, **$p < 0.005$, ***$p < 0.0005$) for the comparisons are marked. NASH: non-alcoholic steatohepatitis. NAFL: nonalcoholic fatty liver. NAFLD: nonalcoholic fatty liver disease.





Supplementary Methods

Lipidomic analysis by UPLC-QTOFMS

An aliquot (10 µL) of the plasma sample was diluted with 10 µL of 0.15 M (0.9%) sodium chloride and 10 µL of internal standard mixture 1A was added. This mixture contained PC(17:0/0:0), PC(17:0/17:0), PE(17:0/17:0), PG(17:0/17:0)[rac], Cer(d18:1/17:0), PS(17:0/17:0) and PA(17:0/17:0) (Avanti Polar Lipids, Inc., Alabaster, AL) as well as monoacylglycerol (17:0/0:0/0:0), diacylglycerol (DG) (17:0/17:0/0:0) and TG(17:0/17:0/17:0). The lipids were extracted using a mixture of HPLC-grade chloroform and methanol (2:1; 100 µL). The lower phase (60 µL) was collected and 10 µL of an internal standard mixture containing labeled PC (16:1/0:0-D_3), PC(16:1/16:1-D_6) and TG(16:0/16:0/16:0-^{13}C3) was added.

The extracts were analyzed on a Waters Q-Tof Premier mass spectrometer combined with an Acquity Ultra Performance LC™. The column (at 50 °C) was an Acquity UPLC™ BEH C18 2.1 × 100 mm with 1.7 µm particles. The solvent system included A. ultrapure water (1% 1 M NH_4Ac, 0.1% HCOOH) and B. LC/MS grade acetonitrile/isopropanol (1:1, 1% 1M NH_4Ac, 0.1% HCOOH). The gradient started from 65% A / 35% B, reached 80% B in 2 min, 100% B in 7 min and remained there for 7 min. The flow rate was 0.400 ml/min and the injected amount was 2.0 µl (Acquity Sample Organizer, at 10 °C). Reserpine was used as the lock spray reference compound. The lipid profiling was carried out using electrospray ionization mode and the data were collected at a mass range of m/z 300-1200 with a scan duration of 0.2 sec.

The data processing included alignment of peaks, peak integration, normalization and identification. Lipids were identified using an internal spectral library. The data were normalized using one or more internal standards representative of each class of lipid present in the samples: the intensity of each identified lipid was normalized by dividing it with the intensity of its corresponding standard and multiplying it by the concentration of the standard. All monoacyl lipids except cholesterol

esters, such as monoacylglycerols and monoacylglycerophospholipids, were normalized with PC(17:0/0:0), all diacyl lipids except ethanolamine phospholipids were normalized with PC(17:0/17:0), all ceramides with Cer(d18:1/17:0), all diacyl ethanolamine phospholipids with PE(17:0/17:0), and TG and cholesterol esters with TG(17:0/17:0/17:0). Other (unidentified) molecular species were normalized with PC(17:0/0:0) for retention times < 300 s, PC(17:0/17:0) for a retention time between 300 s and 410 s, and TG(17:0/17:0/17:0) for longer retention times.

Quality control of the method showed that the day-to-day repeatability of control serum samples, and the relative standard deviation for values identified was on average below 25% and 20% for discovery and validation sets, respectively. The internal standards added to all samples in the study had an average relative standard deviation of 25% and 13 % in the discovery and validation sets.

For further identification of unknown lipids, fractions collected from UPLC run were infused to a LTQ-Orbitrap (Thermo Fischer Scientific, San Jose, CA) mass spectrometer by a TriVersa Nanomate (Advion Biosciences, Ithaca, NY) using chip-based nanoelectrospray in positive and negative ionisation mode. Identifications were based on the exact mass and MS^n spectra. The instrument was calibrated externally according to the instructions of manufacturer. MS^2 and MS^3 were acquired using either low resolution or high resolution up to target mass resolution R = 60 000 at m/z 400. The normalized collision energies of 30-40% were applied in MS^n experiments.

Metabolomic analysis

Polar metabolites are analyzed using using comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry (GC×GC-TOFMS).[1] 400 µl methanol and 10 µl internal standard mixture (C17:0 (93.3 mg/l), valine-d (18.5 mg/l) and succinic acid-d4 (31.5 mg/l)) were added to 30 µl of plasma samples. The samples were vortex mixed (2 minutes at 20 Hz). After 30 minutes at room temperature the samples were centrifuged for 5 min at 10000 rpm. The

supernatant was moved to a gas chromatography (GC) vial and evaporated to dryness under nitrogen. The samples were trimethylsilylated with 25 µl MOX (45°C, 60 minutes) and 25 µl MSTFA (45°C, 60 minutes) and 5 µl of retention index solution in hexane was added to samples (150 mg/l C11, C15, C17, C21 and C25 alkanes).

For metabolomics analysis, a Leco Pegasus 4D GC×GC-TOFMS instrument (Leco Corp., St. Joseph, MI) equipped with a cryogenic modulator was used. The GC part of the instrument was an Agilent 6890 gas chromatograph (Agilent Technologies, Palo Alto, CA), equipped with split/splitless injector. The first-dimension chromatographic column was a 10-m RTX-5 capillary column with an internal diameter of 0.18 mm and a stationary-phase film thickness of 0.20 µm, and the second-dimension chromatographic column was a 1.5 m BPX-50 capillary column with an internal diameter of 100 µm and a film thickness of 0.1 µm. A methyl deactivated retention gap (3 m x 0.53 mm i.d.) was used in the front of the first column. High-purity helium was used as the carrier gas at a constant pressure mode (39.6 psig). A 5-s separation time was used in the second dimension. The MS spectra were measured at 45 – 700 amu with 100 spectra/second. For the injection, a pulsed splitless injection (0.5 µl) at 240 °C was utilized, with pulse pressure of 55 psig for 1 minute. The temperature program was as follows: the first-dimension column oven ramp began at 40 °C with a 2 min hold after which the temperature was programmed to 295 °C at a rate of 7 °C/minute and then held at this temperature for 3 minutes. The second-dimension column temperature was maintained 20 °C higher than the corresponding first-dimension column. The programming rate and hold times were the same for the two columns.

ChromaTOF vendor software (LECO) was used for within-sample data processing, including quantitation of selected target metabolites, and Guineu software was used for alignment, normalization and peak matching across samples. The peaks were first filtered based on number of detected peaks in the total profile of all sample runs. The normalization was performed by

correction for internal standards. Other mass spectra from the GC×GC-TOFMS analysis were searched against The Palisade Complete Mass Spectral Library, 600K Edition (Palisade Mass Spectrometry, Ithaca, NY). Data were processed using the Guineu software.[1]

After non-targeted profiling of the discovery set, thirteen metabolites were selected for quantitative analysis (amino acids, free fatty acids) which were then quantified in both validation and discovery sets using external calibration curves, after normalization with the labelled group-specific internal standards. Quality control of the method showed that the day-to-day repeatability of control serum samples, and the relative standard deviation (RSD) for values identified was on average below 22% and 19% for discovery and validation sets, respectively. The internal standards added to all samples in the study had an average RSD of 20% and 18 % in the discovery and validation sets.

Statistical analyses

Assessment of abundances of TG species. Mean and standard errors of abundances of plasma TG molecular species were calculated. After log2 transformation, the average abundances of TG molecules were compared between the NASH and non-NASH groups by student's t-tests. Multiple comparisons were corrected by using Benjamini-Hochberg's method.[10] The comparisons were illustrated by heatmaps, which plot chain lengths of fatty acid against number of double bonds for each TG. R Package, metadar (http://code.google.com/p/metadar) was used for data analysis.

References

1. Castillo S, Mattila I, Miettinen J, Oresic M, Hyotylainen T. Data analysis tool for comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry. Anal Chem 2011;83:3058-3067.

Supplementary Tables

Supplementary Table 1. Comparison between the clinical characteristics in the estimation and validation group for building of the NASH score

Total	**Estimation group (n=223)**	**Validation group (n=95)**	***p-value***
Liver fat (%)	15 (5-40)	12.5 (0-47.5)	0.9
Grade (0/1/2/3) (n)	176/40/7/0	71/19/4/1	0.41
Stage (0/1/2/3/4) (n)	150/55/12/4/2	65/23/5/1/1	0.99
NASH (n, %)	47 (21.1)	24 (25.3)	0.5
Age (years)	46±1	47±1	0.79
Gender (n, % women)	140 (62.8)	57 (60)	0.73
BMI (kg/m^2)	39.9±0.5	39.8±1.0	0.61
fP-Glucose (mmol/l)	4.9 (4.4-5.8)	5.0 (4.5-5.7)	0.55
HbA_{1C} (%)	5.7 (5.4-6)	5.7 (5.4-6)	0.92
fP-Triglycerides (mmol/l)	1.46 (1.09-2.18)	1.36 (0.96-2.03)	0.39
fP-HDL cholesterol (mmol/l)	1.17 (0.99-1.46)	1.22 (1.02-1.46)	0.37
fP-LDL cholesterol (mmol/l)	2.9 (2.19-3.51)	2.81 (2.24-3.49)	0.72
P-AST (IU/l)	29 (25-42)	32 (24-40)	0.67
P-ALT (IU/l)	40 (31-59)	42 (32-56)	0.64
P-GGT (U/l)	30 (22-54)	31 (23-54)	0.63
P-Albumin (g/l)	41.5±0.4	42.5±0.6	0.25
B-Platelets ($x10^9$/l)	260 (214-311)	256 (215-303)	0.67
PNPLA3 (CC/GC/GG) (n)	108/85/17	50/35/4	0.49
Type 2 Diabetes (n, %)	43 (19.5)	18 (18.9)	1
Metabolic Syndrome (n, %)	145 (65.3)	62 (65.3)	1

Data are in n (%), means ± SEM or median (25th-75th percentile), as appropriate. BMI: body mass index; HDL: High-density lipoprotein; LDL: low-density lipoprotein; AST: aspartate aminotransferase; ALT: alanine aminotransferase; GGT: gamma glutamyl transpeptidase; PNPLA3: patatin-like phospholipase domain-containing protein 3.

Supplementary Table 2. Clinical characteristics of the NASH and the non-NASH groups in the estimation group

Estimation Group	**Non-NASH (n=176)**	**NASH (n=47)**	***p-value***
Liver fat (%)	10 (0, 25)	40 (26.3, 67.5)	<0.001
Grade (0/1/2/3) (n)	176/0/0/0	0/40/7/0	<0.001
Stage (0/1/2/3/4) (n)	145/24/5/0/2	5/31/7/4/0	<0.001
Age (years)	46±1	48±2	0.21
Gender (n, % women)	115 (65.3)	25 (53.2)	0.17
BMI (kg/m^2)	39.5±0.6	41.4±1.1	0.15
fP-Glucose (mmol/l)	4.8 (4.4-5.7)	5.7 (5-6.5)	<0.001
HbA_{1C} (%)	5.6 (5.4-6)	6 (5.8-6.7)	<0.001
fP-Triglycerides (mmol/l)	1.38 (1.07-1.95)	1.84 (1.36-2.54)	0.005
fP-HDL cholesterol (mmol/l)	1.22 (1.01-1.51)	1.07 (0.95-1.25)	0.012
fP-LDL cholesterol (mmol/l)	2.89 (2.17-3.52)	3 (2.32-3.51)	0.46
P-AST (IU/l)	28 (24-38)	43 (28-71)	<0.001
P-ALT (IU/l)	37 (30-52)	53 (40-103)	<0.001
P-GGT (U/l)	28 (21-48)	48 (29-74)	<0.001
P-Albumin (g/l)	41.6±0.5	41.3±0.9	0.89
B-Platelets ($x10^9/l$)	264 (219-314)	243 (199-296)	0.11
PNPLA3 (CC/CG/GG) (n)	90/66/10	18/19/7	0.064
Type 2 Diabetes (n, %)	24 (13.6)	19 (42.2)	<0.001
Metabolic Syndrome (n, %)	106 (60.2)	39 (84.9)	0.0033
Hyperglycemic medication (n, %)	18 (10.2)	15 (31.9)	<0.001

Data are in n (%), means ± SEM or median (25th-75th percentile), as appropriate.

Supplementary Table 3. Clinical characteristics of the NASH and the non-NASH groups in the validation group

Validation Group	**Non-NASH (n=71)**	**NASH (n=24)**	***p-value***
Liver fat (%)	5 (0-31.3)	42.5 (30-75)	<0.001
Grade (0/1/2/3) (n)	71/0/0/0	0/19/4/1	<0.001
Stage (0/1/2/3/4) (n)	63/6/2/0/0	2/17/3/1/1	<0.001
Necroinflammation (0/1/2) (n)	70/1/0	0/15/9	<0.001
Age (years)	45±2	52±2	<0.001
Gender (n, % women)	43 (60.6)	14 (58.3)	1
BMI (kg/m^2)	39.4±1.2	40.8±1.7	0.36
fP-Glucose (mmol/l)	5.0 (4.6-5.6)	5.5 (4.5-6.3)	0.2
HbA_{1C} (%)	5.6 (5.4-6.0)	5.9 (5.6, 6.4)	0.042
fP-Triglycerides (mmol/l)	1.27 (0.94-1.82)	1.86 (1.28-2.77)	0.021
fP-HDL cholesterol (mmol/l)	1.25 (1.07-1.51)	1.05 (0.91-1.29)	0.028
fP-LDL cholesterol (mmol/l)	2.81 (2.19-3.5)	2.7 (2.4-3.48)	0.71
P-AST (IU/l)	31 (24-38)	40 (30-48)	0.0047
P-ALT (IU/l)	39 (31-51)	54 (40-60)	0.012
P-GGT (U/l)	29 (22-56)	41 (29-52)	0.18
P-Albumin (g/l)	42.9±0.7	41.5±1.2	0.46
B-Platelets (x109/l)	256 (219-296)	248 (195-325)	0.65
PNPLA3 (CC/CG/GG) (n)	42/24/0	8/11/4	<0.001
Type 2 Diabetes (n, %)	8 (11.3)	10 (41.7)	0.0028
Metabolic Syndrome (n, %)	41 (57.7)	21 (87.5)	0.016
Hyperglycemic medication (n, %)	6 (8.4)	7 (30.4)	0.021

Data are in n (%), means ± SEM or median (25th-75th percentile), as appropriate.