The altered serum lipidome and its diagnostic potential for Non-Alcoholic Fatty Liver (NAFL)-associated hepatocellular carcinoma

Monika Lewinska, PhD,a, Alvaro Santos-Laso, PhDb, Enara Arretxe, PhDc, Cristina Alonso, PhDb,c, Ekaterina Zhuravleva, PhDb, Raul Jimenez-Aguero, MD, PhDb, Maria Jesus Pareja, MD, PhDb, Manuel Romero-Gomez, MD, PhDb,c, Malte P. Suppli, MD, PhDb, Filip K. Knop, MD, PhDb,i, Stine Karlsen Oversoe, MD, PhDb, Gerda Elisabeth Viladsen, MD, PhDb, Thomas Decaens, MD, PhDb,k,l, Ekaterina Zhuravleva, PhDa, Raul Jimenez-Aguero, PhDb, Filip K. Knop, MD, PhDb,i, Stine Karlsen Oversoe, MD, PhDb, Jesper B. Andersen, PhDa,*

ABSTRACT

Background: Non-alcoholic fatty liver disease (NAFLD) is affecting more people globally. Indeed, NAFLD is a spectrum of metabolic dysfunctions that can progress to hepatocellular carcinoma (NAFLD-HCC). This development can occur in a non-cirrhotic liver and thus, often lack clinical surveillance. The aim of this study was to develop non-invasive surveillance method for NAFLD-HCC.

Methods: Using comprehensive ultra-high-performance liquid chromatography mass-spectrometry, we investigated 1,295 metabolites in serum from 249 patients. Area under the receiver operating characteristic curve was calculated for all detected metabolites and used to establish their diagnostic potential. Logistic regression analysis was used to establish the diagnostic score.

Findings: We show that NAFLD-HCC is characterised by a complete rearrangement of the serum lipidome, which distinguishes NAFLD-HCC from non-cancerous individuals and other HCC patients. We used machine learning to build a diagnostic model for NAFLD-HCC. Patients' metabolic landscapes show a progressive depletion in unsaturated fatty acids and AC, acylcarnitines; AV-HCC, alcohol- and viral-associated HCC; AFP, alpha-fetoprotein; AA, amino acids; ArAA, aromatic amino acids; BA, bile acids; BMI, body mass index; BCAAs, branch-chain amino acids; Cer, ceramides; Chol, cholesterol esters; CTRL, control; DEM, differentially expressed metabolites; DG, diglycerides; LPC, lysophosphatidylcholines; LPE, lysophosphatidylethanolamines; LPL, lipoprotein lipase; MUFAs, monounsaturated fatty acids; OA-NAFLD, morbidly obese bariatric surgery NAFLD; non-alcoholic fatty liver disease; NAFL, non-alcoholic fatty liver; NAE, N-acyl ethanolamines; oxFAs, oxidized fatty acids; PC, phosphatidylcholines; PE, phosphatidylethanolamines; PI, phosphatidylinositols; PUFA, polyunsaturated fatty acids; SFA, saturated fatty acids; SM, sphingomyelins; S1P, sphingosine-1-phosphate; TG, triglycerides
Introduction

Evidence before this study

Non-alcoholic fatty liver disease (NAFLD) is a group of metabolic disorders that can progress to hepatocellular carcinoma (HCC) often without a background of cirrhosis. Non-cirrhotic NAFLD patients are not under surveillance for HCC development. Some metabolomic studies distinguish stages of NAFLD (steatosis, steatohepatitis and cirrhosis). However, NAFLD versus early hepatocarcinogenesis (NAFLD-HCC) in a metabolic background has not been investigated.

Added value of this study

This is the first metabolomic study with comprehensive serum lipidomics, investigating a broad-range of lipid classes in NAFLD and NAFLD-HCC. This study provides evidence of lipidomic utility in NAFLD-HCC detection, presents the NAFLD-HCC Diagnostic Score (NHDS) as a putative screening tool for HCC development in a metabolic background and provides a comprehensive lipidomic landscape of NAFLD-HCC patients serum. We show that fatty acid (FA) uptake is increased in NAFLD-HCC tumours, leading to FA depletion in the circulation.

Implications of all the available evidence

Nearly 30% of newly diagnosed HCC patients have underlying metabolic liver disease. Unlike in the case of viral hepatitis or alcoholic-related HCC, NAFLD-HCC often develops without cirrhosis and thus, patients are not under surveillance. This demonstration of effectiveness using a non-invasive, blood-based assay in detection of resectable (early) NAFLD-HCC should encourage consideration for surveillance of this rapidly increasing, at-risk population.

Non-alcoholic fatty liver disease (NAFLD) involves a spectrum of metabolic diseases affecting 24% of the population globally [1]. NAFLD is associated with obesity and type 2 diabetes [2], and represents a spectrum of non-malignant conditions that range from hepatic steatosis to non-alcoholic steatohepatitis (NASH) with fibrosis and ultimately cirrhosis [3]. NAFLD is emerging as a leading risk factor of hepatocellular carcinoma (HCC) in both men and women (hazard ratio ~17) [4], but tools to detect the progression of NAFLD to HCC remain inadequate.

HCC is the fourth most common cause of cancer-related death worldwide [5] with rapidly increasing incidence and mortality rates [6,7]. The incidence of NAFLD-related HCC (NAFLD-HCC) varies between 0-5% and 19.5%, depending on the disease state (presence or absence of cirrhosis) and geography [7-9]. Indeed, the prevalence of NAFLD-HCC is increasing compared to alcohol- and viral hepatitis-related HCC (AV-HCC) [8,10]. Importantly, accumulating evidence suggests that NAFLD-HCC may develop in a non-cirrhotic background [9,11]. This presents a major clinical challenge as non-cirrhotic NAFLD patients currently are not under surveillance for developing HCC [7,11]. Furthermore, NAFLD-HCC is characterised by unique mutational signatures and an immune environment, which plays an important role in the response to checkpoint inhibitors and may lead to therapeutic failure [12-14]. Therefore, non-invasive serum biomarkers are urgently needed to monitor NAFLD, its progression to HCC, and to distinguish NAFLD-HCC from other aetiologies.

Serum metabolomics provides insight into holistic metabolic changes and can potentially predict hepatic pathological lipid abundance [15] and is reflective of the liver pathology [16]. However, metabolomic profiles are highly dependent on ethnicity and risk factors such as obesity and diabetes [17-20]. The understanding of the biology underlying NAFLD-to-HCC progression is still lacking. Therefore, analysis of the metabolic shifts in the liver and in circulation at the stage of NAFLD and during malignant progression (NAFLD-HCC) is crucial for the development of new diagnostics and therapeutic modalities [21].

The aim of this study was to develop and validate a non-invasive biomarker for surveillance and early detection of NAFLD-HCC. Here, we used a comprehensive serum metabolomics-based approach to identify unique NAFLD-HCC biomarkers, allowing us to distinguish early NAFLD-HCC from patients with NAFLD or HCC on a background of alcohol or viral hepatitis (AV-HCC).

Patients and Methods

2.1. Study population

A total of 249 patient samples were collected in this international and multicentre study. All samples were collected and stored according to REMARK for biomarker analysis. Our study cohort was divided into discovery and validation sets. The proportion of NAFLD-HCC patients in this study (17.3%) is reflecting the reported prevalence of NAFLD-HCC [7-9]. The discovery set included serum samples from 196 patients from Spain and France: 27 patients with NAFLD-HCC, 32 patients with AV-HCC, 102 morbidly obese NAFLD patients undergoing bariatric surgery (OB-NAFLD), and 35 healthy subjects described previously [22]. Based on unsupervised clustering (Supplementary Materials) and clinical information, we classified 9 obese bariatric surgery patients with a NAS<3 and liver fibrosis score <2 as controls (CTRL). The validation set included serum samples from 37 NAFLD patients from Chile and Spain and plasma sampled from 16 NAFLD-HCC patients in Brazil and Denmark. NAFLD patients from the validation set were overweight (body mass index (BMI)=25) or obese (BMI > 30), however, significantly leaner compared to patients in the OB-NAFLD group (p<0.0001). All patients had biopsy-proven NAFLD. The NAFLD-HCC group consisted of patients who were diagnosed non-alcoholic steatohepatitis (NASH) based on liver biopsy, self-reported alcohol consumption (~20 g/day) and hepatitis B or C serology (hepatitis B surface antigen, hepatitis B surface antibody, hepatitis B core antibody, and hepatitis C antibody). All patients were
qualified for curative liver resection. Serum was obtained from all patients prior to surgery, including patients undergoing bariatric surgery or liver resection. The clinical and biochemical representation of the study population is presented in Supplemental Table 1. The NAFLD-HCC patients were diagnosed before surgery with imaging (33%), biopsy (26%), both (15%), or medical concilium (11%). For remaining, 15% of patients’ diagnostic method was not reported. After removal tumours were classified by a skilled pathologist. In all patients’ blood was collected before surgical intervention.

2.2. Ethics

The study was performed following individual patient consent, local institutional review board (IRB) approval, and assessed by the Committee on Health and Research Ethics for the Capital Region of Denmark for use of archival material no. 17029679. All patient data sets were anonymised.

2.3. Metabolomic analyses

Comprehensive metabolomics including 1295 metabolites was performed on serum samples with three platforms as described previously [23]. Briefly, metabolite extraction was accomplished by fractionating the samples into pools of species with similar physicochemical properties, using appropriate combinations of organic solvents. Two UHPLC-time of flight-MB based platforms analysing methanol and chloroform/methanol serum extracts were combined with the amino acid measurement using an UHPLC-single quadrupole-MS based analysis. Platform used for the analysis methanol extraction was optimized for the profiling of fatty acids, oxidized fatty acids, acyl carnitines, lysoglycerophospholipids (monacylglycerophospholipids and monoetherglycerophospholipids), free sphingoid bases, bile acids, and steroid sulfates. The chloroform/methanol extract platform provided coverage over glycerolipids (di- and triglycerides), cholesterol esters, sphingolipids (ceramides and sphingomyelins), and glycerophospholipids (diacylglycerophospholipids and 1-ether, 2-acylglycerophospholipids).

Metabolite extraction procedures, chromatographic separation conditions, and mass spectrometric detection conditions have been previously described [23].

Data pre-processing generated a list of chromatographic peak areas for the metabolites detected in each sample injection. An approximated linear detection range was defined for each identified metabolite, assuming similar detector response levels for all metabolites belonging to a given chemical class represented by a single standard compound. Metabolites for which more than 30% of data points were found outside their corresponding linear detection range were not used for statistical analyses. Intra and inter batch data normalisation was performed following the procedure described by Martinez-Arranz et al [24].

For fatty acid profiling we used tissue samples matched to serum samples from 9 CTRL, 32 NAFLD, 27 surrounding liver tissues, and 27 tumour tissues. Methanol with internal standards was added to liver tissue (30:1, v/w) and the homogenization of the tissues, and 27 tumour tissues. Methanol with internal standards was prepared individually in methanol at a concentration level of approximately 100, 1000 or 10000 µg/mL, depending on the compound. The working standard was prepared by mixing the appropriate amount of each standard solution in methanol to reach a final concentration of approximately 100 µg/mL. Calibration standards were prepared by consecutive dilutions in methanol, ranged between 0-0005-5 µg/mL for FAAx13 and 0-0005-50 µg/mL for the rest of the standards in the calibration curve. Likewise, internal standards (IS) stock solution of Octadecanoyl (18,18,18-D3)-L-Carnitine, Linoleic-9,10,12,13-D4 acid, Hexadecatrienoic 7 (Z),10 (Z),13-D4 acid, 18:1-D7 Lyso PC were prepared at a concentration of 1000 or 5000 µg/mL. The IS working solution was prepared by mixing the appropriate amount of each stock solution in methanol to reach the desire concentration. Quality control (QC) samples (reference sera samples commercially available) were treated according to the same protocol that serum and plasma samples and were analysed against the calibration curve. The intra-day precision was determined by analysing five replicates.

To estimate the concentration of validated metabolites in the discovery set, the measured absolute values of metabolites (normalised to IS and standard curves) were plotted against the normalised chromatographic peak areas for 58 samples. Simple linear regression was applied to generate the equation for each metabolite and estimate R-square (Goodness of fit). The equations used for concentration estimation are presented in Supp. Table 2.

2.4. Gene expression analysis

To estimate expression of fatty acid transport and metabolism genes we used 48 NAFLD-HCC tumour tissues and 47 matched, surrounding tissues. RNA was isolated from 20 mg of frozen tissue with AllPrep DNA/RNA/Protein Mini Kit (Qiagen) according to manufacturers recommendation and sequenced Illumina PE150 (Novogene). Fastq files were trimmed with Trimomatic-0-2 36 (TruSeq-PE illumina adapters cut). Reads were annotated with STAR-2.5.1a to hg38 canonical genome assembly 1 pass mode, with default parameters. Gene reference (Gencode.v30) was used for gene abundancy estimation and is normalised by gene length (transcript per million, TPM). Differential expression between tumour and surrounding tissue was established with t-test (p<0.05). The Level 3 TCGA gene expression data (LIHC.maseqv2__illuminahiseq_maseqv2) from TCGA [25] were downloaded from Firebrowse (http://firebrowse.org/). Clinical data were obtained from (https://www.cbioportal.org/). Total of 12 tumour samples had identified NAFLD as one of aetiologies. For surrounding liver tissue only two matched NAFLD samples were available in the dataset, thus we included ‘Other’ (n=4) and ‘No History of Primary Risk Factors’ (n=19) as controls.

2.5. Statistics

We identified 0-3% of missing values in the matrix with the maximum of missing values in one feature equal to 15%. The missing values were estimated by the k-nearest neighbour method with Metaanoa 4.0 [26]. Data were quantile-normalised, log-transformed before analysis (Metaanoa 4.0). The covariate testing and correction was performed with linear regression prior to analysis, and age, sex and BMI were included as covariates in differential expression analysis (R 4.0.4) The outline of statistical approaches is presented in Figure 1a.

Differences in biochemical parameters were established with one-way ANOVA, Kruskal-Wallis test with Tukey’s multiple comparisons, or Fisher exact test for categorical data (Prism 8.2.0).
The covariate testing and correction was performed with linear regression prior to analysis, and age, sex and BMI were included as covariates in differential expression analysis (R 4.0.4).

Data sets per metabolic class were calculated as the sum of the normalised areas of all the metabolites with the same chemical characteristics. The outlier analysis was performed before one-way ANOVA or Kruskal-Wallis test with Tukey’s or Dunn’s multiple testing (Prism 8.2.0) depending on data normality.

Multivariate analyses were performed with package mixOmics [27] or Metaboanalyst 4.0. The sparse partial least squares discriminant analysis (sPLS-DA model) was used to investigate the best separation between the groups for multiple groups and orthogonal partial least squares discriminant analysis (oPLS-DA) for pair-wise comparison.

Hierarchical cluster analysis was performed with Ward’s algorithm using Euclidean distances.

The receiver operating characteristic (ROC) curves were generated for pair-wise comparison and areas under the curves (AUC) were estimated with Metaboanalyst 4.0 or Prism 8.2.0.

The diagnostic model was constructed using ROC curves generated by Monte-Carlo cross-validation (MCCV) using balanced sub-sampling. In each MCCV, two-thirds (2/3) of the samples were used to evaluate the feature importance and the classification model was validated on 1/3 of the samples that were left out. The procedure was
repeated 50 times to calculate the performance and confidence interval of each model detecting the optimal number of features for best accuracy. The linear support vector machine (SVM) method was used for sample classification. Feature selection was performed with SVM Mean Importance Measure. The average accuracy of the model was based on 100 cross validations. Model performance was measured with 1000 permutations and both area under the ROC curve and predictive accuracy p < 0.001 (Metaboanalyst 4.0).

**NAFLD-HCC diagnostic score (NHDS).** In the discovery set of 196 samples, 8 metabolites [1] had commercial standards allowing absolute quantification, [2] highly correlated between measured concentration and peak areas. These were used to build an SVM model to identify the optimal number of metabolites and select features (Mean Importance Measure) distinguishing NAFLD-HCC. The concentrations of selected metabolites were used to build a logistic regression model (with an additional 10-fold Cross Validation) retaining metabolites in the equation with p < 0.05. Model performance was tested with 1000x permutations and both AUC and predictive accuracy reaching p < 0.001 (Metaboanalyst 4.0). A logistic model was generated based on metabolite concentrations in the discovery set data establishing the NAFLD-HCC diagnostic score (NHDS) and the cut-off value. The NHDS was calculated for the validation set. ROC curves between the groups were computed and compared (Prism 8.2.0).

The odds ratio and relative risk for NHDS, clinical and biochemical features were established using Prism 8.2.0. We generated confusion matrices for the features in discovery and validation sets as well as pulled patient sets using a cut-off value of age (50 years), BMI=30, and reference ranges for biochemical features. Statistical significance was measured with Fisher’s exact test (p < 0.05), the relative risk was calculated with Koopman asymptotic score and odds ratio with the Baptista-Pike method.

**Role of covariates on the performance of NHDS.** Patients were stratified into groups according to the underlying condition (diabetes, dyslipidaemia, cirrhosis (NAFLD-HCC)), reported medication (statin and paracetamol) use or NAFLD risk-genotype of Patatin-like phospholipase domain-containing protein 3 (PNPLA3), Transmembrane 6 Superfamily Member 2 (TM6SF2), Membrane-Bound O-acyltransferase Domain Containing 7 (MBOAT7).

Pearson correlation was applied for pattern hunter analysis to detect metabolites significantly correlating with progressive increase or decline. Metabolites were considered as correlated with pattern when (q < 0.05) and abs (r) > 0.3.

**Pathway overrepresentation analysis** was performed with the Integrated Molecular Pathway Level Analysis (IMPaLa) [28]. As the majority of our metabolites belong to lipids, which to a high extend are not included in metabolic pathways, we used all detected serum metabolites with HMDB identifiers as background in the enrichment analysis. Lipid pathway enrichment analysis was performed with BioPAN software and default settings.

### 2.6. Role of funding sources

The funding sources for this study had no role in study design, data collection, data analysis or interpretation nor writing of the report as it is presented herein.

### 3. Results

#### 3.1. Clinical and biochemical features of studied populations

We analysed the metabolic composition of serum obtained from 249 patients divided into a discovery set comprising patients with NAFLD-HCC (n=27), non-cancerous individuals (n=137) and 32 patients with alcohol- or viral-related HCC (AV-HCC). The control (CTRL) group included 44 individuals (35 healthy subjects [22] and 9 bariatric surgery patients with a NAFLD-activity score (NAS)<3 and fibrosis score <2) and 93 morbidly obese NAFLD patients awaiting bariatric surgery (OB-NAFLD). Furthermore, we independently validated our findings in serum from 37 patients with NAFLD and in plasma from 16 patients with NAFLD-HCC. Importantly, all NAFLD-HCC patients had no prior history of viral hepatitis or excessive alcohol consumption. The clinical, pathological, and biochemical features of all patients are summarized in Supp. Table 1 and Supp. Fig. 1.

We observed a significant difference in mean age (NAFLD-HCC patients were significantly older compared to CTRL, OB-NAFLD and AV-HCC), BMI (significantly higher in OB-NAFLD compared to CTRL, AV-HCC and NAFLD-HCC), as well as gender ratio (significantly higher prevalence of HCC among males, but equal between NAFLD-HCC and AV-HCC). Nevertheless, unsupervised principal component analysis (PCA) showed that the metabolic profiles grouped independently of these covariates (Supp. Fig. 2). Additionally, the presence of underlying diabetes, cirrhosis, and level of fibrosis did not affect patients’ grouping in unsupervised clustering. Further, the analysis of alpha-fetoprotein (AFP) or other liver biochemical features (alanine aminotransferase (ALT), gamma-glutamyl transferase (GGT), alkaline phosphatase (AP), bilirubin, albumin, and prothrombin activity) to follow the liver function showed extensive variability already in non-cancerous patients, but generally remained within the range of their references (Supp. Fig. 1). As such, these markers presented limited potential for diagnosing NAFLD-HCC (Table 1). To investigate whether metabolomic changes in NAFLD-HCC are aetiology-specific, we compared the profiles of NAFLD-HCC patients and 32 patients with AV-HCC. As such, the metabolic profiles of patients with AV-HCC showed significant overlap, but a clear separation from the profiles of patients with NAFLD-HCC (Supp. Fig. 2h). Compared to patients with AV-HCC, NAFLD-HCC patients were older (p < 0.05) and less likely to develop HCC on a cirrhotic background (respectively 90% compared to 30%, Fisher’s exact test p < 0.0001). However, no significant differences were observed between NAFLD-HCC and AV-HCC in measurements of the liver function (AFP, AP, ALT, GGT, and

| Model | AUC |
|-------|-----|
| CTRL | 0.989 |
| OB-NAFLD | 0.997 |
| AV-HCC | 0.999 |

**Table 1** The area under the curve of the receiver operating characteristic curves for the model, biochemical parameters, and individual metabolites

Abbreviations: alpha-fetoprotein (AFP), alanine aminotransferase (ALT), gamma-glutamyl transferase (GGT), alkaline phosphatase (AP), bilirubin, albumin, and prothrombin activity.
bilirubin) or diabetes. Interestingly, tumours obtained from NAFLD-HCC patients were larger in size compared to tumours from patients with AV-HCC (Mann Whitney test, p=0.0006) and more frequently displayed microvascular invasion (41% NAFLD-HCC compared to 20% AV-HCC, Fisher’s exact test p=0.016).

3.2. NAFLD-HCC patients present a disparate serum metabolome

To establish a serum-based metabolomic landscape of NAFLD-HCC, we performed detailed metabolomics using a comprehensive library of 1295 metabolites covering amino acids (AA), glycerophospholipids, fatty acyls, sterols, sphingolipids, and glycerolipids. In total, we detected 470 metabolites. Due to the differences in age, sex and BMI between the groups, we performed covariate correction, including age, sex and BMI, and as a result 43 metabolites were excluded from further analysis (Fig. 1a). Next, sparse partial least squares discriminant analysis (sPLS-DA) revealed that NAFLD-HCC patients metabolically are the most distinct group compared to CTRL, OB-NAFLD, and patients with AV-HCC (Fig. 1b). Besides, NAFLD-HCC patients were the most dissimilar as seen by unsupervised hierarchical clustering (Fig. 1c). Interestingly, AV-HCC and NAFLD-HCC subgroups showed low similarity and clustered far apart (Fig. 1b, Supp. Fig. 2h), suggesting that unique metabolic programs may be driven by differences in the disease aetiology (NAFLD versus alcohol or viral hepatitis background).

3.3. Diagnostic potential of serum metabolomics

Serum metabolomics has been successfully used as a diagnostic tool to discriminate liver diseases [21,22]. Here, we investigated the potential of distinguishing NAFLD-HCC not only from healthy individuals and NAFLD patients, but also from AV-HCC. Thus, to generate a predictive metabolite signature, we first used receiver operating characteristic (ROC) curves and calculated area under the curve (AUC) for each metabolite as a contrast test between NAFLD-HCC and the respective comparative groups (CTRL, OB-NAFLD, and AV-HCC). As such, 89 metabolites presented an AUC > 0.75, distinguishing NAFLD-HCC patients from each of the other control groups (healthy

Figure 2. The diagnostic potential of lipidomics. a. The waterfall plot presenting the 14 metabolites signature distinguishing NAFLD-HCC patients. Metabolites are presented as the fold change to CTRL. b. ROC curves of 5-metabolite based model distinguishing NAFLD-HCC patients from other groups in comparison to biochemical tests. c. The predicted class probabilities (average of the cross-validation) for each sample using the best classifier (based on AUC).
or disease). Among the metabolites in the DEM signature, 14 metabolites presented a superior AUC of 0.9 in all contrast tests (Table 1). These metabolites and their fold change compared to CTRL are presented in Fig. 2a. Importantly, these 14 metabolites individually presented AUCs markedly superior to alpha-fetoprotein (AFP: AUC_{NAFLD-HCC} vs CTRL=0.791 and AUC_{NAFLD-HCC} vs AUC_{NHDS}=0.614) and all biochemical measurements for the liver function (ALT: AUC_{NAFLD-HCC} vs CTRL=0.776, AUC_{NAFLD-HCC} vs AUC_{NHDS}=0.733, and AUC_{NAFLD-HCC} vs AUC_{AV-HCC}=0.570 or GGT: AUC_{NAFLD-HCC} vs CTRL=0.927, AUC_{NAFLD-HCC} vs AUC_{NHDS}=0.865, and AUC_{NAFLD-HCC} vs AUC_{AV-HCC}=0.562), discriminating NAFLD-HCCs from all controls and AV-HCC (HCC of different aetiology other than NAFLD).

Next, we assessed if a combination of the 14 metabolites would increase the diagnostic potential. As such, we used support vector machine (SVM) modelling and determined that a panel of 5 metabolites yielded the optimal predictive accuracy (Supp. Fig. 3a). Indeed, the model based on the 5 metabolites reached an AUC of 0.98 (Fig. 2b) for any of the contrasts (compared to all controls and AV-HCC) with a predictive accuracy greater than 90% (Fig. 2c). Importantly, the accuracy of the diagnostic panel was confirmed in the validation set and a model performance with an AUC=0.91, including matching NAFLD-HCC patients according to BMI (Supp. Fig. 3b).

3.4. Validation and quantification of diagnostic metabolites

To reinforce the clinical relevance of the metabolite panel in the diagnosis of NAFLD-HCC patients, we next validated and quantified the abundance of each metabolite. Also, we established their reference concentration range. Among the 14 metabolites with the greatest diagnostic value, only 10 of them have commercial standards available. Also, each metabolite needs to be detected within the linear range, which allows for absolute quantification. As such, we established the abundance of metabolites in the validation set (Fig. 3a). All metabolites except PC (0/0.22:5) showed a similar trend as in the discovery set. Overall, the validation set showed a great variability, which can be due to difference in geographical sampling or availability of plasma instead of serum. Moreover, we utilised the quantification of each metabolite to generate linear regression models to assess their concentration levels in the discovery set. The linear dependence and estimated concentrations are presented in Supp. Table 2. Lastly, among the validated metabolites in the panel, which are based in SVM modelling and the mean importance measure, we determined the optimal set of metabolites (linoleic acid, osbond acid, monounsaturated fatty acid MUFA (14:1n-Strans), and PC (18:2/0:0)) to build the NAFLD-HCC Diagnostic Score (NHDS). The NHDS is dependent on the absolute concentration level of each metabolite in the score. The NHDS performed with an AUC=0.75 for any given contrast (Fig. 3b-c) as well as presented suitable power according to sample size (Supp. Table 3). Furthermore, we established the odds ratio (OR) and relative risk (RR) for the NHDS (at cut-off value 0) in the discovery, validation, and combined cohort. The NHDS=0 was predictive of NAFLD-HCC in the discovery set (Fishier’s exact test, p=0.01) with relative risk RR=1.18, odds ratio OR=8.87 and in validation set (Fishier’s exact test, p=0.0001) with RR=2.23 and infinite OR. Taken together, the NHDS=0 was predictive of NAFLD-HCC with the highest odds ratio OR=97.55 (95% CI=16.40 to 999.7) (Fig. 3d) and relative risk RR=1.67 (95% CI=1.45 to 1.99) (Fig. 3e). Moreover, we next tested if underlying conditions, such as diabetes, dyslipidaemias, cirrhosis effect the performance of NHDS (Supp. Table 4). Interestingly, no DEMs were detected between cirrhotic and non-cirrhotic NAFLD-HCC patients in direct pair-wise comparison. Next, we considered if the use of medication (statin and paracetamol) and the presence of known NAFLD risk-genotypes (Patatin-like phospholipase domain-containing protein 3 (PNPLA3), Transmembrane 6 Superfamily Member 2 (TM6SF2), Membrane Bound O-Acyltransferase Domain Containing 7 (MBOAT7)) affected the performance of NHDS. Indeed, the model showed significant performance (AUC= 0.888-1, p<0.0001) regardless of the underlying condition, medication used and genotypes (Supp. Table 4). Also, the NHDS predicted NAFLD-HCC independent of tumour stage (BCLC and TNM), degree of NAFLD and sex (Supp. Table 4).

3.5. Unique metabolomic profile of NAFLD-HCC patients

To investigate differences in the metabolic profiles between NAFLD-HCC patients and the other groups, we performed a series of pair-wise tests. As such, after the covariate correction (age, sex, BMI) orthogonal partial least squares discriminant analysis (oPLS-DA) showed substantial separation of NAFLD-HCC patients from CTRL (R²X=0.178, R²Y=0.751, Q²=0.723) (Fig. 4a). The differential expression analysis identified 274 significantly different (t-test, false-discovery rate (FDR) corrected p<0.05) metabolites (DEMs), among which 152 metabolites were depleted and 122 metabolites were increased (Fig. 4b, Supp. Table 5). Next, we performed pathway over-representation analysis of the depleted metabolites using integrated molecular pathway level analysis (IMPala) [28] and identified linoleic acid metabolism and G protein-coupled receptor (GPCR) signalling as the most impaired networks (Supp. Table 6). Contrary, cholesterol synthesis, membrane fluidity and trafficking, and glycerophospholipid metabolism were among the most upregulated pathways (Supp. Table 6). In addition to relate individual metabolites to their processes, we compared unique classes of metabolites with the same chemical characteristics. As such, we defined a unique depletion of acylcarnities (AC), sterol lipids (ST), and fatty acids (FA); especially oxidized fatty acids (oxFA) and omega-6 FA in NAFLD-HCC, while saturated triglycerides (TG) were upregulated (Fig. 4c, Supp. Fig. 4). Furthermore, we utilised Bioinformatics Methodology For Pathway Analysis (BioPAN) [29] for lipid pathway enrichment analysis. We observed a significant activation of reactions converting sphingomyelins (SM) to ceramides (Cer), a process catalysed by sphingomyelin phosphodiesterases (SMPD2 and SMPD3), as well as phosphatidylcholines (PC) to diglycerides (DG), which is catalysed by the sphingomyelin synthases (SGMS1 and SGMS2) (Supp. Fig. 5a). Similarly, significant alterations were revealed in the activity of FA desaturases and elongases with specific activation of fatty acid desaturase 1 (FADS1) and impairment in FADS2, stearyol-CoA desaturase 1 (SCD1) and elongation of very long chain fatty acid (ELOVL) elongases (ELOVL2, and ELOVL5) (Supp. Fig. 6a).

Next, we compared OB-NAFLD and NAFLD-HCC patients. Indeed, oPLS-DA analysis showed moderate separation (R²X=0.173, R²Y=0.715, Q²=0.697) (Fig. 4d). The differential expression analysis identified 316 DEMs with 154 metabolites upregulated and 162 downregulated (Fig. 4e, Supp. Table 7). The overrepresentation analysis identified cholesterol and bile acid (BA) metabolism, aminoacyl-tRNA biosynthesis, as well as protein and glucose metabolism as significantly upregulated in NAFLD-HCC compared to OB-NAFLD. Contrary, FA biosynthesis and GPCR signalling were both downregulated (Supp. Table 8). Furthermore, OB-NAFLD significantly differed from NAFLD-HCC in specific classes of metabolites. As such, AA, BA, Cer, TG, PC, and phosphatidylethanolamines (PE) were all significantly upregulated in NAFLD-HCC compared to OB-NAFLD. Contrary, FA biosynthesis and GPCR signalling were both downregulated (Supp. Table 8). Furthermore, OB-NAFLD significantly differed from NAFLD-HCC in specific classes of metabolites. As such, AA, BA, Cer, TG, PC, and phosphatidylethanolamines (PE) were all significantly upregulated in NAFLD-HCC compared to OB-NAFLD. Contrary, FA biosynthesis and GPCR signalling were both downregulated (Supp. Table 8). Moreover, ELOVL3 and ELOVL6 presented an increased activity, whereas ELOVL2, SCD1, and SCD3 were suppressed (Supp. Fig. 6b). Contrary, reactions catalysed by FADS1, FADS2, and ELOVL5 displayed a mixed activity.

Lastly, we compared the metabolomes of NAFLD-HCC to HCCs with alcohol and/or viral aetiology (AV-HCC). The oPLS-DA model showed moderate separation (R²X=0.171, R²Y=0.739, Q²=0.694) (Fig. 4f). As expected, the AV-HCC group showed greater
Figure 3. Validation of metabolites and NAFLD-HCC Diagnostic Score (NHDS).

(a) The relative abundances of 10 validated metabolites in discovery (blue) and validation (purple) sets are expressed as a fold change against CTRL. Metabolites included in NHDS are marked in red.

(b) ROC curves of NHDS in the discovery set.

(c) ROC curves of NHDS in the NAFLD-HCC validation set.

(d) The forest plot presenting odds ratios (Fisher’s exact test) of NHDS at the established cut-off value as well as clinical and biochemical characteristics for discovery, validation, and pulled (discovery and validation) sets. AFP and GGT were available only in the discovery set. Error bars present 95% CI, bold font was used for statistically significant (Fisher’s exact test, p<0.05) factors.

(e) The forest plot representing the relative risk (Fisher’s exact test) of NHDS at the established cut-off value as well as clinical and biochemical characteristics for discovery, validation and pulled (discovery and validation) sets. AFP and GGT were available only in the discovery set. Error bars present 95% CI, bold font was used for statistically significant (Fisher’s exact test, p<0.05) factors.
Figure 4. Pair-wise comparison between NAFLD-HCC and other groups. 

a. The orthogonal partial least squares discriminant analysis (oPLS-DA) showing separation between NAFLD-HCC and CTRL. 
b. Volcano plot representing DEMs between NAFLD-HCC and CTRL samples. The FDR p=0.05 is marked with a red dashed line. 
c. Differences between NAFLD-HCC and other groups across major metabolic classes. Statistical significance between NAFLD-HCC and other groups was established with one-way ANOVA or Kruskal-Wallis tests *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. 
d. The oPLA-DA model showing separation between NAFLD-HCC and OB-NAFLD. 
e. Volcano plot representing DEMs between NAFLD-HCC and OB-NAFLD. 
f. The oPLS-DA model showing separation between NAFLD-HCC and AV-HCC. 
g. Volcano plot representing DEMs between NAFLD-HCC and AV-HCC. 
h. Simplified relationship between major metabolic classes. Classes marked in blue are downregulated and red upregulated in NAFLD-HCC patients. DEMs were identified by t-test followed by FDR correction. 

Abbreviations: acylcarnitines (AC), amino acids (AA), aromatic amino acids (ArAAs), bile acids (BA), branched-chain amino acids (BCAAs), ceramides (Cer), cholesteryl esters (ChoE), diacylglycerols (DG), diacylglycerolphosphoethanolamine (DAPE), fatty acids (FA), fatty acids omega 3 (FA omega 3), fatty acids omega 6 (FA omega 6), fatty acids omega 9 (FA omega 9), free bile acids (FBA), free sphingoid bases (FSB), glycine-conjugated bile acids (GCBA), lysophosphatidylcholines (LPC).
heterogeneity compared to NAFLD-HCC, which partly can be due to the mixed etiology. Differential expression analysis identified 257 DEMs (125 downregulated and 132 upregulated) in NAFLD-HCC compared to AV-HCC (Fig. 4g, Supp. Table 9), showing an overall impairment in the free FA (FFA) biosynthesis. Contrary, insulin resistance, glycophospholipid, and choline metabolism were upregulated in NAFLD-HCC (Supp. Table 10). Next, we compared metabolic classes between NAFLD-HCC and AV-HCC patients and observed significant depletion of AC, BA, FA, ST, lysophosphatidylethanolamines (LPE), and phosphatidylinositols (PI) in NAFLD-HCC. Contrary, cholesteryl esters (ChoE), SM, Cer, and TG were all increased in NAFLD-HCC compared to AV-HCC (Fig. 2c, Supp. Fig. 4). Moreover, lipid pathway enrichment analysis showed an increased reaction activity in the path SM→Cer (Supp. Fig. 5c) with the enzyme activities of FADS2, ELOVL5, and ELOVL2 significantly reduced (Supp. Fig. 6c).

Taken together, the serum of NAFLD-HCC patients is characterized by a significant depletion of FA reflective of a significantly lower FA biosynthesis with decreased FA desaturase and elongase activities. A depletion in both AC and ST with concurrent higher TG and Cer abundance is suggestive of a unique metabolic reprogramming in NAFLD-HCC patients. The altered SM:Cer ratio could be the result of an increased activity in the enzymes SMPD2 and SMPD3 or reduced activity of SGMS1, SGMS2, CERT1 in NAFLD-HCC patients. A simplified association between the lipid classes and their deregulation in NAFLD-HCC is presented in Fig. 4h.

3.6. NAFLD-HCC tumours are responsible for fatty acid deterioration in circulation

We observed significant depletion of omega-6 unsaturated fatty acids in the serum of NAFLD-HCC patients. Interestingly, essential linoleic acid (LA) as well as its metabolite osbond acid showed significant difference and diagnostic potential. As such, we investigated the concentration of LA and its elongation and desaturation pathway (Fig. 5a). We first pooled the NAFLD discovery and validation patient sets and based on the liver histology, we stratified them into NAFL (NAS<6, n=58) and NASH subsets (n=72). Indeed, we observed significant depletion of LA, arachidonic acid, ardenic acid and osbond acid in the serum from NAFLD-HCC patients (Fig. 5b). Interestingly, these FAs remained unchanged in serum-matched tumours compared to the normal tumour-adjacent tissues, healthy CTRLs as well as NAFL and NASH patients (Fig. 5c). As such, we hypothesized that tumor tissue might be responsible for the metabolic sink of FFA. To that end, we measured the expression of FA transporters, FA binding proteins, G-protein coupled receptor for FA, FA elongases, FA desaturases and genes involved in oxylipin and prostaglandin synthesis (Fig. 5d). NAFLD-HCC tumours showed a significant increase in the expression of FA transporters (CD36, SLC27A1, SLC27A3, SLC27A4), FA binding proteins (FABP2, FABP4, FABP5, FABP6), FA elongases (ELOVL1, ELOVL2, ELOVL5, ELOVL6), FA desaturases (FADS1-3), Similarly, LA lipoxynegenases (ALOX15 family) were upregulated in tumours while cytochrome P450 enzymes (CYPs) with LA affinity were downregulated. Furthermore, we validated overexpression of CD36 and FA elongases in NAFLD-HCC samples in The Cancer Genome Atlas (Fig. 5e). Collectively, these data suggest that NAFLD-HCC tumours significantly depend on exogenous FAs, increasing their uptake and metabolism.

3.7. The serum lipidome landscape reflects the progression to NAFLD-HCC

We investigated whether the serum lipidome reflects the progressive nature of NAFLD-HCC development. It has been previously reported, that major lipidomic alterations are observed in the development of steatosis, but not progression to NASH [16]. We performed pair-wise comparisons between NAFL and NASH adjusting for age, sex and BMI and found no DEMs, which is in accordance with recent findings [16]. As such, we investigated the CTRL→NAFL→NAFLD-HCC axis employing a pattern hunter approach with Pearson correlation to designate the metabolic rearrangements in the subgroups as part of the developing disease (Fig. 5f). A total of 412 lipids were detected in the pooled discovery and validation NAFLD samples.

As such, we found a total of 121 lipids progressively altered (Pearson’s correlation FDR p<0.05, r>0.3; 66 negatively, and 55 positively) following this axis (Supp. Table 11). The negatively correlated metabolites include AC, ChoE, PUFA, LPC, and ST subclasses. Among the metabolites positively correlated, we found an overrepresentation of TG (42 out of 55 metabolites). Importantly, 18 TGs significantly correlate with increasing NAS and fibrosis scores in OB-NAFLD patients, suggesting their importance in the progressive deterioration of the liver (Fig. 5g). Lastly, we tested whether any of these metabolites correlated with tumour size, recurrence, microvascular invasion, or liver cirrhosis in NAFLD-HCC patients, however, none of the metabolites reached statistical significance. This suggests that these metabolites are associated with NAFLD-HCC, but not directly involved the cancer progression.

4. Discussion

A sedentary lifestyle and overnutrition have led to an epidemic of obesity, diabetes, and NAFLD that soon may become the leading causes of HCC development. Metabolic reprogramming is at the core of NAFLD progression to HCC [21]. However, due to the need of invasive techniques for NAFLD diagnosis, our understanding of NAFLD-HCC is limited and the underlying metabolomic landscape remains elusive. The present study provides the first comprehensive analysis covering 22 classes of lipids and amino acids and identifying complex relationships in the gradual progression of NAFLD versus NAFLD-HCC patient subsets. Importantly, this study was performed in patients with biopsy-proven liver histology before therapy. To our knowledge, it is the first study aiming to non-invasively differentially diagnose NAFLD-HCC from HCC of other aetiologies. The current study is focused on early-stage, resected NAFLD-HCC patients, which is important for detection of metabolic changes at early-stage of the disease and for its application in screening of patients at-risk. Our findings warrant investigation in patients with advanced disease.

The diagnosis of NAFLD-HCC is challenging as patients with metabolic syndrome are not screened for HCC [7,71]. Indeed, often only patients with chronic liver disease of viral aetiology are offered HCC screening. Increasing levels of liver enzymes are suggestive of liver damage but are not specific to hepatocarcinogenesis. Additionally, the diagnostic capacity of AFP [30] for HCC is limited and was observed to be particularly inadequate for NAFLD-HCC patients (Fig. 3d, e, Supp. Fig. 1). As such, we have applied machine learning approaches to pin-point a combination of metabolites that offers the best diagnostic potential for NAFLD-HCC patients. We found that a combination of 5 metabolites accurately distinguishes NAFLD-HCC patients from healthy individuals (AUC=0.989), OB-NAFLD patients

[1] M. Lewinska et al. / EBioMedicine 73 (2021) 103661
Figure 5. The metabolism of linoleic acid in serum and tissue and the progressive metabolic perturbations in CTRL to NAFLD to NAFLD-HCC trajectory. 

a. The schematic representation of LA metabolism (Fatty Acid Desaturase 1 (FADS1), Fatty Acid Desaturase 2 (FADS2), Elongation of Very Long chain fatty acids (ELOVLs)).

b. Violin plots representing the abundance of LA metabolites in serum. Expression was tested with one-way ANOVA or Kruskal-Wallice test *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.

c. Violin plots representing the abundance of LA metabolites in tissue (NAFLD-HCC surrounding liver (SL, n=27); NAFLD-HCC tumour tissue (T, n=27); CTRL n=9; NAFL n=32; NASH n=20).

d. The heatmap representing gene expression of fatty acid transport and metabolism genes in NAFLD-HCC surrounding liver tissue and NAFLD-HCC tumours. Genes marked in red are significantly (t-test, p < 0.05) upregulated and blue downregulated in tumour compared to matched surrounding liver tissue.

e. The mRNA expression of CD36, ELOVL1 and ELOVL2 in NAFLD-HCC tumours in TCGA (n=12) compared to SL (n=25).

f. The heatmap presenting the expression of metabolites statistically significantly correlated with disease progression.

The bubble plot presenting Pearson’s correlation of metabolites with NAS and fibrosis level in NAFLD patients.

M. Lewinska et al. / EBioMedicine 73 (2021) 103661
metabolism of LA and arachidonic acid have been implicated in viral-

sucrantly decreased levels of serum acylcarnitines suggests an increased organ for carnitine synthesis and metabolism [40]. As such, the signif-

Interestingly, the LA abundance in tissues is restored (Fig. 5c) likely emphasizing the potential limitation of murine studies in this model not fully mimic the human disease, and possibly further NAFLD-HCC patients (Fig. 2a, Fig. 5b), suggesting that the murine model increases oxidative stress and causes a selective loss of intrahepatic CD4+ T cells, leading to NAFLD-mediated hepatocarcinogenesis [38]. Interestingly, serum LA is significantly depleted in NAFLD-HCC patients (Fig. 2a, Fig. 5b), suggesting that the murine model not fully mimic the human disease, and possibly further emphasizing the potential limitation of murine studies in this field. Interestingly, the LA abundance in tissues is restored (Fig. 5c) likely due to increased uptake by the tumour (Fig. 5d). Indeed, increased metabolism of LA and arachidonic acid have been implicated in viral-associated HCC in the Korean population [39]. The liver is the key organ for carnitine synthesis and metabolism [40]. As such, the significantly decreased levels of serum acylcarnitines suggests an increased β-oxidation of FFA [41], which could be caused by increased energy demands of NAFLD-HCC tumours. Furthermore, we found elevated levels of bile acids in NAFLD-HCC patients compared to OB-NAFLD patients, but lower than AV-HCC (Fig. 4c). That can be result of altered BA metabolism in cirrhotic patients owning to altered gut micro-

The unsaturated fatty acids (mono- (MUFA) and polyunsaturated (PUFA)) significantly differed between NAFLD-HCC and other groups (Fig. 4) with lower levels in NAFLD-HCC. Indeed, cancer cells show higher dependency on unsaturated fatty acids to proliferate [33], and lipid-dependence was previously reported in HCC [34]. Thus, we can speculate that NAFLD-HCC tumours act as sinks for unsaturated fatty acids from the blood. Indeed, the increased transport of FA by CD36 has been observed in widely investigated in the context of NAFLD and NASH [35]. As such, its further augmentation in NAFLD-HCC tumours might be responsible for reduced serum FA making it potential therapeutic target. The essential omega-6 FA (linoleic acid/LA) is a precursor to long-chain metabolites, including arachidonic acid, which is a substrate in the prostaglandin synthesis and thus, deple-

tion of LA may cause an altered signalling and inflammatory response [36]. The role of LA in carcinogenesis remains controversial. On the one hand, LA and its derivatives have shown a tumour-suppressive role in colorectal cancer [37]. Conversely, LA accumulation (in a murine model) increases oxidative stress and causes a selective loss of intrahepatic CD4+ T cells, leading to NAFLD-mediated hepatocarci-

nogenesis [38]. Interestingly, serum LA is significantly depleted in NAFLD-HCC patients (Fig. 2a, Fig. 5b), suggesting that the murine model not fully mimic the human disease, and possibly further emphasizing the potential limitation of murine studies in this field. Interestingly, the LA abundance in tissues is restored (Fig. 5c) likely due to increased uptake by the tumour (Fig. 5d). Indeed, increased metabolism of LA and arachidonic acid have been implicated in viral-associated HCC in the Korean population [39]. The liver is the key organ for carnitine synthesis and metabolism [40]. As such, the significantly decreased levels of serum acylcarnitines suggests an increased β-oxidation of FFA [41], which could be caused by increased energy demands of NAFLD-HCC tumours. Furthermore, we found elevated levels of bile acids in NAFLD-HCC patients compared to OB-NAFLD patients, but lower than AV-HCC (Fig. 4c). That can be result of altered BA metabolism in cirrhotic patients owning to altered gut micro-

biome in cirrhosis [42].

HCC is a progressive disease developing over decades. We observed a gradual attenuation of acylcarnitines, lysophosphatidyl-

cholines, and unsaturated FA and an increase in phosphatidylcholines and triglycerides during the progression of liver disease. Interest-

ingly, some of the triglycerides that were found to be augmented in NAFLD versus NAFLD-HCC were previously associated with NAS and fibrosis scores in NASH [43,44] (Fig. 5g), but were not associated with tumour size. Although, we have presented a successful metric, demon-

strating the ability of the metabolite panel in diagnosing NAFLD-

HCC patients and in distinguishing these patients from AV-HCC, results must be interpreted with caution given the sample size. The NHDS needs to be tested in a large cohort, including advanced nonre-

sectable and ethnically diverse HCC patients in the prospective longi-

tudinal cohorts. However, stratification of HCC based on their metabolome may in the future be an approach to evaluate crypto-

genic HCC patients. It is notable that NAFLD-HCC patients were significantly older compared to other groups which limited the power to detect metabolic differences.

In conclusion, the depletion of unsaturated FA, and the increase of triglycerides are at the core of deregulated metabolic networks in NAFLD-HCC, leading to altered signalling and likely different nutrient utilisation by cancer cells. These changes can be exploited for non-

invasive surveillance of the ‘at risk’ population for early HCC detection in the background of metabolic syndrome.

Declaration of Competing Interest

Drs. Alonso and Arretxe are employed by OWL Metabolomics (One Way Liver, S.L.). Prof. Banales is a member of the scientific advisory board of OWL Metabolomics. Remaining authors declare no conflict of interest.

Contributors

Conceptualization: ML, JBA. Data curation: ML, ASI, EZ, RJA, EE, MJP, MPS, FKK, MRC, MAJ, SKO, GEV, TD, CPO, BS, RIR, JMB, JBA. For-

mal Analysis: ML, EA, EZ, CA. Writing – original draft: ML, JBA. All authors read and approved the manuscript. ML, JBA have verified the underlying data.

Acknowledgment

We thank the participating patients and their families as well as clinical staff at all contributing centres. We thank the French Liver Biobanks network – INCA (BB-0033-00085) for the contribution of NAFLD-HCC samples (participating hospitals: Beaujon, Bordeaux, Grenoble, Henri-Mondor, Paris-SUD, and Rennes). We thank the section of Biostatistics at Department of Public Health, University of Copenhagen for statistical advice. The results are in part based upon data generated by the TCGA Research Network: https://www.cancer. gov/tcga. We thank all funding sources: The laboratory of JBA is supported by the Novo Nordisk Foundation (14040, 0058419), Danish Cancer Society (R98-A6446, R167-A10784, R278-A16638), and the Danish Medical Research Council (4183-01118A, 1030-00070B). Data used for validation in this study provided by JMB was funded by the Spanish Ministry of Economy and Competitiveness and ‘Institut de Salud Carlos III’ grants (PI18/01075, Miguel Servet Programme CON14/00129 and CPI19/00008) co-financed by ‘Fondo Europeo de Desarrollo Regional’ (FEDER); CIBERehd, Spain; IKERBASQUE, Basque foundation for Science, Spain; BIOEF (Basque Foundation for Innovation and Health Research: EITB Maratobia BIO15/CA/016/BD); Department of Health of the Basque Country (2017111010), Euskadi RIS3 (2019222054, 2020333010); Department of Industry of the Basque Country (Elkartek: KK-2020/00008), AECC Scientific Foundation and European Commission Horizon 2020 program (ESCALON project no.: 825510). Similarly, MAJ was funded by grants from the Fondo Nacio-

nal de Ciencia y Tecnología de Chile (FONDECYT #1191145 to M.A.) and the Comisión Nacional de Investigación, Ciencia y Tecnología (CONICYT, AFB170005, CARE Chile UC).

Data sharing

The data used in this study will according to IRBs be made available upon request if the requestor has an approved protocol.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2021.103661.
