Plasma lipidomics distinguishes NASH and fibrosis from simple fatty liver in children

Running head: Lipidomics in paediatric NAFLD

Jake P. Mann[1], Samuel Furse[2], Stuart G. Snowden[2], Anna Alisi[3], Laura G. Draijer[4], EU-PNAFLD investigators*, Benjamin Jenkins[2], Kylie Karnebeek[5], Deirdre A. Kelly[6], Bart G. Koot[4], Antonella Mosca[7], Camilla Salvestrini[8], Indra van Mourik[6], Anita Vreugdenhil[5], Matthias Zilbauer[8,9], Albert Koulman[2]

Affiliations

1. Institute of Metabolic Science, University of Cambridge, Cambridge, UK
2. Core Metabolomics and Lipidomics Laboratory, Wellcome Trust-MRC Institute of Metabolic Science, University of Cambridge, L4 Addenbrooke's Treatment Centre, Keith Day Road, Cambridge, CB2 0QQ, UK.
3. Research Unit of Molecular Genetics of Complex Phenotypes, Bambino Gesù Children's Hospital - IRCCS, Rome, Italy.
4. Department of Paediatric Gastroenterology and Nutrition, Amsterdam University Medical Centers, Academic Medical Center/Emma Children's Hospital, Meibergdreef 9, 1100 AZ, Amsterdam, Netherlands
5. Centre for Overweight Adolescent and Children's Healthcare (COACH), School of Nutrition and Translational Research in Metabolism (NUTRIM), Maastricht University Medical Centre, The Netherlands
6. Liver Unit, Birmingham Children's Hospital, Birmingham, UK
7. Hepatology Gastroenterology and Nutrition, Bambino Gesù Children's Hospital, Rome, Italy
8. Department of Paediatric Gastroenterology and Nutrition, Addenbrooke's Hospital, Cambridge, UK
9. Department of Paediatrics, University of Cambridge, UK

*A list of investigators is included in the Acknowledgements section

Correspondence:
Dr. Jake P Mann
Metabolic Research Laboratories
Level 4, Institute of Metabolic Science
Addenbrooke's Hospital
Cambridge, CB2 0QQ
jm2032@cam.ac.uk
Tel: +44 1223 336792

Fax: +44 1223 330598

Financial Support

Authors' declaration of personal interests:

The authors have no conflicts of interest to declare.

Declaration of funding interests:

JPM is supported by a Wellcome Trust fellowship (216329/Z/19/Z), a European Society for Paediatric Research (ESPR) Young Investigator Award, and a Children's Liver Disease Foundation Grant. EU-PNAFLD Registry is supported by an European Association for Study of the Liver (EASL) Registry Grant. MZ is supported by a New Investigator Research Grant from the MRC (MR/T001917/1). BK is supported by grants from Van den Broek Lohman Foundation, Virtutis Opus Foundation and For Wishdom Foundation. SF, SGS & AK are supported by the BBSRC (BB/M027252/1 & BB/P028195/1), BJJ & AK are supported by the National Institute for Health Research (NIHR146281).

Acknowledgements

The authors are grateful to the young people and their families for taking part in the EU-PNAFLD Registry. The other EU-PNAFLD investigators are : Piotr Socha, Wojciech Jańczyk, Ulrich Baumann, Donatella Comparcola, Sanjay Rajwal, Florence Lacaille, Myriam Dabbas, Quentin M. Anstee and the late Valerio Nobili.

Guarantor of the article: Jake P. Mann

Specific author contributions: Concept and design - JPM & AK; Recruitment & data collection - JPM, AA, LGD, KK, DAK, BGK, AM, CS, IvM, AV, MZ; Analysis & data processing - SF, SGS, BJ, AK; Drafting manuscript - JPM, SF, SGS, AK. All authors have approved the final version of the manuscript.

Summary

Background: Non-alcoholic fatty liver disease (NAFLD) is a condition characterised by insulin resistance and altered lipid metabolism. Evaluation of the serum lipid profile of children with NAFLD offers the opportunity to understand these perturbations and identify biomarkers to distinguish non-alcoholic steatohepatitis (NASH) and fibrosis from 'simple' fatty liver (NAFL).

Aim: To investigate the plasma lipid profile of children with NAFLD to: (a) gain insight into lipid metabolism in NAFL and NASH; (b) explore the potential of their use as biomarkers.

Methods: Untargeted lipidomics was performed on fasting plasma samples were collected from children: 19 lean control, 65 obese controls (without NAFLD), and 222 children with NAFLD, of whom 120 had undergone liver biopsy to identify 34 with NAFL and 86 with NASH. Associations between lipid species and liver histology were assessed using regression adjusted for age and sex.

Results: Histological severity of paediatric NAFLD was associated with altered in lipid metabolism in multiple classes, including: higher free cholesterol, higher phosphatidylserine, higher *lyso*-phosphatidylcholines, and lower polyunsaturated di- and triglycerides. Many of these associations remained after adjusting for insulin resistance and were not replicated in obese controls. The plasma lipid profile generated a diagnostic model that accurately predicted NAFL versus NASH or fibrosis on liver biopsy.

Conclusions: Histological severity of NAFLD in children is associated with perturbation of multiple lipid pathways, several apparently independent of insulin resistance. These data provide proof-of-concept that plasma lipidomics can distinguish simple fatty liver from NASH or fibrosis in children.

Keywords

Steatohepatitis, biomarker, metabolomics, portal inflammation, liver biopsy.

1. Introduction

Non-alcoholic fatty liver disease (NAFLD) is a common, chronic disorder that is closely linked to obesity and insulin resistance. The majority of morbidity and mortality in patients with NAFLD occurs due to complications of cardiovascular disease though a small, but significant, proportion develop cirrhosis[1]. Individuals with higher stages of liver fibrosis[2], or more active inflammation (non-alcoholic steatohepatitis (NASH))[3] are more likely to progress to end-stage liver disease.

NAFLD is also associated with the metabolic syndrome in children[4] and, whilst the long-term outcomes have not yet been formally established with the same degree of confidence as in adults[5,6], they are believed to be similar. However paediatric NAFLD has several unique features, in particular that peri-portal inflammation is a more prominent feature, particularly in younger children, boys, and in those of Hispanic ethnicity[7,8]. It is not yet understood what causes increased peri-portal inflammation, though it does appear to be associated with degree of insulin resistance[9].

Lipidomics, a technique that measures the relative concentration (or abundance) of multiple lipid species, has been used by several groups to gain insight into altered lipid metabolism in NAFLD. Liver samples[10–12], venous-[13,14] and portal- blood[15] have been studied in humans, all showing specific lipid species to associate with histological severity of NAFLD in adults. This work has identified perturbation of pathways related to hepatic *de novo* lipogenesis, desaturase activity, and phospholipid metabolism. To date, lipidomic studies in children have focused on species that differentiate NAFLD patients from healthy controls [16,17].

In this study we used plasma lipidomics to investigate lipid metabolism in children with NAFLD. Specifically, we aimed to identify trends in lipid species that were associated with the

histological severity of NASH, separate to those affected by obesity in the absence of NAFLD. We hypothesised that peri-portal inflammation would have a unique pattern of lipid metabolism, distinct from that of ‘adult’ NASH.

2. Methods

2.1 Participants

Three groups of participants were included in this cross-sectional study: cases (children with NAFLD), obese controls (BMI >2 SD without NAFLD), and lean controls were included in the study. All participants (or their parents) gave written informed consent and were recruited between 2014-2019.

Children (<18 years) with NAFLD were recruited from the European Paediatric NAFLD Registry (EU-PNAFLD, Clintrials.gov NCT:04190849)[18], which was approved by the East Midlands - Nottingham 2 Research Ethics Committee (17/EM/0084), were recruited from five European specialist centres for paediatric gastroenterology and/or hepatology.

Obese controls were recruited from paediatric obesity clinics at Maastricht Children's Hospital (under ethical approval METC 13-4-130) and Amsterdam Medical Centre (under ethical approval MEC 2017_306 and MEC 07/141). Children were referred to these clinics due to obesity and were then subsequently investigated for co-morbidities (and secondary causes of obesity), including NAFLD. Children with obesity but no evidence of liver disease (as described below) were included in this study as obese controls.

Recruitment from multiple different clinics aimed to reduce the bias associated with a sample from a single hospital population.

Lean controls were recruited as part of the Translational Research in Intestinal Physiology and Pathology (TRIPP) Study at Cambridge University (UK), which was approved by East of England - Cambridge South Research Ethics Committee (REC 17/EE/0265). These children underwent endoscopy to rule out gastrointestinal disease and were found to have no evidence of pathology and complete resolution of any symptoms.

As an exploratory analysis, we utilised data from the maximum number of available participants therefore no formal sample size calculation was performed.

2.2 Inclusion and exclusion criteria

All participants were 5-18 years old at the time of inclusion.

Cases: NAFLD was diagnosed by liver histology, or by radiological (ultrasound or magnetic resonance imaging) identification of steatosis with exclusion of secondary causes using bloods: normal range caeruloplasmin, ferritin, alpha-1-antitrypsin; normal immunoglobulins, negative antinuclear antibody, and negative anti-smooth muscle antibody; and negative screening for chronic viral hepatitis.

Controls: NAFLD was excluded by absence of hepatic steatosis on ultrasound or MRI, plus aminotransferases below age-specific diagnostic cut-offs[19]. Controls with an age/sex-specific BMI z-score >2 were included as 'obese controls'.

2.3 Clinical and laboratory investigations

NAFLD cases and obese controls underwent baseline characterisation with clinical assessment and anthropometric measurements including waist/hip circumferences and blood pressure. . Fasting blood was obtained and plasma was frozen at -80 degree Celsius for use in all analyses. The homeostatic model assessment of insulin resistance (HOMA-IR) was derived using fasting insulin (μU/L) x fasting glucose (nmol/L) / 22.5.

Participants who underwent liver biopsy were tested for single nucleotide variants implicated in the pathogenesis in NAFLD, which are also known to perturb plasma lipid profiles. DNA was extracted from whole blood (using Qiagen DNeasy kit #69504). Genotypes of three common variants was tested by quantitative polymerase chain reaction using TaqMan assays (ThermoFisher #4351379): rs738409C>G in *PNPLA3* , rs58542926C>T in *TM6SF2*, and rs641738C>T near *MBOAT7*.

Lean controls had height and weight measured and fasting blood obtained but no further samples or investigations were performed for the purpose of this study.

2.4 Liver biopsies

A subset of children with NAFLD underwent liver biopsy. Liver biopsy was not performed in control participants. Liver biopsy was performed in accordance with ESPGHAN guidelines[20], particularly in cases of diagnostic uncertainty. Biopsies were obtained percutaneously and considered adequate if there was minimal fragmentation and at least 20 mm long. Liver biopsy was scored according to NASH CRN criteria[8,21], including assessment of portal inflammation as none (0), mild (1), or more than mild (2), and calculation of the NAFLD Activity Score (NAS). Non-alcoholic steatohepatitis (NASH) was diagnosed according to the presence of both ballooning and lobular inflammation (the Fatty Liver Inhibition of Progression (FLIP) algorithm)[22]. Samples from liver biopsy were used for histological assessment only; no tissue was available for lipidomic (or transcriptomic) analysis.

2.5 Plasma lipidomics analysis

Reagents and standards

Solvents were purchased from *Sigma-Aldrich Ltd* (Gillingham, Dorset, UK) of at least HPLC grade and were not purified further. Lipid standards were purchased from Avanti Polar lipids (Alabaster, AL; through Instruchemie, Delfzijl, NL) and used without purification. Consumables were purchased from Sarstedt AG & Co (Leicester, UK).

Mass Spectrometry

Isolation—Lipids, triglycerides and sterols were extracted together from samples placed in random order using a high throughput technique described recently[23]. Briefly, aliquots of serum (20 µL) were placed along with blank and quality control (QC) samples in the wells of a glass-coated 2.4 mL/well 96w plate (Plate+™, Esslab, Hadleigh, UK). Methanol (150 µL, spiked with Internal Standards, Supplementary Table 1), was added to each of the wells, followed by water (500 µL) and a mixture of solvents (500 µL) comprising dichloromethane and methanol (3:1) doped with triethylammonium chloride (500 mg/L). The mixture was agitated (96 channel pipette) before being centrifuged (3·2k × *g*, 2 min). A portion of the organic solution (20 µL) was transferred to a high throughput plate (384w, glass-coated, Esslab Plate+™) before being dried ($N_{2\,(g)}$). The dried films were re-dissolved (tert butyl methyl ester (TBME), 30 µL/well) and diluted with a stock mixture of alcohols and ammonium acetate (100 µL/well; propan-2-ol : methanol, 2:1; $CH_3COO.NH_4$ 7·5 mM). The analytical plate was heat-sealed and analysed immediately.

Instrument—All samples were infused into an Exactive Orbitrap (Thermo, Hemel Hampstead, UK), using a Triversa Nanomate (Advion, Ithaca, NY, USA). The Nanomate infusion mandrel was used to pierce the seal of each well before an aliquot of the solution (15 µL) was collected with an air gap (1·5 µL). The tip was pressed against a fresh nozzle and the sample was dispensed using 0·2 psi ($N_{2\,(g)}$). Ionisation was achieved at a 1·2 kV. The Exactive started acquiring data 20 s after sample aspiration began. The Exactive acquired data with a scan rate of 1 Hz (resulting in a mass resolution of 65,000 full width at half-maximum (fwhm) at 400 m/z). After 72 s of acquisition in positive mode the Nanomate and the Exactive switched over to negative mode, decreasing the voltage to −1·5 kV. The spray was maintained for another 66 s,

after which Collision-Induced Dissociation commenced, with a mass window of 50–1000 Da, and was stopped after another 66 s. The analysis was then stopped, and the tip discarded before the analysis of the next sample began. The sample plate was kept at 10 °C throughout the data acquisition. Samples were run in row order and repeated several times, if necessary, to ensure accuracy.

Data processing—Raw high-resolution mass-spectrometry data were processed using XCMS (www.bioconductor.org) and Peakpicker v 2.0 (an in-house R script [24]). Lists of known species (by *m/z*) were used for both positive ion and negative ion mode (~8·5k species). Variables whose mass deviated by more than 9 ppm from the expected value, had a signal-to-noise ratio of <3 and had signals for fewer than 50% of samples were discarded. The correlation of signal intensity to concentration of human placenta, mouse liver, human serum and pooled human seminal plasma samples as QCs (0.25, 0.5, 1.0×) were used to identify the lipid signals the strength of which was linearly proportional to abundance (threshold for acceptance was a correlation of 0.75). Signals were then divided by the sum of signals for that sample and expressed per mille (‰). Zero values were interpreted as not measured. All statistical calculations were done on these finalised values.

2.6 Statistical analyses

Participants were initially separated into lean controls, obese controls, and NAFLD cases. Clinical and biochemical characteristics of obese controls and NAFLD cases were compared using two-way unpaired T-tests for continuous data and chi-squared test for categorical data. Children with NAFLD who had liver histology available were initially separated into those with NASH and those with NAFL. Genotype differences were compared between groups using an allelic contrast model. p-values were converted to q-values by adjusting for multiple testing using the Benjamini-Hochberg method.

The association between clinical and biochemical characteristics of participants and liver histology was tested using regression analysis. Continuous independent variables (e.g. age, HOMA-IR), were logarithmically transformed and standardised (to mean = 0, standard deviation (SD) = 1). Univariable linear regression was performed for NAS, portal inflammation score, and fibrosis stage (dependent variable) and each independent variable. Logistic regression was performed for the presence of NASH (where 0 = NAFL, 1 = NASH). p-values were converted to q-values by adjusting for multiple testing using the Benjamini-Hochberg method. Critical q-value for significance was $q<0.05$.

Following initial processing as described above, relative abundances for each lipid were logarithmically transformed and standardised (to mean = 0, SD = 1). Missing values were ignored and no imputation was used. Univariable regression analyses adjusted for sex and age were run to test for associations between lipid species and outcomes. Categorical outcomes were tested using logistic regression: lean versus obese controls, obese controls versus NAFLD, and NAFL versus NASH. Linear regression analyses were performed for NAS, portal inflammation score, and fibrosis stage. Due to high correlation between lipid species (Supplementary Figure 1), the critical p-value for significance was defined by $0.05/\sqrt{n}$, where n

is the number of identified species from each analysis method. The critical p-value for each lipid species is listed in Supplementary Table 2. For all lipid species significantly associated with histological features (NAS, fibrosis, portal inflammation, or NASH diagnosis) on linear regression adjusted for age and sex, we repeated the regression additionally adjusting for BMI z-score and HOMA-IR.

For graphical representation, box plots (with Tukey's method) were generated for species that were significantly associated with outcomes on regression analysis. All data shown used in box plots was logarithmically adjusted and standardised (to mean = 0, SD = 1). p-values for box-plots were calculated using one-way ANOVA adjusted for multiple comparisons.

Heatmaps were generated using Morpheus (https://software.broadinstitute.org/morpheus).

Finally, random forest analysis[25], a supervised clustering algorithm, was used to determine the predictive capacity of lipidomics in this sample. We aimed to distinguish: (a) obese controls from NAFLD; and (b) simple fatty liver (NAFL) from NAFLD with either NASH or fibrosis ("NASH/fibrosis"). For each outcome, fandom forest analysis was first performed using clinical/biochemical characteristics, then lipidomics, and then lipidomics plus clinical/biochemistry. For lipidomics, lipid species were only included if relative abundance was recorded for >90% of participants (245/387 species (63%)). Blanks were then set to -9 standard deviations below to mean.

Participants were randomly separated into training (70%) and validation (30%) datasets. A training forest was built using the optimum number of trees (as determined by the 'bestmtry' function in R package randomForest) and the relative importance of each variable was extracted. The training algorithm was then tested in the independent validation set, from which receiver-operator characteristic (ROC) curves were generated.

Regression analysis was performed in Stata v16.1 (StataCorp) and random forest analysis was performed using R 3.6.1[26], in addition, graphs were produced in GraphPad v8.0 for Mac,

GraphPad Software, La Jolla California, USA). Code for analyses run in STATA and R are available in the Supplementary Data.

3. Results

306 children were recruited to the study: 19 lean controls, 65 obese controls, and 222 children with NAFLD (Table 1). Data were available and suitable for analysis from all participants. The lean controls were younger than the other two groups (mean age 11.0 years versus 12.9 years for NAFLD, p=.01) and their mean BMI z-score was normal (-0.04).

3.1 Characteristics of children with obesity versus NAFLD

The obese controls were older than children with NAFLD (mean age 14.1 years vs 12.9 years q=0.01) though had a similar degree of insulin resistance as assessed by HOMA-IR (3.1 (control) versus 3.7 (NAFLD), q=.18). However children with NAFLD were more dyslipidaemic, with higher total cholesterol and triglycerides (mean 0.9 mmol/L (control) versus 1.3 mmol/L (NAFLD),q=.02).

Of those with NAFLD, 120/222 (54%) had undergone liver biopsy, which allowed identification of children with NASH (Supplementary Table 3). Children with NASH had more severe NAFLD with a higher grade of steatosis (q= .002) and more fibrosis ($q=7.6x10^{-4}$). Compared to children with NAFL, patients with NASH were older (12.3 years vs 10.0 yrs, $q=6.5x10^{-6}$) and more insulin resistant (HOMA-IR 3.1 3.5 versus 2.6 (NAFLD) , q=0.01). There was no difference in liver enzymes (ALT, AST, GGT) or serum lipids (high/low density lipoprotein cholesterol, triglycerides, or total cholesterol) when comparing NAFL and NASH. Though when histological

features were analysed as continuous traits using linear regression, higher ALT, higher AST, and higher triglycerides were positively associated with NAS, portal inflammation, and fibrosis (Supplementary Table 4).

3.2 Lipid species associate with histological severity of NAFLD

To gain insight into the lipid metabolism in paediatric NAFLD we performed untargeted lipidomics on fasted plasma samples from our study sample. 383 lipid species were identified across 17 broad classes (Supplementary Figure 2).

From this heatmap, some patterns could be identified, such as lower phosphatidylinositols (PI) and lower phosphatidylcholines (PC) in children with NASH/fibrosis. However in order to more reliably identify which lipids were associated with NAFLD independent of obesity, we performed logistic regression to assess for differences between lean versus obese controls, and then obese controls versus children with NAFLD (Table 2, Supplementary Table 2, and Supplementary Figures 3-14). Then, to gain insight into how lipid metabolism was altered with increasing severity of NAFLD, we performed regression to test the association between lipid species and components of NAFLD histology (Table 3).

After adjustment for age and sex, 16 lipid species distinguished obese from lean controls and 66 lipids distinguished NAFLD from obese controls. For some species (and classes) there was a consistent trend from lean control to obese control to NAFLD, such as phosphatidylcholines (PC, Figure 1). Whereas for several other lipids demonstrated specific associations with NAFLD (and NASH), for example *lyso*-PC(18:0); levels were lower in obese controls than lean controls but higher in NASH or fibrosis (Figure 1B). Higher *lyso*-PC(18:0) was also positively associated

with NAS, portal inflammation score, and fibrosis stage. For NAS, this relationship remained after adjusting for BMI and HOMA-IR (β 1.0 (SE 0.3), p=0.002).

Multiple individual phospholipid species and summed scores for classes of phospholipids were associated with histological severity of NAFLD. Participants with NASH or fibrosis had the most perturbed concentrations of phospholipids, in particular lower phosphatidylinositols (PI, Figure 1C). The direction of effect was not altered by adjusting for BMI or HOMA-IR, for example the association between NAS and total PI: β -0.53 (SE 0.1), $p=2.5x10^{-4}$.

Plasma fatty acids were positively associated with the histological severity of NAFLD. This trend was consistent across saturated-, monounsaturated-, and polyunsaturated-fatty acids (PUFA) as well as fatty acid hydroxides (Figure 2A). Several of these associations were attenuated by adjusting for BMI and HOMA-IR, however positive associations remained for FA(22:4) with NAS and with diagnosis of NASH (β 1.0 (SE .3), p=0.002).

We also observed that with increasing severity of NAFLD histology, for example increasing NAFLD Activity Score, there was a positive associated with total free cholesterol and a negative association with cholesterol esters (Figure 2B). The association between free cholesterol and NAS remained after adjusting for BMI and HOMA-IR (β .67 (SE .2), $p=9.5x10^{-4}$), i.e. for a difference in 1 standard deviation of free cholesterol, NAS was 0.7 higher, independent of BMI, age, sex, and HOMA-IR.

From our results of baseline participant characteristics we had identified that triglycerides were positively associated with NAFLD and NASH. Our lipid profiling demonstrated several triglyceride (TG) and diglyceride (DG) species to be associated with histological severity of

NAFLD. We observed that polyunsaturated species (≥4 double-bonds (db)) were negatively associated with NAS, portal inflammation and NASH/fibrosis (Figure 3). These effects were almost completely attenuated by adjustment for insulin resistance (BMI and HOMA-IR).

3.3 Lipid species distinguish NAFL from NASH or fibrosis

Finally, we tested whether these differences in lipids could be combined into an algorithm to predict presence of NAFLD and severity of NAFLD. We used random forest analysis, a form of supervised clustering, to first distinguish NAFLD from obese controls. Using only clinical parameters, participants with NAFLD could be effectively separated from obese controls (Supplementary Figure 15) with an area under the ROC curve (AUROC) of 0.98 (95% CI 0.96-1.0) in the validation cohort. ALT, BMI, triglycerides, age, and AST contributed most to the model. Obese controls and NAFLD patients could be separated using lipidomics however it was less effective than using clinical parameters alone (AUROC .81 (95% CI .72-.90), Supplementary Table 5).

Using clinical and biochemical characteristics was only moderately effective in separating patients with NAFL from those with NASH/fibrosis (Figure 4), with an AUROC of 0.85 (95% CI .74-.95) in the validation cohort. Lipidomics showed good efficacy in separating children with NASH or fibrosis and those with NAFL alone, with AUROC .93 (95% CI 0.87-1.0) in the validation cohort. Unsaturated *lyso*-PC species made the greatest contribution to the model.

4. Discussion

Paediatric NAFLD is a common condition strongly associated with obesity and insulin resistance, however, only a minority of children are thought to go on to develop clinical end-stage liver disease later in life. Therefore, it is key to identify children with early evidence of more aggressive NAFLD and to understand why some young people progress whilst others have a more indolent liver phenotype. In this study we have used plasma lipidomics to identify perturbed lipid metabolism associated with NASH, separately from 'simple' fatty liver (NAFL) or obesity without NAFLD. Moreover, we then found that this lipid profile could predict which children would have NASH or fibrosis on liver biopsy with high accuracy.

Our analysis identified perturbation of multiple lipid pathways, including phospholipids and *lyso*-phospholipids. Many of the strongest associations with individual histological components were PC or *lyso*-PC species. The majority of other lipidomic studies in NAFLD have also identified substantial associations between PC species and NASH. Whilst much of this work to date has been in adults[11,12,27], the similarities between findings suggest similar alterations in lipid metabolism between adult and paediatric NAFLD.

Phosphatidylcholines are membrane-forming lipids and therefore in circulation their abundance is influenced by the concentration of lipoprotein particles. Many of the associations we observed are specific to NAFLD (i.e. they were not shared by obese, insulin resistant but non-steatotic children), which suggests that hepatic steatosis alters the membrane composition of circulating lipoparticles. Hartley *et al.* observed lower concentrations of large- and medium-sized HDL in children with NAFLD, which could account for lower PC[17]. We also found higher levels of several *lyso*-PCs in NASH, similar to the findings by Puri *et al.* (in the livers of adults with NASH)[10]. In addition, *lyso*-PC(18:1) was one of the top species identified in a separate study

that differentiated controls from children with NAFLD[16]. The *lyso*-PCs identified in the present study (LPC(16:0, 18:0)) are consistent with the activity of phospholipase A_2 (PLA_2) on PCs[28], therefore increased PLA_2 activity might account for these observations. There are several potential sources of increased PLA_2 activity, including lipoprotein-associated PLA_2 (Lp-PLA_2) and secreted forms of PLA_2[29]. PLA_2 is of particular interest in NAFLD (and cardio-metabolic disease) as its activity is thought to correlate with pro-inflammatory mediators, and presence of oxidised-LDL[30,31]. Whether (paediatric) NASH is an independent risk factor for atherosclerosis is a complex question[32] however cardiovascular disease is the top cause of mortality in NAFLD[33] and these lipid perturbations are consistent with a pro-atherogenic phenotype.

Similar to findings from lipidomics of liver biopsy samples by Puri *et al*.[10], we observed an increase in free cholesterol with a reduction in esterified cholesterol. In adults this has been attributed to reduced LDL-receptor expression, resulting in reduced uptake of free cholesterol from the plasma, as well as increased hepatic synthesis via HMG CoA reductase[34]. There is a large body of evidence implicating free cholesterol in the pathogenesis of NASH[35–37], with which our data appear to be consistent, again suggesting similarities between paediatric and adult NASH.

There is a strong body of work implicating hepatic *de novo* lipogenesis (DNL) in NAFLD[38] and as both a cause and consequence of insulin resistance[39,40]. Lipidomics has been suggested to offer some indirect evidence for changes in DNL[41]. Hepatic DNL is classically associated with a higher abundance of shorter, saturated TG and lower, longer, unsaturated TG[42], which correlates closely with the effects of (hepatic) insulin resistance[43]. We found a negative association with TG(54:2), TG(54:4), and DG(37:4), which were attenuated by including BMI and HOMA-IR as covariates in regression analyses, though did not find positive associations with TG(48:0) or TG(50:0), as might have been expected. In addition, whilst we found a few

weak associations between fatty acids (from phospholipid fractions) that might be suggestive of DNL (e.g. positive association between NAS and ratio of FA(16:0) to FA(16:1))[44], there is evidence to suggest that this might not be a reliable method to derive conclusions about hepatic DNL[45]. Overall, we consider these results to be reflective of hepatic insulin resistance but an alternative methodology, for instance using isotopically labeled substrates, would be needed to formally investigate DNL in paediatric NASH.

More prominent peri-portal inflammation on liver biopsy is a characteristic feature of paediatric NAFLD[8,9]. In this analysis we found similar lipid species to be associated with NASH, NAS, and portal inflammation score, with the same direction of effect. This suggests that the perturbation of lipid metabolism that occurs in children with portal inflammation is similar to that of lobular inflammation and ballooning (i.e. classical 'adult' NASH).

A wide range of associations have been identified in previous metabolite (or lipid) profiling studies in children with NAFLD. Several have found higher levels of (branched-chain) amino acids[16,17,46–50]. There is strong evidence that levels of circulating amino acids, particularly branched chain amino acids, are highly correlated to (and even causal of) insulin resistance[51,52]. Other studies have focused on gastro-intestinal tract-derived metabolites[53], which also appear to show utility in differentiating controls from NAFLD, though are less effective in separating NAFL and NASH[54].

Metabolite profiling studies in adults with NAFLD have been able to develop prediction models that effectively distinguish NAFL from NASH. Gorden *et al*. performed lipidomics on both plasma and liver; they found a profile of plasma phospholipids and ceramides could completely separate NAFLD from NASH[12]. A recent paper including 27 patients with biopsy-staged NAFLD found a combination of PC, ceramides, diglycerides, and triglycerides separated NAFL from

NASH[55]. Our current work provides proof of concept that untargeted metabolite profiling can effectively non-invasively diagnose NASH in children. Therefore we propose that this panel of lipids can be considered a candidate panel of biomarkers that requires validation in an external cohort.

The main strength of this study was the inclusion of participants from across the spectrum of paediatric NAFLD plus obese and lean controls. This allowed identification of patterns specific to NASH and NAFL separate to obesity. The observation that some lipids (for example *lyso*-PC(18:0)) were negatively associated with obesity but positively associated with NASH suggest that hepatic steatosis (or steatohepatitis) has a unique effect on the serum lipid profile beyond insulin resistance *per se*. Our sample size was large enough to facilitate identification of associations in all tested outcomes. Lastly, use of an untargeted approach allowed us to interrogate multiple pathways of lipid metabolism.

The principal limitation of this work is the lack of liver biopsy samples for lipidomics, which would have improved our understanding of lipid metabolism at the level of the hepatocyte. However, studies that have included paired liver and plasma samples have found considerable overlap between significantly associated species[12]. We also identified a number of changes in the plasma consistent with previous reports from liver samples in adults[10]. It should also be noted that our participants were primarily of non-Finnish European descent and therefore it is unclear to what extent these findings are generalisable to other ethnicities. Finally, whilst we have illustrated proof of concept for this technique for non-invasive stratification of NAFL versus NASH, it is not currently realistic to suggest untargeted lipidomics as a routine clinical test. Further work, both practical and conceptual, would be required to progress this technique towards clinical utility.

Conclusion

Lipid profiling of plasma in children with obesity and NAFLD shows perturbation of multiple lipid pathways, several apparently independent of insulin resistance and associated with histological severity of liver disease. Such a plasma lipidomics technique can reliably distinguish simple fatty liver from NASH or fibrosis in children and external validation of this candidate biomarker panel is of high clinical relevance.

5. References

1. Sanyal, A. J. Past, present and future perspectives in nonalcoholic fatty liver disease. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 377–386 (2019).
2. Taylor, R. S. *et al.* Association Between Fibrosis Stage and Outcomes of Patients With Nonalcoholic Fatty Liver Disease: A Systematic Review and Meta-Analysis. *Gastroenterology* (2020) doi:10.1053/j.gastro.2020.01.043.
3. Brunt, E. M. *et al.* Improvements in Histologic Features and Diagnosis Associated With Improvement in Fibrosis in Nonalcoholic Steatohepatitis: Results From the Nonalcoholic Steatohepatitis Clinical Research Network Treatment Trials. *Hepatology* **70**, 522–531 (2019).
4. Goldner, D. & Lavine, J. E. NAFLD in Children: Unique Considerations and Challenges. *Gastroenterology* (2020) doi:10.1053/j.gastro.2020.01.048.
5. Lavine, J. E. *et al.* Effect of Vitamin E or Metformin for Treatment of Nonalcoholic Fatty Liver Disease in Children and Adolescents: The TONIC Randomized Controlled Trial. *JAMA* **305**, 1659–1668 (2011).
6. Feldstein, A. E. *et al.* The natural history of nonalcholic fatty liver disease in children: a follow-up study for up to 20-years. *Gut* **58**, 1538–1544 (2010).
7. Carter-kent, C. *et al.* Nonalcoholic steatohepatitis in children: A multicentre clinicopathological study. *Hepatology* **50**, 1113–1120 (2009).
8. Brunt, E. M. *et al.* Portal chronic inflammation in nonalcoholic fatty liver disease (NAFLD): a histologic marker of advanced NAFLD-clinicopathologic correlation from the Nonalcoholic Steatohepatitis Clinical Research Network. *Hepatology* **49**, 809–820 (2009).
9. Mann, J. P. *et al.* Portal inflammation is independently associated with fibrosis and metabolic syndrome in pediatric nonalcoholic fatty liver disease. *Hepatology* **63**, 745–753

(2016).

10. Puri, P. *et al.* A lipidomic analysis of nonalcoholic fatty liver disease. *Hepatology* **46**, 1081–1090 (2007).
11. Chiappini, F. *et al.* Metabolism dysregulation induces a specific lipid signature of nonalcoholic steatohepatitis in patients. *Sci. Rep.* **7**, 46658 (2017).
12. Gorden, D. L. *et al.* Biomarkers of NAFLD progression: a lipidomics approach to an epidemic. *J. Lipid Res.* **56**, 722–736 (2015).
13. Puri, P. *et al.* The plasma lipidomic signature of nonalcoholic steatohepatitis. *Hepatology* **50**, 1827–1838 (2009).
14. Zhou, Y. *et al.* Non-invasive Detection of Non-alcoholic Steatohepatitis Using Clinical Markers and Circulating Levels of Lipids and Metabolites. *Clin. Gastroenterol. Hepatol.* **14**, 1463–1472.e6 (2016).
15. Anjani, K. *et al.* Circulating phospholipid profiling identifies portal contribution to NASH signature in obesity. *J. Hepatol.* **62**, 905–912 (2015).
16. Khusial, R. D. *et al.* Development of a Plasma Screening Panel for Pediatric Nonalcoholic Fatty Liver Disease Using Metabolomics. *Hepatol Commun* **3**, 1311–1321 (2019).
17. Hartley, A., Santos Ferreira, D. L., Anderson, E. L. & Lawlor, D. A. Metabolic profiling of adolescent non-alcoholic fatty liver disease. *Wellcome Open Research* **3**, 166 (2018).
18. Mann, J. P. *et al.* European paediatric non-alcoholic fatty liver disease registry (EU-PNAFLD): Design and rationale. *Contemp. Clin. Trials* **75**, 67–71 (2018).
19. Vos, M. B. *et al.* NASPGHAN Clinical Practice Guideline for the Diagnosis and Treatment of Nonalcoholic Fatty Liver Disease in Children: Recommendations from the Expert Committee on NAFLD (ECON) and the North American Society of Pediatric Gastroenterology, Hepatology and Nu. *J. Pediatr. Gastroenterol. Nutr.* **64**, 319–334 (2017).
20. Vajro, P. *et al.* Diagnosis of nonalcoholic fatty liver disease in children and adolescents: position paper of the ESPGHAN Hepatology Committee. *J. Pediatr. Gastroenterol. Nutr.* **54**,

700–713 (2012).

21. Kleiner, D. E. *et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41**, 1313–1321 (2005).
22. Bedossa, P. *et al.* Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. *Hepatology* **60**, 565–575 (2014).
23. Furse, S. *et al.* A high-throughput platform for detailed lipidomic analysis of a range of mouse and human tissues. *Anal. Bioanal. Chem.* (2020) doi:10.1007/s00216-020-02511-0.
24. Harshfield, E. L. *et al.* An Unbiased Lipid Phenotyping Approach To Study the Genetic Determinants of Lipids and Their Association with Coronary Heart Disease Risk Factors. *J. Proteome Res.* **18**, 2397–2410 (2019).
25. Liaw, A., Wiener, M. & Others. Classification and regression by randomForest. *R news* **2**, 18–22 (2002).
26. R Core Team. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. (2019).
27. Feldman, A. *et al.* Clinical and metabolic characterization of obese subjects without non-alcoholic fatty liver: A targeted metabolomics approach. *Diabetes Metab.* **45**, 132–139 (2019).
28. Steinbrecher, U. P. & Pritchard, P. H. Hydrolysis of phosphatidylcholine during LDL oxidation is mediated by platelet-activating factor acetylhydrolase. *J. Lipid Res.* **30**, 305–315 (1989).
29. Lambeau, G. & Gelb, M. H. Biochemistry and physiology of mammalian secreted phospholipases A2. *Annu. Rev. Biochem.* **77**, 495–520 (2008).
30. Kim, J. Y. *et al.* Lipoprotein-associated phospholipase A2 activity is associated with coronary artery disease and markers of oxidative stress: a case-control study. *Am. J. Clin. Nutr.* **88**, 630–637 (2008).

31. Ha, C. Y. *et al.* The association of specific metabolites of lipid metabolism with markers of oxidative stress, inflammation and arterial stiffness in men with newly diagnosed type 2 diabetes. *Clin. Endocrinol.* **76**, 674–682 (2012).
32. Alexander, M. *et al.* Non-alcoholic fatty liver disease and risk of incident acute myocardial infarction and stroke: findings from matched cohort study of 18 million European adults. *BMJ* **367**, l5367 (2019).
33. Angulo, P. *et al.* Liver Fibrosis, but No Other Histologic Features, Is Associated With Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology* **149**, 389–397.e10 (2015).
34. Min, H.-K. *et al.* Increased hepatic synthesis and dysregulation of cholesterol metabolism is associated with the severity of nonalcoholic fatty liver disease. *Cell Metab.* **15**, 665–674 (2012).
35. Caballero, F. *et al.* Enhanced free cholesterol, SREBP-2 and StAR expression in human NASH. *J. Hepatol.* **50**, 789–796 (2009).
36. Van Rooyen, D. M. *et al.* Hepatic free cholesterol accumulates in obese, diabetic mice and causes nonalcoholic steatohepatitis. *Gastroenterology* **141**, 1393–1403.e5 (2011).
37. Ioannou, G. N., Haigh, W. G., Thorning, D. & Savard, C. Hepatic cholesterol crystals and crown-like structures distinguish NASH from simple steatosis. *J. Lipid Res.* **54**, 1326–1334 (2013).
38. Donnelly, K. L. *et al.* Sources of fatty acids stored in liver and secreted via lipoproteins in patients with nonalcoholic fatty liver disease. *J. Clin. Invest.* **115**, 1343–1351 (2005).
39. Smith, G. I. *et al.* Insulin resistance drives hepatic de novo lipogenesis in nonalcoholic fatty liver disease. *J. Clin. Invest.* **130**, 1453–1460 (2020).
40. Lambert, J. E., Ramos-Roman, M. A., Browning, J. D. & Parks, E. J. Increased de novo lipogenesis is a distinct characteristic of individuals with nonalcoholic fatty liver disease. *Gastroenterology* **146**, 726–735 (2014).

41. Paglialunga, S. & Dehn, C. A. Clinical assessment of hepatic de novo lipogenesis in non-alcoholic fatty liver disease. *Lipids Health Dis.* **15**, 159 (2016).
42. Sanders, F. W. B. *et al.* Hepatic steatosis risk is partly driven by increased de novo lipogenesis following carbohydrate consumption. *Genome Biol.* **19**, 79 (2018).
43. Eiden, M. *et al.* Mechanistic insights revealed by lipid profiling in monogenic insulin resistance syndromes. *Genome Med.* **7**, 63 (2015).
44. Lee, J. J. *et al.* Palmitoleic acid is elevated in fatty liver disease and reflects hepatic lipogenesis. *Am. J. Clin. Nutr.* **101**, 34–43 (2015).
45. Rosqvist, F. *et al.* Fasting hepatic de novo lipogenesis is not reliably assessed using circulating fatty acid markers. *Am. J. Clin. Nutr.* **109**, 260–268 (2019).
46. Miccheli, A. *et al.* Urinary (1)H-NMR-based metabolic profiling of children with NAFLD undergoing VSL#3 treatment. *Int. J. Obes.* **39**, 1118–1125 (2015).
47. Jin, R. *et al.* Amino Acid Metabolism is Altered in Adolescents with Nonalcoholic Fatty Liver Disease-An Untargeted, High Resolution Metabolomics Study. *J. Pediatr.* **172**, 14–19.e5 (2016).
48. Troisi, J. *et al.* Urinary Metabolomics in Pediatric Obesity and NAFLD Identifies Metabolic Pathways/Metabolites Related to Dietary Habits and Gut-Liver Axis Perturbations. *Nutrients* **9**, (2017).
49. Kaikkonen, J. E. *et al.* Metabolic profiling of fatty liver in young and middle-aged adults: Cross-sectional and prospective analyses of the Young Finns Study. *Hepatology* **65**, 491–500 (2017).
50. Goffredo, M. *et al.* A Branched-Chain Amino Acid-Related Metabolic Signature Characterizes Obese Adolescents with Non-Alcoholic Fatty Liver Disease. *Nutrients* **9**, (2017).
51. Würtz, P. *et al.* Branched-chain and aromatic amino acids are predictors of insulin resistance in young adults. *Diabetes Care* **36**, 648–655 (2013).

52. Lotta, L. A. *et al.* Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis. *PLoS Med.* **13**, e1002179 (2016).
53. Alkhouri, N. *et al.* Analysis of breath volatile organic compounds as a noninvasive tool to diagnose nonalcoholic fatty liver disease in children. *Eur. J. Gastroenterol. Hepatol.* **26**, 82–87 (2014).
54. Del Chierico, F. *et al.* Gut microbiota profiling of pediatric nonalcoholic fatty liver disease and obese patients unveiled by an integrated meta-omics-based approach. *Hepatology* **65**, 451–464 (2017).
55. Perakakis, N. *et al.* Non-invasive diagnosis of non-alcoholic steatohepatitis and fibrosis with the use of omics and supervised learning: A proof of concept study. *Metabolism* **101**, 154005 (2019).

Tables

	Lean controls	Obese controls	NAFLD	p-value	q-value
	N=19	N=65	N=222		
Female, N (%)	9 (47)	31 (48)	103 (46)	.85	0.85
Age (years)	11.0 (3.4)	14.1 (2.1)	12.9 (2.8)	.002	0.01
BMI z-score	-.04 (1.2)	3.2 (.4)	2.5 (.8)	7.7×10^{-10}	4.1×10^{-9}
Waist circumference (cm)	-	101.8 (12.2)	102.6 (15.3)	0.76	0.82
Hip circumference (cm)	-	120.1 (11.4)	124.5 (13.6)	.11	0.18
Systolic BP (mmHg)	-	121.7 (16.3)	127.6 (15.6)	.05	0.10
Diastolic BP (mmHg)	-	78.7 (9.6)	79.4 (9.7)	.70	0.82
HOMA-IR	-	3.1 (2.1)	3.7 (2.2)	.10	0.18
ALT (IU/L)	-	24.0 (11.8)	63.4 (43.0)	3.0×10^{-11}	4.8×10^{-10}
AST (IU/L)	-	24.2 (7.3)	42.6 (21.2)	3.2×10^{-10}	2.6×10^{-9}
GGT (IU/L)	-	19.9 (9.0)	29.3 (14.1)	1.7×10^{-5}	5.4×10^{-5}
ALP (IU/L)	-	172 (110)	198 (127)	.77	0.82
Total cholesterol (mmol/L)	-	3.9 (.8)	4.2 (.9)	.01	0.02
HDL (mmol/L)	-	1.1 (.2)	1.1 (.2)	.53	0.71
LDL (mmol/L)	-	2.4 (.8)	2.5 (.5)	.30	0.44
Triglycerides (mmol/L)	-	0.9 (.4)	1.3 (.7)	4.3×10^{-6}	1.7×10^{-5}

Table 1. Participant demographics and biochemistry. Continuous variables are given as mean (standard deviation). p- and q-values are given for comparison between obese controls and NAFLD patients. BMI, body mass index; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma glutamyl transferase; ALP, alkaline phosphatase; HDL, high density lipoprotein; LDL, low density lipoprotein; HOMA-IR, homeostatic model assessment of insulin resistance.

Lipid species associated with outcomes	Measured m/z	Direction of association	Unadjusted fold change	Age, sex-adjusted β (SE)	p-value
Lean (n=19) vs. Obese controls (n= 65)					
PC(34:2)[$M^+Cl]^-$	792.532	Positive	1.27	2.5 (.7)	$5.1x10^{-4}$
PC(34:1)[$M^+H]^+$	760.585	Negative	0.77	-1.8 (.5)	$5.1x10^{-4}$
LPC(22:2)[$M^+OAC]^-$	634.409	Positive	6.33	2.2 (.6)	$5.4x10^{-4}$
PC(37:5)[$M^+OAC]^-$	852.576	Negative	0.13	-4.0 (1.2)	$7.1x10^{-4}$
PC(34:1)[$M^+OAC]^-$	818.592	Negative	0.71	-3.2 (.9)	$8.2x10^{-4}$
Obese control (n=65) vs. NAFLD (n=222)					
LPC(18:1)[$M^-H]^-$	478.294	Positive	4.32	1.15 (.2)	$1.6x10^{-8}$
LPC(17:0)[$M^+OAC]^-$	568.362	Positive	1.81	1.07 (.19)	$2.4x10^{-8}$
PC(34:3)[$M^+H]^+$	756.554	Positive	1.23	2.1 (.39)	$7.9x10^{-8}$
LPC(20:0)[$M^-H]^-$	508.3409	Positive	1.26	.95 (.18)	$1.3x10^{-8}$
PC(35:4)[$M^+OAC]^-$	826.560	Negative	0.80	-1.46 (.28)	$1.4x10^{-7}$
NAFL (n=34) vs. NASH (n=86)					
PC(35:2)[$M^+OAC]^-$	830.592	Negative	0.91	-1.3 (.38)	$5.5x10^{-4}$
FA(22:4)[$M^-H]^-$	331.264	Positive	1.04	1.0 (.33)	$1.4x10^{-3}$
PS-O(35:2)[$M^-H]^-$	758.534	Positive	1.08	1.1 (.37)	$2.0x10^{-3}$
TG(54:02)[$M^+NH_4]^+$	904.833	Negative	0.93	-1.7 (.57)	$2.7x10^{-3}$
PI Sum	Sum PI species	Negative	0.79	-.83 (.32)	$8.8x10^{-3}$

Table 2. Top associations between plasma lipid species and dichotomous outcomes. Association between lipid species and outcomes were calculated using logistic regression using log-normalised relative lipid abundance with adjustment for age and sex. β (SE, standard error) represents the change in log-odds per standard deviation change in lipid species. For all outcomes, control was coded as 0 and cases as 1 (e.g. NAFL=0, NASH=1), therefore a positive

beta correlation coefficient indicates that an increase in the lipid species is associated with higher odds of NASH, for example. FA, fatty acid; FA-OH, hydroxylated fatty acid;LPC, *lyso*-phosphatidylcholine; PC, phosphatidylcholine; PC-O phosphatidylcholine plasmalogen; PI, phosphatidylinositol; PS-O, phosphatidylserine oxide; TG, triglyceride.

Lipid species associated with outcomes	m/z	Direction of association	Unadjusted fold change	Age, sex-adjusted β (SE)	p-value
NAS [0-9] (n=120)					
PC(34:3)[$M^+H]^+$	756.5538	Positive	2.89	2.2 (.11)	$7.9x10^{-41}$
LPC(16:0)[$M^+H]^+$	496.3398	Positive	1.06	2.1 (.13)	$3.5x10^{-35}$
FA-OH(18:2)[$M^-H]^-$	295.2279	Positive	1.38	2.0 (.12)	$3.7x10^{-35}$
LPC(18:0)[$M^+H]^+$	524.3711	Positive	1.08	2.1 (.14)	$7.9x10^{-34}$
LPC(17:0)[$M^-H]^-$	568.362	Positive	1.04	2.0 (.14)	$4.3x10^{-29}$
Portal inflammation [0-2] (n=118)					
LPC(16:0)[$M^+H]^+$	496.3398	Positive	1.01	.36 (.05)	$6.7x10^{-13}$
LPC(18:1)[$M^+H]^+$	522.3554	Positive	1.10	.36 (.05)	$8.3x10^{-13}$
LPC(18:0)[$M^+H]^+$	524.3711	Positive	1.00	.35 (.05)	$5.1x10^{-12}$
LPC(17:0)[$M^-H]^-$	568.362	Positive	1.00	.33 (.05)	$1.5x10^{-11}$
FA-OH(18:2)[$M^-H]^-$	295.2279	Positive	1.00	.31 (.04)	$1.0x10^{-10}$
Fibrosis [0-3] (n=120)					
LPC(17:0)[$M^-H]^-$	568.363458	Positive	0.95	.46 (.06)	$1.7x10^{-13}$
LPC(18:0)[$M^+H]^+$	524.3711	Positive	0.97	.47 (.06)	$3.5x10^{-13}$
PC(36:4)[$M^+Cl]^-$	816.5316	Negative	0.45	-.41 (.05)	$6.2x10^{-13}$
FA-OH(18:2)[$M^-H]^-$	295.2279	Positive	1.02	.42 (.05)	$6.5x10^{-13}$
LPC(16:0)[$M^+H]^+$	496.3398	Positive	0.98	.43 (.06)	$9.6x10^{-12}$

Table 3. Top associations between plasma lipid species and histological severity of NAFLD. Association between lipid species and outcomes were calculated using linear regression using log-normalised relative lipid abundance with adjustment for age and sex. Unadjusted fold changes calculated as ratios between mean abundance for: NAS 1-2 vs NAS 6-7; portal inflammation 0 vs 2; fibrosis 0 vs. 2-3. β (SE, standard error) represents a unit change in each outcome (range indicated in square brackets for each outcome) per standard deviation change in lipid species. FA, fatty acid; FA-OH, hydroxylated fatty acid;LPC, *lyso*-

phosphatidylcholine; PC, phosphatidylcholine; PC-O phosphatidylcholine plasmalogen; PI, phosphatidylinositol; PS-O, phosphatidylserine oxide; TG, triglyceride.

Figures

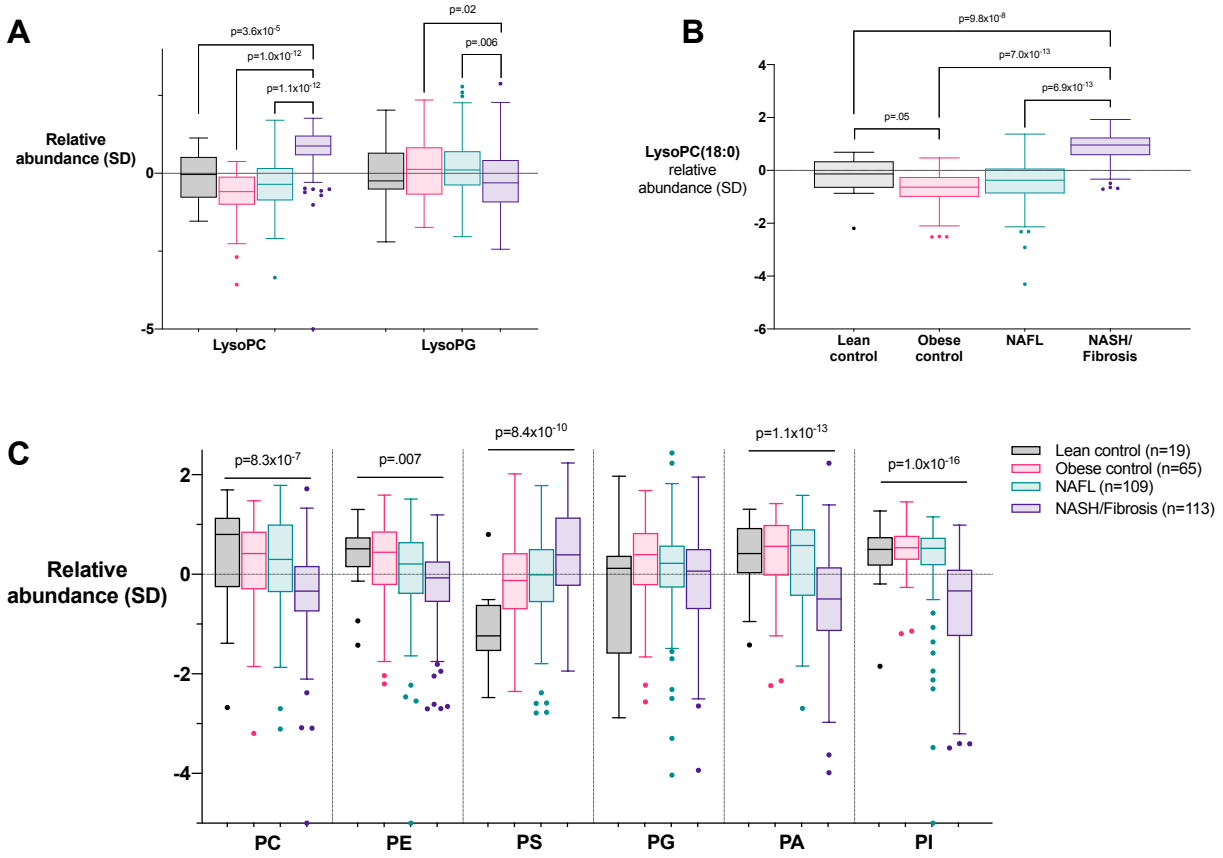

Figure 1. Paediatric NAFLD is associated with changes in phospholipids. Total (sum) for each *lyso*-phospholipid subclass (A) as well as individual *lyso*-phospholipids (B) differed between groups. Almost all classes of total (sum) plasma phospholipids were associated with NAFLD (C). The relative abundance of each lipid species was normalised (mean = 0, and standard deviation (SD) = 1). Box plots are shown using Tukey's method for outliers. p-values represent results for one-way ANOVA with adjustment for multiple comparisons. NAFL, non-alcoholic fatty liver; NASH, non-alcoholic steatohepatitis; PA, phosphatidic acid; PC, phosphatidylcholine; PE, phosphoethanolamine; PG, phosphatidylglycerol; PI, phosphatidylinositol; PS, phosphatidylserine.

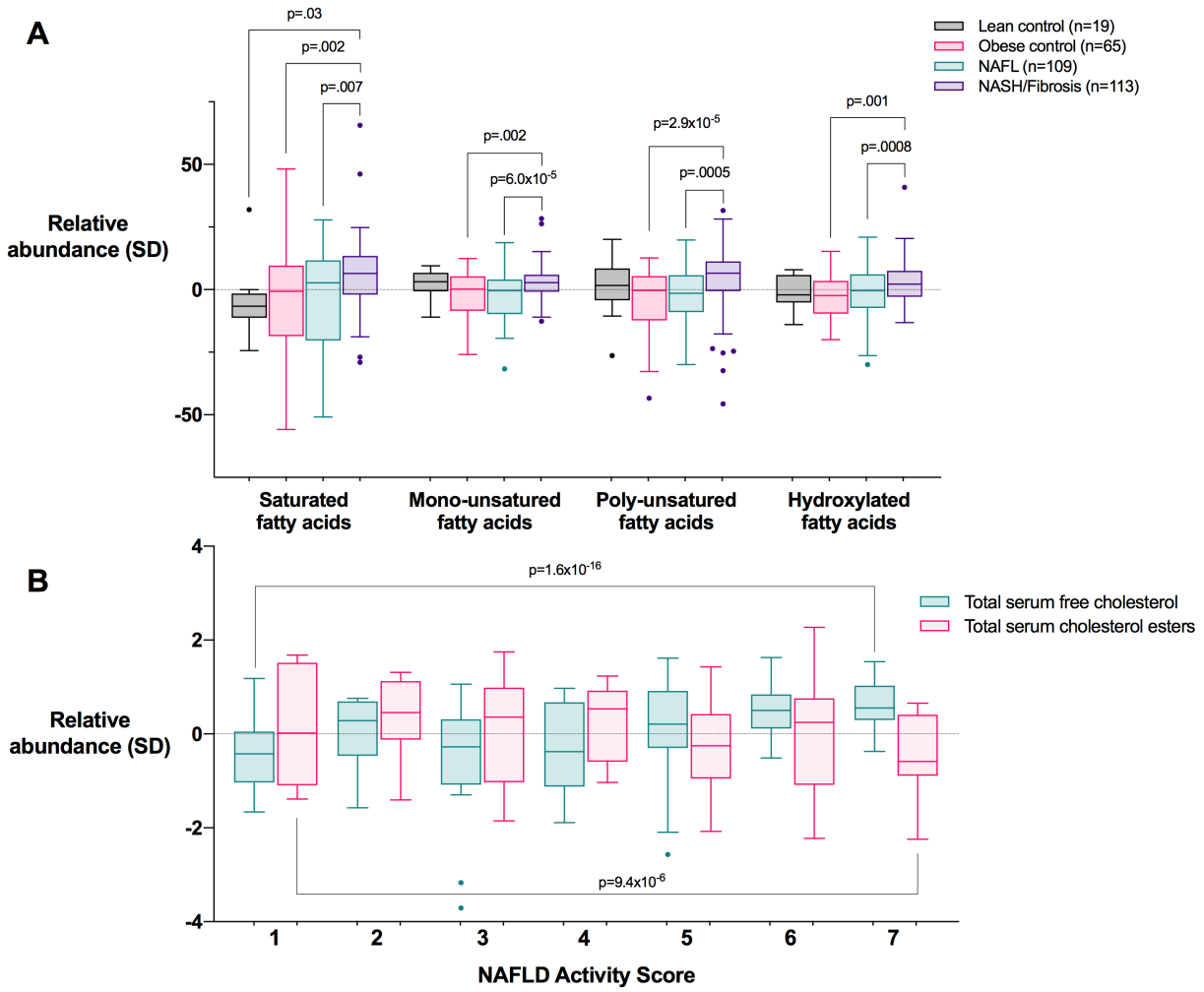

Figure 2. Severity of paediatric NAFLD is associated with changes in fatty acids (A) and cholesterol (B). Total (sum) for each subclass of fatty acid, derived from the phospholipid fraction of plasma (A). Box plots are shown using Tukey's method for outliers. p-values represent results for one-way ANOVA with adjustment for multiple comparisons. Total (sum) free cholesterol (green) and cholesteryl esters (red) for children with liver biopsies scored by the NAFLD Activity Score (n=120, B). p-values represent the result of linear regression adjusted for age and sex. The relative abundance of each lipid species was normalised (mean = 0, and standard deviation (SD) = 1). NAFL, non-alcoholic fatty liver; NASH, non-alcoholic steatohepatitis.

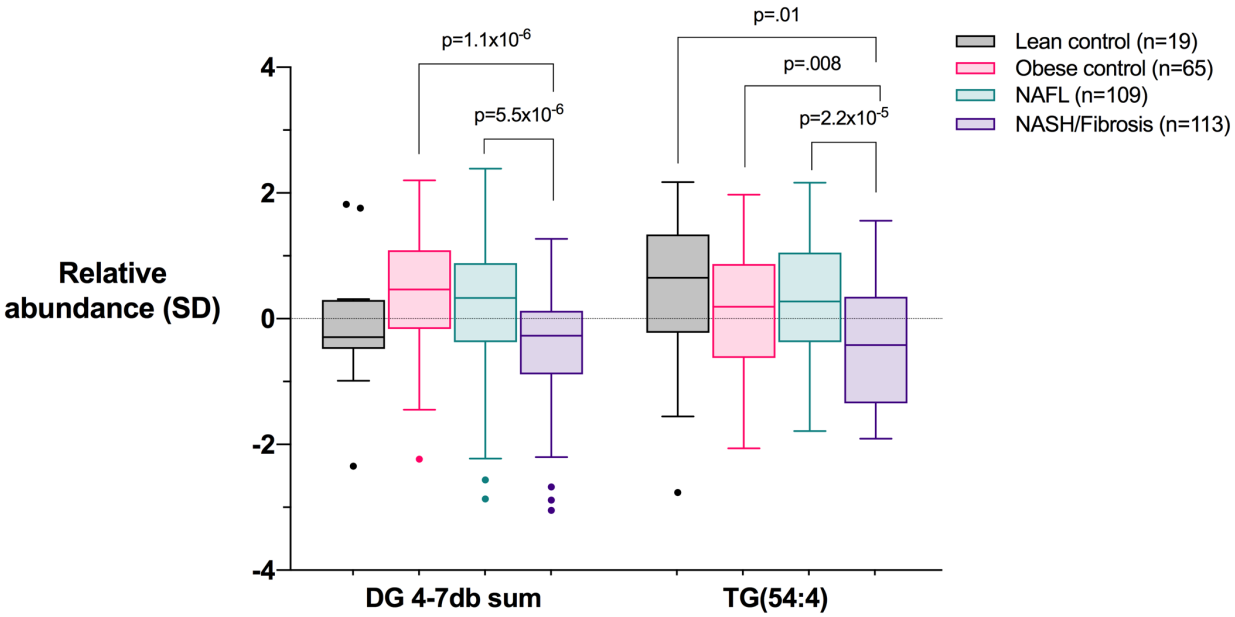

Figure 3. Paediatric non-alcoholic steatohepatitis (NASH) is negatively associated with polyunsaturated diglycerides (DG) and triglycerides (TG). Box plots are shown using Tukey's method for outliers. p-values represent results for one-way ANOVA with adjustment for multiple comparisons. db, double-bonds; NAFL, non-alcoholic fatty liver.

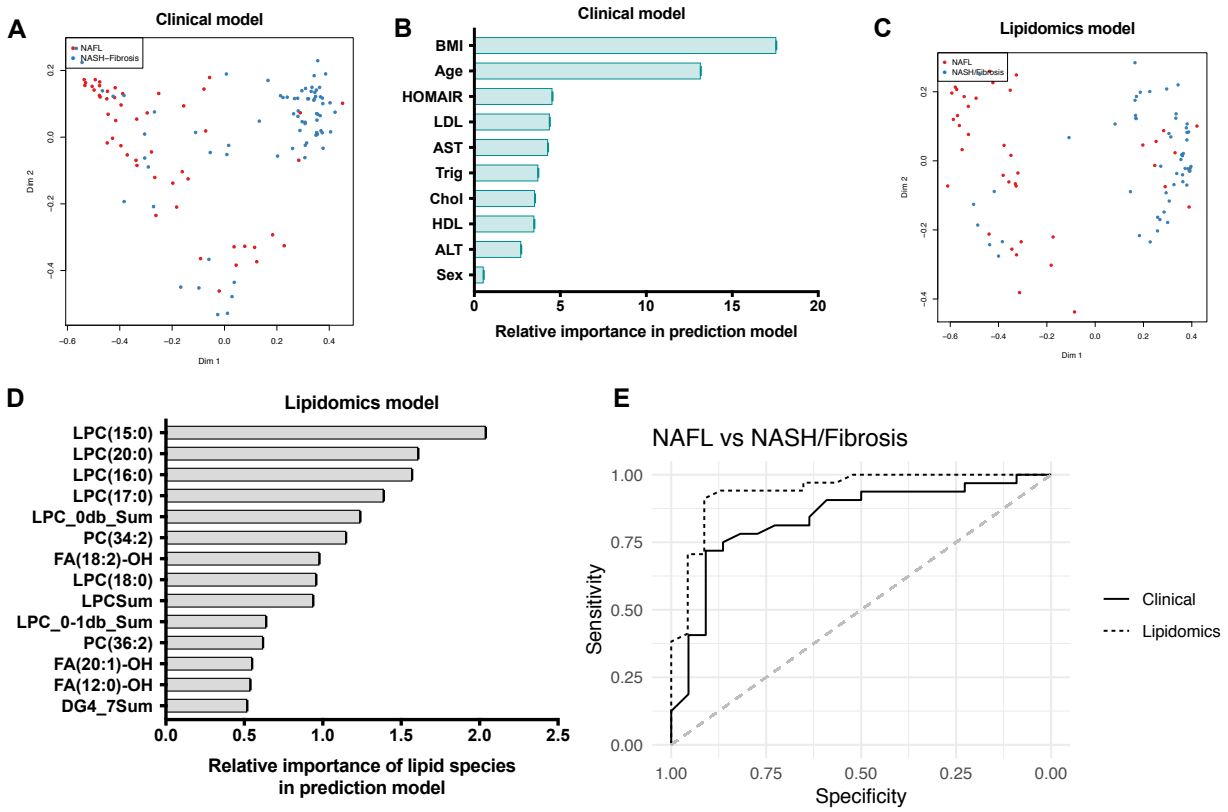

Figure 4. Plasma lipidomics distinguishes non-alcoholic fatty liver (NAFL) from non-alcoholic steatohepatitis (NASH) or fibrosis. Random forest analysis was used to build a predictive model using clinical and biochemical parameters in a training set (A), where BMI, age, and HOMA-IR were the most important variables (B). The process was replicated using lipidomics in a training set (C), with phospholipids having the greatest importance in the model (D). These models were tested in separate validation sets (E), where the lipidomics model showed complete separation of NAFL and NASH/fibrosis participants. DG4_7Sum, sum of diglycerides with 4-7 double-bonds; FA-OH, fatty acid hydroxide; LPC, *lyso*-phosphatiydlcholines; LPC_0db_sum, sum of saturated LPC species; LPC_0+1db_sum, sum of saturated and monounsaturated LPC species; LPCSum, sum of all LPC species; PC, phosphatidylcholine.

STROBE statement checklist

	Item No	Recommendation	Page no.
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	3
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	8
		(*b*) For matched studies, give matching criteria and the number of controls per case	-

Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	9-10
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*8-9*
Bias	9	Describe any efforts to address potential sources of bias	7
Study size	10	Explain how the study size was arrived at	8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	9-11
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	11-12
		(*b*) Describe any methods used to examine subgroups and interactions	12
		(*c*) Explain how missing data were addressed	12
		(*d*) If applicable, explain how matching of cases and controls was addressed	-
		(*e*) Describe any sensitivity analyses	-
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	14

		(b) Give reasons for non-participation at each stage	-
		(c) Consider use of a flow diagram	-
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	14 Table 1
		(b) Indicate number of participants with missing data for each variable of interest	Table 1
Outcome data	15*	Report numbers in each exposure category, or summary measures of exposure	Table 1
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Tables 2-3
		(*b*) Report category boundaries when continuous variables were categorized	Tables 2-3
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	-
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	17
Discussion			
Key results	18	Summarise key results with reference to study objectives	18

Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	21
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	22
Generalisability	21	Discuss the generalisability (external validity) of the study results	22
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	2