Contents lists available at ScienceDirect



Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj



# Distinct changes in serum metabolites and lipid species in the onset and progression of NAFLD in Obese Chinese

Check fo

Jiarui Chen <sup>a,b,c,1</sup>, Ronald Siyi Lu <sup>a,b,1</sup>, Candela Diaz-Canestro <sup>a,b</sup>, Erfei Song <sup>a,b,d</sup>, Xi Jia <sup>a,b</sup>, Yan Liu <sup>a,b</sup>, Cunchuan Wang <sup>d</sup>, Cynthia K.Y. Cheung <sup>a,b</sup>, Gianni Panagiotou <sup>b,c,f,g,\*</sup>, Aimin Xu <sup>a,b,e,\*\*,2</sup>

<sup>a</sup> State Key Laboratory of Pharmaceutical Biotechnology, The University of Hong Kong, Hong Kong Special Administrative Region

<sup>b</sup> Department of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region

<sup>c</sup> Leibniz Institute for Natural Product Research and Infection Biology, Microbiome Dynamics, Hans Knöll Institute, Jena, Germany

<sup>d</sup> Department of Metabolic and Bariatric Surgery, The First Affiliated Hospital of Jinan University, Guangzhou, China

<sup>e</sup> Department of Pharmacology and Pharmacy, the University of Hong Kong, Hong Kong Special Administrative Region

<sup>f</sup> Friedrich Schiller University, Faculty of Biological Sciences, Jena, Germany

<sup>8</sup> Cluster of Excellence Balance of the Microverse, Friedrich-Schiller-University Jena, Jena, Germany

ARTICLE INFO

Keywords: Non-alcoholic fatty liver disease Simple steatosis Nonalcoholic steatohepatitis Metabolomics Lipidomics Biomarkers Prediction model Chinese Obesity Multi-omics

#### ABSTRACTS

*Introduction:* Metabolic disturbances are major contributors to the onset and progression of non-alcoholic fatty liver disease (NAFLD), which includes a histological spectrum ranging from single steatosis (SS) to non-alcoholic steatohepatitis (NASH). This study aimed to identify serum metabolites and lipids enriched in different histological stages of NAFLD and to explore metabolites/lipids as non-invasive biomarkers in risk prediction of NAFLD and NASH in obese Chinese.

*Methods*: Serum samples and liver biopsies were obtained from 250 NAFLD subjects. Untargeted metabolomic and lipidomic profiling were performed using Liquid Chromatography-Mass Spectrometry. Significantly altered metabolites and lipids were identified by MaAsLin2. Pathway enrichment was conducted with MetaboAnalyst and LIPEA. WGCNA was implemented to construct the co-expression network. Logistic regression models were developed to classify different histological stages of NAFLD.

*Results*: A total of 263 metabolites and 550 lipid species were detected in serum samples. Differential analysis and pathway enrichment analysis revealed the progressive patterns in metabolic mechanisms during the transition from normal liver to SS and to NASH, including N-palmitoyltaurine, tridecylic acid, and branched-chain amino acid signaling pathways. The co-expression network showed a distinct correlation between different triglyceride and phosphatidylcholine species with disease severity. Multiple models classifying NAFLD versus normal liver and NASH versus SS identified important metabolic features associated with significant improvement in disease prediction compared to conventional clinical parameters.

\* Corresponding author at: Department of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region.

\*\* Corresponding author at: State Key Laboratory of Pharmaceutical Biotechnology, The University of Hong Kong, Hong Kong Special Administrative Region.

E-mail addresses: Gianni.Panagiotou@hki-jena.de (G. Panagiotou), amxu@hku.hk (A. Xu).

<sup>1</sup> These authors contributed equally.

<sup>2</sup> Lead contact.

#### https://doi.org/10.1016/j.csbj.2024.01.007

Received 28 August 2023; Received in revised form 8 January 2024; Accepted 14 January 2024 Available online 20 January 2024

2001-0370/© 2024 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

*Abbreviations:* NAFLD, non-alcoholic fatty liver disease; SS, simple steatosis; NASH, non-alcoholic steatohepatitis; CT, computed tomography; MRI, magnetic resonance imaging; MAFLD, metabolic associated fatty liver disease; MASLD, metabolic dysfunction associated steatotic liver disease; AST, aspartate aminotransferase; ALT, alanine aminotransferase; CRN, Clinical Research Network; FLIP, Fatty Liver Inhibition of Progression; UHPLC, ultra-high performance liquid chromatography; SD, standard deviation; BMI, Body Mass Index; KEGG, Kyoto Encyclopedia of Genes and Genomes; TOM, topological overlap measure; AUROC, area under the receiver operating characteristic; HOMA-IR, Homeostatic Model Assessment for Insulin Resistance; APOA, Apolipoprotein A; APOB, Apolipoprotein B; SBP, systolic blood pressure; CRP, C-Reactive Protein; HbA1c, haemoglobin A1c; LC-MS, liquid chromatography-mass spectrometry; SM, Sphingomyelin; PAF, Platelet-activating Factor; PC, Phosphatidylcholine; PG, Phosphatidylglycerol; TG, Triglyceride; PE, Phosphatidylethanolamine; DBP, diastolic blood pressure; TP, total protein; AUC, area under the curve; VAT, visceral adipose tissue; PIP2, phosphatidylinositol 4,5-bisphosphate; PIP3, phosphatidylinositol 3,4,5-trisphosphate; PI3K, phosphatidylinositol-3 kinase; Akt, Protein kinase B; EPA, omega-3 eicosapentaenoic acid; LPCAT, lysophosphatidylcholine acyltransferase; BCAA, branched-chain amino acid; NAT, N-acyl taurine; GRP84, immunomodulating receptor G protein-coupled 84; SFA, saturated fatty acid.

*Conclusion:* Different histological stages of NAFLD are enriched with distinct sets of metabolites, lipids, and metabolic pathways. Integrated algorithms highlight the important metabolic and lipidomic features for diagnosis and staging of NAFLD in obese individuals.

#### 1. Introduction

Non-alcoholic fatty liver disease (NAFLD) is the most common chronic liver disease globally, affecting 25–33% of the world's population [1]. NAFLD encompasses a wide histological spectrum ranging from simple steatosis (SS), which often has a benign course, to non-alcoholic steatohepatitis (NASH), the advanced form with a much greater risk of progression into fatal hepatic complications such as cirrhosis and hepatocellular carcinoma [2]. Furthermore, NASH is a leading cause of liver transplantation and is closely associated with a high incidence of cardiovascular mortality independent of traditional risk factors [3]. Despite the rapidly increasing prevalence of NAFLD and NASH, there is currently no approved pharmacotherapy for this disease. Early detection of high-risk individuals, followed by lifestyle interventions, remains the most effective strategy.

Various imaging techniques, including ultrasonography, computed tomography (CT), and magnetic resonance imaging (MRI), are currently used to measure hepatic fat content. However, these imaging methods cannot differentiate SS from NASH, which is typically characterized by not only steatosis but also hepatocyte ballooning and lobular inflammation [2]. Currently, both the diagnosis and staging of NASH rely heavily on the histological evaluation of liver biopsies, a protocol that has many pitfalls, including sampling error, high cost, invasiveness, risk of complications, and inter-observer variation among pathologists. [4] Although several circulating biomarkers have been identified as potential tools for the non-invasive diagnosis and/or prognosis of NASH, none of them have been clinically implemented at present.

Metabolic disturbance is a primary driver for the onset and progression of NAFLD. NAFLD often co-occurs with obesity and other cardiometabolic disorders such as insulin resistance, type 2 diabetes, dyslipidemia, hypertension, and atherosclerosis [5]. Recognizing its close association with metabolic disorders, international experts have suggested renaming it to metabolic associated fatty liver disease (MAFLD) [6], and more recently, to metabolic dysfunction associated steatotic liver disease (MASLD) [7]. A large number of both targeted and untargeted metabolomic studies have been performed to identify metabolic biomarkers associated with the pathophysiology of NAFLD and NASH [8-11]. Distinct alterations in amino acids, glutathione metabolism, and various lipid species in NAFLD and NASH have been reported in different clinical cohorts [12]. Although these studies provided insightful evidence of dysregulated metabolic pathways in NAFLD and NASH, there are large variations and even inconsistencies due to limited sample size, differences in the study population, and varying diagnostic criteria. Few studies have been performed in the Asian population with liver biopsy-confirmed NAFLD and NASH.

In this study, we aimed to leverage our unique clinical cohort of obese Chinese, which included the entire disease spectrum from normal liver to SS to borderline NASH to NASH, to systematically search for metabolites and lipid species associated with different stages of the disease and to explore the possibility of using metabolites and lipid biomarkers for predicting the risk of NAFLD and NASH in the obese population.

#### 2. Experimental procedures

#### 2.1. Study population

Five hundred and nine patients with severe obesity were recruited at Huaqiao Hospital, affiliated with Jinan University, Guangzhou, China, during the period from January 6, 2017, to October 23, 2020. They were then further screened for eligibility (see Supplementary methods). A total of 250 subjects spanning the full spectrum of NAFLD were finally enrolled and subjected to untargeted metabolomic and lipidomic profiling. Demographic, clinical, and laboratory data (after an overnight fast of at least 10 h) for each patient were collected within one week prior to surgery, and liver biopsies were obtained from the middle of the right lobes during the bariatric surgery.

#### 2.2. Histological assessment of liver biopsies

Thin sections of liver biopsy specimens were stained with H&E or Masson trichrome and were evaluated by three independent liver pathologists as described previously (Qiu H, CMGH, 2022). The scores of steatosis, hepatocellular ballooning, lobular inflammation, and fibrosis were evaluated according to the NASH Clinical Research Network scoring system [13]. The definitions of histologically normal liver, SS, borderline NASH, and NASH were based on the Fatty Liver Inhibition of Progression (FLIP) algorithm [14].

Further experimental procedures (e.g., metabolomics, lipidomics, and statistical analysis) are provided in the supplementary methods.

#### 3. Results

# 3.1. Distinct clinical characteristics in subjects with normal liver, SS, and NASH

The clinical characteristics of the study cohort (n = 250) are shown in Table 1, including 44 obese individuals with normal livers and 196 NAFLD patients (55 SS, 75 NASH, and 76 borderline NASH). There were no significant differences in the prevalence of different NAFLD stages between males and females (p > 0.05, Chi-square test). We identified 17 out of 20 metabolic-related parameters that were different between normal liver and NAFLD (p < 0.05, Student t-test). As expected, the levels of body mass index (BMI), alanine transaminase (ALT), aspartate aminotransferase (AST), insulin, and Homeostatic Model Assessment for Insulin Resistance (HOMA-IR) increased progressively from normal liver to SS and to NASH. Another seven parameters, including Apolipoprotein A (APOA), Apolipoprotein B (APOB), systolic blood pressure (SBP), C-Reactive Protein (CRP), uric acid, amylase, and haemoglobin A1C (HbA1c), were higher in SS compared to normal liver, whereas the differences between the SS and NASH groups were not significant. Furthermore, high-density lipoprotein (HDL) was higher in SS compared to NASH, and the level of total protein was highest in SS compared to both normal liver and NASH.

# 3.2. Differentially changed metabolites and lipid species across different histological stages of NAFLD

In total, 263 metabolites and 550 lipid species from 12 lipid classes were detected in serum samples. Comparison between normal liver and SS showed 12 metabolites and 63 lipid species enriched in normal liver, while 9 metabolites and 83 lipid species were elevated in SS (p < 0.05, MaAsLin2, Fig. 1A & Table S1). Meanwhile, a comparison between SS and NASH identified 17 metabolites and 78 lipid species enriched in SS and 13 metabolites and 5 lipid species increased in NASH patients, respectively (p < 0.05, MaAsLin2, Fig. 1B & Suppl. Table 1). Further analysis across the entire NAFLD spectrum demonstrated several lipid species and metabolites were progressively altered across disease severity (p < 0.05, ordinal regression, Suppl. Table 2). Phosphatidic acid (PA), 12 phosphatidylcholine (PC), 8 phosphatidylethanolamine (PE), 5 phosphatidylglycerol (PG), 3 phosphatidylinositol (PI), 6 sphingomyelin (SM), and tridecylic acid displayed stepwise decreases from normal liver to SS to NASH (Fig. 2A). In contrast, 10 triglyceride (TG) species, 1 PE, xanthine, L-valine, creatine, and N-palmitoyltaurine exhibited progressive elevations from normal liver to SS to NASH (Fig. 2B). Interestingly, L-cystine, a precursor for the synthesis of glutathione, an important antioxidant, showed a significant decrease in SS compared to both normal liver and NASH (Fig. 2C), while another 2 PC and 1 PG were upregulated in SS (Fig. 2D).

# 3.3. Metabolomic/lipidomic pathway enrichment analysis for different stages of NAFLD

To delineate the potential metabolic mechanisms underlying the onset and/or progression of NAFLD, we performed pathway enrichment analysis using the significantly altered metabolites and lipid species in different disease stages. This analysis found that the pathway related to arginine biosynthesis was enriched in normal liver and SS, while pathways for D-arginine and D-ornithine metabolism were enriched in SS compared to NASH (Fig. 3 & Table 2). In contrast, the sphingolipid signaling pathway, sphingolipid metabolism, and necroptosis were enriched in SS compared to normal liver, suggesting the potential involvement of these pathways in triggering the initiation of NAFLD.

#### Table 1

#### Clinical characteristics in different stages.

Notably, the metabolism of purine encoded by xanthine and allantoin and the GnRH signaling pathway encoded by PA were uniquely enriched in NASH compared to SS. Retrograde endocannabinoid signaling, the phosphatidylinositol signaling system, and biosynthesis of unsaturated fatty acids were enriched in SS compared to NASH. Importantly, valine, leucine, and isoleucine biosynthetic pathways and pantothenate and CoA biosynthesis pathways were progressively enriched from normal liver to SS to NASH, indicating that increased branched-chain amino acids and pantothenic acid are key contributors to the pathogenesis of NAFLD.

# 3.4. Co-expression network analysis of lipid species in relation to NAFLD and its associated clinical parameters

In light of the complexity of dynamic changes in lipid species and metabolites during NAFLD progression, we created co-expression networks for metabolites and lipid species, respectively. Eight cooccurrence modules represented in different colors were constructed successfully among 481 lipid species (Fig. 4A). The remaining 78 lipid species failed to cluster and were marked in grey; they were not included for further analysis. The 8 modules were named according to the assigned color from module blue to module yellow. The number of lipid species among the 12 lipid categories in each module was included in

Characteristics	<b>Total Population</b>	Normal	NAFLD		P value			
			Total	SS	NASH	Normal vs NAFLD	Normal vs SS	SS vs NASH
Male (%)	107 (41.3%)	12 (27.2%)	95 (44.2%)	27 (48.2%)	31 (39.2%)	0.056	0.054	0.389
APOB (g/L)	$1.04 \pm 0.23$	$0.92 \pm 0.2$	$1.06 \pm 0.22$	$1.01 \pm 0.21$	$1.08 \pm 0.23$	< 0.001	0.034	0.1
Age (years)	30 (23, 36)	30 (23, 39)	30 (24.5, 30)	30.5 (25.25, 35.25)	30 (23, 35)	0.418	0.034	0.089
BMI (kg/m2)	38.33 (33.58,	34.15 (31.8, 38.5)	39.27 (34.71,	37.95 (33.1,	41.2 (36.25,	< 0.001	0.013	0.008
	43.62)		44.86)	41.82)	46.96)			
Systolic BP (MmHg)	126 (118.5, 137)	120 (113.75, 128 25)	127 (120, 138)	128 (119, 139)	125 (119.5, 134)	< 0.001	0.004	0.323
Diastolic BP (MmHg)	80 (72, 89)	75.18 (71.75,	81 (74, 90)	80 (72, 87)	82 (74, 89)	0.012	0.22	0.257
-		84.25)						
CRP (mg/L)	5.68 (3.01, 9.01)	3.72 (1.35, 6.89)	5.95 (3.48,	5.28 (3.03,	7.25 (3.86,	< 0.001	0.025	0.069
			10.8)	8.61)	11.95)			
Uric Acid (µmol/L)	453.2 (376.05,	402.02 (357.48,	462 (382.3,	460.35 (385,	465 (379.65,	0.003	0.041	0.527
Amylase (II/L)	341.2) 41 (32 7 51 2)	475.2)	397 (31 85	40 (36 22	40 (32 15 46 5)	< 0.001	0.011	0 385
1111 Jube (0, 2)	11 (0217, 0112)	1510 (05) 01120)	48.6)	48.75)	10 (02110) 1010)	0.001	01011	0.000
ALT (U/L)	43 (24, 68)	22.7 (17, 30.25)	50 (29.5, 71)	39 (23, 53.75)	55 (35, 74)	< 0.001	< 0.001	0.004
AST (U/L)	25 (19, 37)	20 (16, 22.25)	28 (20, 39.5)	23 (18.75, 30.25)	32 (22, 46)	< 0.001	0.004	0.002
Total_Protein (g/L)	72.36 (67.8,	71.2 (66.8, 73.82)	72.6 (68.1,	73.65 (70.06,	71.1 (67.65,	0.035	0.003	0.043
	75.45)		76.05)	76.65)	75.45)			
Total Cholesterol (mmol/L)	4.93 (4.5, 5.5)	4.89 (4.32, 5.36)	4.95 (4.51, 5.52)	4.9 (4.4, 5.48)	5.03 (4.58, 5.46)	0.242	0.567	0.453
Total Triglycerides	1.78 (1.32, 2.47)	1.34 (1.08, 1.91)	1.82 (1.39,	1.69 (1.26,	1.81 (1.41,	0.002	0.114	0.235
(mmol/L)			2.65)	2.18)	2.54)			
HDL (mmol/L)	0.98 (0.86, 1.15)	1.04 (0.95, 1.24)	0.97 (0.84, 1.1)	1 (0.92, 1.18)	0.91 (0.82, 1.08)	0.001	0.134	0.021
LDL (mmol/L)	3.01 (2.66, 3.46)	2.77 (2.48, 3.3)	3.04 (2.7, 3.47)	2.97 (2.65, 3.43)	3.1 (2.7, 3.52)	0.038	0.29	0.391
APOA (g/L)	1.25 (1.12, 1.37)	1.32 (1.21, 1.5)	1.23 (1.1, 1.35)	1.22 (1.03, 1.35)	1.23 (1.12, 1.33)	0.013	0.031	0.815
Insulin (mIU/L)	19.89 (13.3, 29.15)	12.5 (9.02, 18.08)	21.41 (15.02, 30.11)	18.59 (13.77, 24.17)	25.66 (18.53, 33.68)	< 0.001	0.005	0.001
HbAIc (%)	5.8 (5.4, 6.45)	5.4 (5.2, 5.9)	5.8 (5.5, 6.67)	5.8 (5.47, 6.23)	5.87 (5.6, 6.67)	< 0.001	0.006	0.241
Fasting Glucose (mmol/	5.47 (4.94, 6.56)	5.15 (4.88, 5.62)	5.52 (4.96,	5.42 (4.96,	5.66 (5.22,	0.007	0.114	0.089
L)			6.72)	6.32)	6.66)			
HOMA-IR	5.25 (3.29, 8.25)	3.09 (1.93, 5.31)	5.65 (3.82, 8.58)	4.48 (3.05, 6.57)	6.65 (4.68, 10.37)	< 0.001	0.008	< 0.001

Data are presented as mean  $\pm$  SD or median (interquartile range)

APOB, Apolipoprotein B; BMI, Body mass index; BP, Blood pressure; CRP, C-reactive protein; ALT, Alanine aminotransferase; AST, Aspartate aminotransferase; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; APOA, Apolipoprotein A; HbA1c, Haemoglobin A1C; HOMA-IR, Homeostatic Model Assessment for Insulin Resistance.

#### contribute to NAFLD development.

The correlation between modules and clinical parameters showed that module blue, mainly consisting of PC and SM species, and module pink, with PG species, were negatively correlated with the occurrence of NAFLD. These two modules also exhibited negative correlations with different metabolic risk factors for NAFLD, including APOB, ALT, AST, TG, Insulin, HbA1c, fasting glucose, and HOMA-IR, and positive correlations with amylase, suggesting that these two lipid clusters might be protective against NAFLD. Meanwhile, module red (mainly TG species) and module turquoise (mainly PC, PE, and PG species) were positively associated with the occurrence of NAFLD. Module red exhibited positive correlations with DBP, TG, insulin, HbA1, fasting glucose, and HOMA-IR, while module turquoise displayed positive associations with SBP, DBP, insulin, and HOMA-IR, suggesting that these two lipid clusters may The same correlation analysis was performed using the metabolite profile (Fig. S2). However, only 113 out of 263 metabolites (~43.0%) can be grouped into two clusters. Module blue, containing 48 metabolites, showed negative correlations with ALT, APOA, SBP, AST, and insulin and a positive correlation with amylase (Table S5). By performing a pathway analysis with these 48 metabolites, the metabolism of glycine, serine, and threonine was significantly enriched. Further investigation showed that the abundance of these three metabolites was negatively correlated with both ALT and AST, suggesting the metabolism of NAFLD by suppressing the level of ALT and AST.



**Fig. 1.** Significantly changed metabolites and lipid species from Normal to SS and from SS to NASH. The volcano plot shows the significant metabolites and lipid species in the comparisons between normal liver and simple steatosis (A) and between simple steatosis and NASH (B). Variables with adjusted q value higher than 0.25 are labelled in text. The two horizontal dotted lines represent the two cut-offs for significance (p = 0.05, and q = 0.25). Variables coloured in dark green (q < 0.25) and light green (p < 0.05) represent the enrichment in normal liver (A) and simple steatosis (B), respectively. Variables coloured in red (q < 0.25) and light yellow (p < 0.05) represent the enrichment in simple steatosis (A) and NASH (B), respectively.

# 3.5. Integrated machine learning (ML) model for identifying important metabolites and lipids features in NAFLD and NASH

To further investigate the roles of circulating metabolites and lipid species in the onset and progression of NAFLD, we next attempted to construct classification models for differentiating NAFLD from normal liver and NASH from SS, respectively, using clinical parameters, metabolites, and lipids species. We used a stepwise model to select the best clinical variables for these two purposes (Suppl. Table 3). For the comparisons between normal liver and NAFLD, ALT, APOA, APOB, direct bilirubin, and amylase were selected, while total protein (TP), AST, ALT, and insulin were identified for differentiating NASH from SS. Furthermore, 22 metabolites and 27 lipid species were selected for distinguishing NAFLD from normal liver, while 10 metabolites and 3 lipid species were identified for differentiating NASH from SS using Boruta analysis (Table S4).

By using the selected clinical features, metabolites, and lipid species, we constructed different logistic regression models for classifying differences between NAFLD and normal liver and between NASH and SS, respectively. DeLong tests were implemented to compare the area under the curve (AUC) among models using different panels. For the classification between normal liver and NAFLD (Fig. 5A), the model integrating the five routine clinical parameters yielded the AUC of 0.882, which increased significantly to 0.940 and 0.954 by addition of the selected metabolites and lipid species, respectively (p < 0.05, DeLong test). Furthermore, a model integrating the selected metabolites, lipid species, and clinical parameters achieved an 11.6% improvement in the AUC compared to the clinical model (p < 0.05, DeLong test). For stratification between NASH and SS, the model built using the four clinical parameters (TP, AST, ALT, and insulin) yielded an AUC of 0.785, which was further improved to 0.868 and 0.893 by adding the selected metabolites and lipid species, respectively (p < 0.05, DeLong test, Fig. 5B). When adding both the selected metabolites and lipid species in combination, the AUC of the final model significantly improves to 0.915.

#### 4. Discussion

This study revealed the lipidomic and metabolomic biomarkers associated with the onset and progression of NAFLD in a cohort of 250 obese Chinese individuals with the full histological spectrum of NAFLD. The main findings include: (i) the identification of previously overlooked bioactive lipids such as N-palmitoyltaurine, which increase progressively with NAFLD severity, and (ii) a distinct correlation of different circulating TG and PC species with disease severity as determined by co-expression network analyses: saturated and monounsaturated TGs correlate positively with NAFLD, whereas a negative association links PC plasmalogens to NAFLD. Furthermore, we have built and compared multiple ML models classifying differences between NAFLD and normal liver and between NASH and SS to identify important metabolic and lipidomic features associated with significant improvement in disease prediction compared to conventional clinical parameters.

In the last decade, the lipidomic and metabolomic characterization of histological stages of liver biopsy-proven NAFLD was accomplished mainly in clinical cohorts of Caucasian ethnicity [15]. However, extrapolation to non-Caucasian individuals is questionable as an ethnic-specific metabolic phenotype is generally expected. In this regard, it is well established that ethnicity intrinsically influences body fat content and distribution [16]. Herein, Asians exhibit a greater amount of body fat and visceral adipose tissue (VAT) than Caucasians at the same BMI [16]. In turn, VAT is an independent predictor of liver inflammation and fibrosis, leaving definite traces on the serum lipidome [17]. The potentially distinct metabolic phenotype of the Chinese population, along with the projected increase in the prevalence of NASH in China, [18] highlights the need for a large cohort study assessing the lipidomic and metabolomic profiles of well-characterized NAFLD Chinese patients. Thus far, scarce evidence is available in this population, with only a few small cohorts in which serum metabolites and circulating bile acids have been described [19].

In the present study, widespread lipidomic between-group differences were detected in NAFLD Chinese patients. Lipidomic changes, which included increases in the majority of TG species, were of greater magnitude in individuals with normal liver versus patients with SS than between the former and NASH patients. This is consistent with recently reported changes in liver and plasma lipidomes at the early phase of NAFLD, which were relatively stable in the transition from SS to NASH [12]. Moreover, higher phosphatidylinositol 4,5-bisphosphate (PIP2) was observed in SS patients compared to controls. PIP2 is the precursor of phosphatidylinositol 3,4,5-trisphosphate (PIP3), the second messenger in the phosphatidylinositol-3 kinase (PI3K)/protein kinase B (Akt) pathway. The increased level of PIP2 contributed to the lower abundance of PIP3, which attenuates PI3K signaling and thus induces insulin resistance, an important driving factor for NAFLD development [20]. Unexpectedly, omega-3 eicosapentaenoic acid (EPA) was elevated in SS patients relative to normal liver individuals. Interestingly, supplementation with EPA has been associated with decreased accumulation of hepatic TG via increasing hepatic  $\beta$ -oxidation and reducing de novo lipogenesis [21]. Therefore, elevated EPA may represent a compensatory response to defend the early progression from normal liver to SS.

Another notable observation is the significantly outnumbered changes in non-lipid metabolites between SS and NASH compared to the predominant alternations in lipid species between normal liver and SS. Consistently, a recent study in the Caucasian population found that hepatic lipidomic remodeling occurs predominantly in obese individuals with steatosis but not with non-alcoholic steatohepatitis. These findings

#### Table 2

Detailed information of the enriched pathways at each stage. The enriched pathways and the corresponding.

encoding metabolites are presented.					
Pathway	Encoded_metabolites	P_value	Enrichment	Туре	Comparisons
Arginine biosynthesis	Citrulline	0.044	Normal	Metabolites	Normal_vs_SS
Sphingolipid signaling pathway	Cer, SM	0.001	SS	Lipids	Normal_vs_SS
Necroptosis	Cer, SM	0.004	SS	Lipids	Normal_vs_SS
Sphingolipid metabolism	Cer, SM	0.014	SS	Lipids	Normal_vs_SS
Valine, leucine and isolueucine biosynthesis	L-Valine	0.021	SS	Metabolites	Normal_vs_SS
Pantothenate and CoA biosynthesis	L-Valine	0.048	SS	Metabolites	Normal_vs_SS
Pantothenate and CoA biosynthesis	Pantothenic acid, L-Valine	0.002	NASH	Metabolites	SS_vs_NASH
Purine metabolism	Xanthine, Allantoin	0.024	NASH	Metabolites	SS_vs_NASH
Valine, leucine and isoleucine biosynthesis	L-Valine	0.031	NASH	Metabolites	SS_vs_NASH
GnRH signaling pathway	PA	0.033	NASH	Lipids	SS_vs_NASH
D-Arginine and D-ornithine metabolism	D-Ornithine	0.008	SS	Metabolites	SS_vs_NASH
Retrograde endocannabinoid signaling	PC, PE	0.014	SS	Lipids	SS_vs_NASH
Phosphatidylinositol signaling system	PA, PI	0.027	SS	Lipids	SS_vs_NASH
Biosynthesis of unsaturated fatty acids	Eicosapentaenoic acid	0.049	SS	Metabolites	SS_vs_NASH



Fig. 2. Comparisons of the significant metabolites and lipid species among different NAFLD stages. The bar chart shows the significant metabolites and lipids that are progressively decreased from normal liver to simple steatosis to NASH (A), that are progressively increased from normal liver to simple steatosis to NASH (B), that are decreased in simple steatosis compared to normal liver and NASH (C), and that are enriched in simple steatosis compared to normal liver and NASH (D).



**Fig. 3.** Pathway enrichment analysis for different histological stages of NAFLD using the significant metabolites and lipid species. The bar charts show the enriched pathways in each NAFLD stage with the significant metabolites and lipid species in the comparisons between normal and simple steatosis (A) and between simple steatosis and NASH (B). The length of the bar represents the p value of each pathway. The colour of the bar represents the type of pathway. The stages of enrichment are labelled in text.

suggest that the dynamics of metabolic alteration starts with lipidomic species during the progression from normal liver to SS but shifts to other metabolites in the later stages of NAFLD development [12]. Specifically, our results found significantly higher levels of L-cystine and citrate in NASH than in SS patients. In this respect, oxidative stress is a well-established pathogenic insult that drives the progression of NAFLD to NASH [22]. Increased L-cystine may reflect a compensatory mechanism to counteract the production of ROS by stimulating the synthesis of

glutathione. In contrast, citrate contributes to the formation of hydrogen peroxide in the presence of iron [23], suggesting that pro-oxidant and antioxidant mechanisms may be activated in the progression to NASH. Futhermore, 1-stearoyl-sn-glycerol 3-phosphocholine (18:0-LPC) and 1-stearoyl-sn-glycerol were decreased in NASH compared to SS. In line with our findings in patients, a previous study found decreased levels of 18:0-LPC in mice with NASH, resulting from hepatic up-regulation of lysophosphatidylcholine acyltransferase (LPCAT)1–4 [24]. Meanwhile,





**Fig. 4.** The co-expression clusters of lipid species and their correlation with NAFLD-related parameters. (A) the cluster dendrogram shows the constructed modules in different colours. The module coloured in grey represents the lipid species that failed to be grouped. (B) The heatmap on the left shows the correlation results between the eigenvalues of each module and the significant clinical variables. The rho value and the p value are labelled in each cell. The colour of each cell indicates the rho value. The significance of each parameter in the two comparisons is labelled with a star in the upper cell. The heatmap on the right shows the number of lipid species in each category for the corresponding modules.

we observed a significant alteration in 1-stearoyl-sn-glycerol, the precursor of 18:0-LPC, suggesting a possible involvement of this pathway in NASH development.

Our pathway enrichment analysis found that branched-chain amino acid (BCAA) signaling pathways were progressively upregulated with increasing disease severity, mainly due to the increased levels of Lvaline. Consistently, plasma levels of BCAAs (e.g., isoleucine and valine) were shown to be closely associated with the degree of liver inflammation and ballooning in NAFLD/NASH patients [25]. Increased circulating BCAAs in NAFLD have been attributed to excessive BCAA intake, impairment in BCAA catabolism, as well as changes in the composition and function of the gut microbiota. [26] Of particular interest is the stepwise increase in N-palmitoyltaurine, a bioactive lipid that belongs to the family of N-acyl taurine (NAT). A previous study reported that the members of this bioactive lipid family regulate insulin secretion in pancreatic  $\beta$ -cells [27]. Yet, there is only one study that linked N-palmitoyltaurine to asthma severity [28]. Further investigations are justified to address the potential role of N-palmitoyltaurine in the pathogenesis of NAFLD in both animals and humans. Moreover, we identified tridecylic acid, a medium-chain fatty acid, was progressively decreased from normal liver to SS to NASH. Tridecylic acid has been shown to activate the immunomodulating receptor G protein-coupled 84 (GPR84) to suppress lipotoxicity-induced hepatic fibrosis [29] Therefore, further studies are warranted to explore whether supplementation with tridecylic acid ameliorates NAFLD development.

Although the accumulation of TG in hepatocytes is generally regarded as the main hallmark of NAFLD, our results from co-expression network analyses demonstrated substantial heterogeneity among circulating TG species. Notably, TGs in the red module contained saturated and monounsaturated fatty acids and were positively associated with NAFLD, whereas TGs in the black module included mostly polyunsaturated fatty acids and were not associated with NAFLD (Table 3). Saturated and monounsaturated TGs in serum may to some extent mirror the hepatic triglyceride content, since TGs containing saturated fatty acids are elevated in the liver of NAFLD patients mainly due to hepatic de novo lipogenesis [30]. Similarly, we observed two different clusters of PC species: PCs in module blue, mainly PC plasmalogens, were negatively associated with NAFLD, whereas PCs without these moieties were positively correlated with the progression of NAFLD (Table 3). Of note, circulating levels of plasmalogens decreased in a stepwise manner in a cohort of biopsy-proven NAFLD patients [31]. Furthermore, administration of a plasmalogen precursor prevented SS and NASH in mice [32], suggesting that certain PC species may be protective against the development of NAFLD.

Currently, the diagnosis and staging of NAFLD rely heavily on the histological evaluation of liver biopsies, which is invasive, subjective,

#### Table 3

TG and PC species enriched in different modules with different contribution to NAFLD progression.

TG species		PC species				
Red Module	Black Mdule	Blue Module	Turqoise Module			
TG	TG	PC(15:0/20:4	PC(10:0/26:2(5E,9Z))			
(14:0/	(16:0/	(5E,8E,11E,14E))				
16:0/	18:2/					
TG	22.0) TG	PC(17:0/0:0)	PC(14:0/0:0)			
(14:0/	(16:0/					
16:0/	20:4/					
18:1) TG	22:6) TG	PC(0-16·1(77)/0·0)	PC(16.0.0.0)			
(14:0/	(16:0/	16(0-10.1(72)/0.0)	16(10.0/0.0)			
16:1/	22:6/					
18:1)	22:6)	DC(0.17.0.(0.0)	DO(1(-1(77)(0.0)			
(16:0/	(16:1/	PC(0-17:0/0:0)	PC(16:1(7Z)/0:0)			
16:0/	18:2/					
16:0)	18:3)					
TG (16:07	TG (16:1/	PC(O-18:0/0:0)	PC(18:0/0:0)			
16:0/	18:2/					
18:1)	22:6)					
TG	TG	PC(O-24:0/0:0)	PC(18:0/18:1(11E))			
(16:0/ 16:0/	(16:1/					
18:2)	22:6)					
TG	TG	PC(P-16:0/13:0)	PC(18:1(11E)/0:0)			
(16:0/	(17:0/					
16:1/	22:6/					
TG	TG	PC(P-16:0/14:0)	PC(18:1(11Z)/16:1(7Z))			
(16:0/	(18:0/					
16:1/	22:5/					
18:2) TG	22:0) TG	PC(P-16:0/16:0)	PC(18:3(6Z.9Z.12Z)/0:0)			
(16:0/	(18:1/		( (,,,,),)			
18:0/	18:2/					
16:0) TG	22:6) TG	PC(P-16-0/18-2(2F 4F))	DC(18·3(97 127 157)/18·1			
(16:0/	(18:1/	16(1-10.0/10.2(22,42))	(13Z))			
18:0/	20:4/					
18:0)	22:6)	DC(D 1(.0 /00.0	DC(00-0 (0-0)			
(16:0/	(18:1/	(57.87.117))	PG(20:0/0:0)			
18:0/	22:5/	(,,))				
18:1)	22:6)					
TG (16:0/	TG (18·17	PC(P-16:0/20:4 (5F 8F 11F 14F))	PC(20:1(11E)/0:0)			
18:0/	22:6/	(JE, 0E, 11E, 14E))				
22:0)	22:6)					
TG	TG	PC(P-16:0/22:5	PC(22:0/0:0)			
(16:0/	(18:2/	(42,72,102,132,162))				
20:4)	18:3)					
TG	TG	PC(P-16:0/23:0)	PC(22:6			
(16:0/	(18:2/		(4Z,7Z,10Z,13Z,16Z,19Z)/			
18:2/	22:6)		10.2(2E,4E))			
TG	TG	PC(P-16:0/24:0)	PC(7:0/18:2(9Z,12Z))			
(16:0/	(18:2/					
18:2/ 20:4)	18:3/ 22:6)					
TG	TG	PC(P-16:0/26:1(5Z))	PC(O-17:1(9Z)/0:0)			
(16:1/	(18:2/					
16:1/	22:6/					
18:2) TG	22:0) TG	PC(P-18:0/16:0)	PC(O-18:1(11E)/0:0)			
(16:1/	(18:3/					
17:1/	20:5/					
18:2) TG	22:6) TG	DC(D-18.0/18.1(11E))	PC(0-20.070.00			
(17:0/	(20:4/	1 G(1-10.0/ 10.1(11EJJ	1 0(0-20.0/ 0.0)			

TG species		PC species		
Red Module	Black Mdule	Blue Module	Turqoise Module	
18:1/	20:5/			
18:2)	22:6)			
TG	TG	PC(P-18:0/18:2(2E,4E))	PC(O-22:0/0:0)	
(18:0/	(20:4/			
18:1/	22:6/			
20:4)	22:6)			
TG	TG	PC(P-18:0/20:3	PC(O-22:1(13Z)/0:0)	
(18:1/	(20:5/	(5Z,8Z,11Z))		
18:1/	22:6/			
18:3)	22:6)			
	TG	PC(P-18:0/22:2(13Z,16Z))	PC(O-24:1(15Z)/0:0)	
	(22:5/			
	22:6/			
	22:6)			
		PC(P-18:0/22:5	PC(P-16:0/2:0)	
		(4Z,7Z,10Z,13Z,16Z))		
		PC(P-18:0/22:6	PC(P-18:0/26:2(5E,9Z))	
		(4Z,7Z,10Z,13Z,16Z,19Z))		
		PC(P-20:0/20:3	PC(P-20:0/17:2(9Z,12Z))	
		(5Z,8Z,11Z))		
		PC(P-20:0/20:4		
		(5E,8E,11E,14E))		
		PC(P-20:0/22:4		
		(7Z,10Z,13Z,16Z))		
		PC(P-20:0/22:5		
		(4Z,7Z,10Z,13Z,16Z))		
		PC(P-20:0/22:6		
		(4Z,7Z,10Z,13Z,16Z,19Z))		
		PC(P-20:0/24:4		
		(5Z,8Z,11Z,14Z))		

and expensive [4]. There is an unmet need to discover novel biomarkers to distinguish benign SS from NASH, an advanced form of NAFLD with a much higher risk of developing cirrhosis and hepatocarcinoma. In this study, we developed and compared multiple ML models classifying NAFLD versus normal liver and NASH versus SS by integrating common clinical parameters with important metabolomic and lipidomic biomarkers. Notably, a saturated TG (16:0/16:0/18:1) was the top informative feature to discriminate NAFLD patients from individuals with normal liver. A previous study has observed a close association between circulating levels of TG (16:0/18:0/18:1) and liver fat content in NAFLD patients assessed by proton magnetic resonance spectroscopy [33]. Interestingly, among the ten TG species retrieved from our prediction models, seven of them were also selected by Mayo, et al. for the construction of their model to diagnose NASH [11]. Furthermore, the odd-chain saturated fatty acid (SFA) heptadecanoic acid (17:0) was the strongest factor in discerning NASH from SS. Odd-chain SFAs partly reflect dairy fat intake, and their levels were inversely correlated with liver function enzymes as well as T2D [34]. Further studies are warranted to investigate the role of SFAs in NAFLD development and to validate the clinical value of metabolite/lipid species-based ML models as a non-invasive diagnostic tool for NAFLD and NASH in independent cohorts with larger sample sizes.

#### **Statements & Declarations**

none.

#### Ethics approval and consent to participate

All participants had given written informed consent. Ethics approval was obtained from the Institutional Review Board of the University of Hong Kong/Hospital Authority, Hong Kong West Cluster.



**Fig. 5.** Establishment of machine learning models for risk prediction of NAFLD and NASH in obese individuals. (A) Receiver operating characteristic (ROC) curves for classifying normal liver and NAFLD using 5 clinical parameters (ALT, APOA, APOB, direct Bilirubin, and amylase), 5 clinical parameters + selected metabolites, 5 clinical parameters + selected lipid species, and 5 clinical parameters + selected metabolites and lipid species, respectively. (B) ROC curves for the differentiation of NASH from simple steatosis using 4 clinical parameters (total protein, AST, ALT, and insulin), 4 clinical parameters + selected metabolites or lipid species, and 4 clinical parameters + selected metabolites, respectively.

### CRediT authorship contribution statement

Jiarui Chen and Ronald Siyi Lu researched the data. Jiarui Chen, Ronald Siyi Lu, and Candela Diaz-Canestro wrote the manuscript. Jiarui Chen performed statistical analyses. Aimin Xu critically reviewed and edited the manuscript. Aimin Xu, Gianni Panagiotou, Erfei Song, Xi Jia, Yan Liu, Cunchuan Wang, and Cynthia K.Y. Cheung initiated and supervised the study. All authors reviewed the manuscript. Aimin Xu and Gianni Panagiotou are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

### Availability of data and materials

The datasets generated during and/or analysed during the current study are available from the corresponding author upon reasonable request.

### Funding

This work was supported by the Merck Investigator Studies Program (MSD MISP 59615), Hong Kong Research Grants Council/Area of Excellence (AoE/M/707-18), Germany's Excellence Strategy – EXC 2051 – Project-ID 390713860, and the BMBF-funded "PerMiCCion" project (Project ID 01KD2101D).

### Author Statement

During the preparation of this work, the author(s) used no generative AI and AI-assisted technologies.

## **Declaration of Competing Interest**

None.

### Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.01.007.

#### References

- Quek J, et al. Global prevalence of non-alcoholic fatty liver disease and nonalcoholic steatohepatitis in the overweight and obese population: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol 2022.
- [2] Chalasani N, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology 2018;67:328–57.
- [3] Starley BQ, Calcagno CJ, Harrison SA. Nonalcoholic fatty liver disease and hepatocellular carcinoma: a weighty connection. Hepatology 2010;51:1820–32.
- [4] Sumida Y, Nakajima A, Itoh Y. Limitations of liver biopsy and non-invasive diagnostic tests for the diagnosis of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. World J Gastroenterol: WJG 2014;20:475.
- [5] Arab JP, Arrese M, Trauner M. Recent insights into the pathogenesis of nonalcoholic fatty liver disease. Annu Rev Pathol 2018;13:321–50.
- [6] Eslam M, et al. A new definition for metabolic dysfunction-associated fatty liver disease: An international expert consensus statement. J Hepatol 2020;73:202-9.
  [2] Diselb ME et al. A multi-associated biological biolo
- [7] Rinella ME, et al. A multi-society Delphi consensus statement on new fatty liver disease nomenclature. Ann Hepatol 2023;101133.
- [8] Masoodi M, et al. Metabolomics and lipidomics in NAFLD: biomarkers and noninvasive diagnostic tests. *Nature reviews Gastroenterology* &. hepatology 2021;18: 835–56.
- [9] Noureddin M, et al. Serum identification of at-risk MASH: The metabolomicsadvanced steatohepatitis fibrosis score (MASEF). Hepatology 2024;79:135–48.
- [10] Michelena J, et al. Metabolomics discloses a new non-invasive method for the diagnosis and prognosis of patients with alcoholic hepatitis. Ann Hepatol 2019;18: 144–54.
- [11] Mayo R, et al. Metabolomic-based noninvasive serum test to diagnose nonalcoholic steatohepatitis: results from discovery and validation cohorts. Hepatol Commun 2018;2:807–20.
- [12] Ooi GJ, et al. Hepatic lipidomic remodeling in severe obesity manifests with steatosis and does not evolve with non-alcoholic steatohepatitis. J Hepatol 2021; 75:524–35.
- [13] Juluri R, et al. Generalizability of the nonalcoholic steatohepatitis Clinical Research Network histologic scoring system for nonalcoholic fatty liver disease. J Clin Gastroenterol 2011;45:55–8.
- [14] Bedossa P. 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 2014;60:565–75.
- [15] Mayo R, et al. Metabolomic-based noninvasive serum test to diagnose nonalcoholic steatohepatitis: Results from discovery and validation cohorts. Hepatol Commun 2018;2:807–20.

#### J. Chen et al.

#### Computational and Structural Biotechnology Journal 23 (2024) 791–800

- [16] Lear SA, Humphries KH, Kohli S, Birmingham CL. The use of BMI and waist circumference as surrogates of body fat differs by ethnicity. Obes (Silver Spring) 2007;15:2817–24.
- [17] van der Poorten D, et al. Visceral fat: a key mediator of steatohepatitis in metabolic liver disease. Hepatology 2008;48:449–57.
- [18] Yip TC, Lee HW, Chan WK, Wong GL, Wong VW. Asian perspective on NAFLDassociated HCC. J Hepatol 2022;76:726–34.
- [19] Chen J, et al. Ratio of Conjugated Chenodeoxycholic to Muricholic Acids is Associated with Severity of Nonalcoholic Steatohepatitis. Obes (Silver Spring) 2019;27:2055–66.
- [20] Peyrou M, Bourgoin L, Foti M. PTEN in non-alcoholic fatty liver disease/nonalcoholic steatohepatitis and cancer. Dig Dis 2010;28:236–46.
- [21] Calder PC. Omega-3 fatty acids and metabolic partitioning of fatty acids within the liver in the context of nonalcoholic fatty liver disease. Curr Opin Clin Nutr Metab Care 2022;25:248–55.
- [22] Chen Z, Tian R, She Z, Cai J, Li H. Role of oxidative stress in the pathogenesis of nonalcoholic fatty liver disease. Free Radic Biol Med 2020;152:116–41.
- [23] van de Wier B, et al. Elevated citrate levels in non-alcoholic fatty liver disease: the potential of citrate to promote radical production. FEBS Lett 2013;587:2461–6.
- [24] Tanaka N, Matsubara T, Krausz KW, Patterson AD, Gonzalez FJ. Disruption of phospholipid and bile acid homeostasis in mice with nonalcoholic steatohepatitis. Hepatology 2012;56:118–29.

- [25] Gaggini M, et al. Altered amino acid concentrations in NAFLD: Impact of obesity and insulin resistance. Hepatology 2018;67:145–58.
- [26] Siomkajlo M, Daroszewski J. Branched chain amino acids: Passive biomarkers or the key to the pathogenesis of cardiometabolic diseases? Adv Clin Exp Med 2019; 28:1263–9.
- [27] Waluk DP, Vielfort K, Derakhshan S, Aro H, Hunt MC. N-Acyl taurines trigger insulin secretion by increasing calcium flux in pancreatic beta-cells. Biochem Biophys Res Commun 2013;430:54–9.
- [28] Reinke SN, et al. Metabolomics analysis identifies different metabotypes of asthma severity. Eur Respir J 2017;49.
- [29] Ohue-Kitano R, et al. Medium-chain fatty acids suppress lipotoxicity-induced hepatic fibrosis via the immunomodulating receptor GPR84. JCI Insight 2023;8.
- [30] Lambert JE, Ramos-Roman MA, Browning JD, Parks EJ. Increased de novo lipogenesis is a distinct characteristic of individuals with nonalcoholic fatty liver disease. Gastroenterology 2014;146:726–35.
- [31] Puri P, et al. The plasma lipidomic signature of nonalcoholic steatohepatitis. Hepatology 2009;50:1827–38.
- [32] Jang JE, et al. Protective role of endogenous plasmalogens against hepatic steatosis and steatohepatitis in mice. Hepatology 2017;66:416–31.
- [33] Oresic M, et al. Prediction of non-alcoholic fatty-liver disease and liver fat content by serum molecular lipids. Diabetologia 2013;56:2266–74.
- [34] Yoo W, et al. Fatty acids in non-alcoholic steatohepatitis: Focus on pentadecanoic acid. PLoS One 2017;12:e0189965.