

RESEARCH ARTICLE

Open Access



Serum metabolite profiles are associated with the presence of advanced liver fibrosis in Chinese patients with chronic hepatitis B viral infection

Guoxiang Xie^{1,2†}, Xiaoning Wang^{1,3†}, Runmin Wei^{4†}, Jingye Wang^{4†}, Aihua Zhao⁵, Tianlu Chen⁵, Yixing Wang³, Hua Zhang^{1,3}, Zhun Xiao^{1,3}, Xinzhu Liu^{1,3}, Youping Deng⁴, Linda Wong⁴, Cynthia Rajani⁴, Sandi Kwee⁴, Hua Bian⁶, Xin Gao⁶, Ping Liu^{1,3,7*} and Wei Jia^{1,4,8*}

Abstract

Background: Accurate and noninvasive diagnosis and staging of liver fibrosis are essential for effective clinical management of chronic liver disease (CLD). We aimed to identify serum metabolite markers that reliably predict the stage of fibrosis in CLD patients.

Methods: We quantitatively profiled serum metabolites of participants in 2 independent cohorts. Based on the metabolomics data from cohort 1 (504 HBV associated liver fibrosis patients and 502 normal controls, NC), we selected a panel of 4 predictive metabolite markers. Consequently, we constructed 3 machine learning models with the 4 metabolite markers using random forest (RF), to differentiate CLD patients from normal controls (NC), to differentiate cirrhosis patients from fibrosis patients, and to differentiate advanced fibrosis from early fibrosis, respectively.

Results: The panel of 4 metabolite markers consisted of taurocholate, tyrosine, valine, and linoelaidic acid. The RF models of the metabolite panel demonstrated the strongest stratification ability in cohort 1 to diagnose CLD patients from NC (area under the receiver operating characteristic curve (AUROC) = 0.997 and the precision-recall curve (AUPR) = 0.994), to differentiate fibrosis from cirrhosis (0.941, 0.870), and to stage liver fibrosis (0.918, 0.892). The diagnostic accuracy of the models was further validated in an independent cohort 2 consisting of 300 CLD patients with chronic HBV infection and 90 NC. The AUCs of the models were consistently higher than APRI, FIB-4, and AST/ALT ratio, with both greater sensitivity and specificity.

Conclusions: Our study showed that this 4-metabolite panel has potential usefulness in clinical assessments of CLD progression in patients with chronic hepatitis B virus infection.

Keywords: Bile acids, Free fatty acids, Amino acids, Hepatitis B, Chronic liver disease, Liver fibrosis, Metabolomics, Random forest

* Correspondence: liuliver@vip.sina.com; wjia@cc.hawaii.edu

[†]Guoxiang Xie, Xiaoning Wang, Runmin Wei and Jingye Wang contributed equally to this work.

¹E-Institute of Shanghai Municipal Education Committee, Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Liver fibrosis is a wound-healing response to damage caused by chronic liver disease (CLD) [1]. Liver fibrosis can progress to cirrhosis over years or decades [2], and results in liver function decline and increased risk of hepatocellular carcinoma (HCC). Liver biopsy has been the gold standard for evaluating the presence and degree of liver fibrosis, but its clinical application is limited by inherent limitations such as invasiveness, sampling errors, and intra- and inter-observer variability [3]. Recent studies indicated that liver fibrosis could be reversed [1], creating the need for less invasive clinical tools to monitor and assess the responses of CLD patients to treatments. A number of scoring systems, such as the FibroTest [4], the aspartate transaminase/alanine transaminase (AST/ALT) ratio [5], the AST/Platelet Ratio Index (APRI) [6], FIB-4 (patient age, AST, ALT, and platelet) [7], *Wisteria floribunda* agglutinin-positive Mac-2 binding protein (WFA⁺-M2BP) [8], and machine learning-based clinical predictive models [9], have recently been used to stage CLD and predict the development of liver fibrosis and cirrhosis. Imaging techniques, such as computed tomography, magnetic resonance imaging [10], and two recently approved ultrasound-based systems, shear wave elastography and transient elastography (FibroScan) [11], have also been used clinically to assess the degree of liver fibrosis. However, these imaging modalities have limited accuracy in some patients, such as those with ascites, elevated central venous pressure, and obesity [12].

Developing noninvasive, accurate, and reliable markers to assess the severity and progression of liver fibrosis in CLD patients has become increasingly important for treatment decisions, for continuous monitoring of patients who have mild liver disease and are not under treatment [13], and for risk stratification and longitudinal follow-up in clinical trials.

Alterations of bile acids (BAs) [13–15], free fatty acids (FFAs) [16], and amino acids (AAs) [17] are closely associated with CLD regardless of etiology. However, the relationship between serum AAs, BAs, and FFAs and the stages of liver fibrosis have not been thoroughly investigated. The aim of this study was to identify serum metabolite markers that reliably predict the stage of fibrosis in CLD patients with chronic hepatitis B virus (HBV) infection, a leading cause of CLD worldwide. We used a targeted metabolomics approach to quantify serum BAs, AAs, and FFAs in 1006 participants in cohort 1 (504 biopsy-proven fibrosis and cirrhosis CLD patients with chronic HBV infection and 502 normal controls, NC), and selected four predictive metabolite markers to construct three machine learning models using random forest (RF). Model 1 diagnosed CLD patients from NC, model 2 differentiated cirrhosis patients from fibrosis patients, and model 3 differentiated advanced fibrosis

and early fibrosis patients. The diagnostic accuracy of the three models was further validated in an independent cohort consisting of 300 HBV-CLD patients and 90 NC.

Methods

Study design and participants

Two datasets were enrolled in this study. Cohort 1 was recruited between April 2013 and June 2015 at Shuguang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine, consisted of 1006 participants, including 504 CLD patients with chronic HBV infection and 502 NC as our training cohort to identify serum metabolite markers and establish predictive models (Table 1). All the patients were tested positive for HBV-DNA or positive for hepatitis B surface antigen (HBsAg). Infection with chronic HBV was diagnosed according to the “Guideline on prevention and treatment of chronic hepatitis B in China” [18]. More detailed inclusion and exclusion criteria can be found in Additional file 1.

Cohort 2, recruited between December 2016 and December 2017 at Xiamen Hospital of Traditional Chinese Medicine, consisted of 300 CLD patients with chronic HBV infection and 90 NC. Data obtained from cohort 2 were used as a validation set to further verify the performance of the models established from the cohort 1. Detailed information about this cohort can be found in Additional file 1. Sample size was not determined by statistical methods and was comparable to other studies in the field [4–8, 17, 19].

In this study, the diagnosis and the sample collection were performed using exactly the same protocols to avoid “external” influences. The samples were provided to lab staffs blind samples with respect to patient identity and other clinical information.

The study was organized and led by Shuguang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine, and participated by Xiamen Hospital of Traditional Chinese Medicine. The study was approved by the institutional review board of Shuguang Hospital first (approval no. 2012-206-22-01) and endorsed by the ethics committee of Xiamen Hospital. All participants provided written informed consent.

Liver biopsy

All patients, except those diagnosed with decompensated cirrhosis (presence of any of the following complications in cirrhosis: variceal hemorrhage, ascites, encephalopathy, and jaundice), received a liver biopsy directed by ultrasonography within 1 week after enrollment. The biopsy specimens were fixed with 10% formalin, embedded in paraffin, and stained with hematoxylin/eosin and Masson’s trichrome stain. Examination of a minimum

Table 1 Demographic and clinical data of patients with CLD and NC in cohort 1 (training set) and cohort 2 (validation set)

Dataset	Cohort 1 training set						Cohort 2 validation set					
	Control	CLD	S0-2	S3-4	Fibrosis	Cirrhosis	Control	CLD	S0-2	S3-4	Fibrosis	Cirrhosis
n	502	504	349	155	400	104	90	300	134	166	141	159
Sex (MF)	365/137	361/143	257/92	104/51	299/101	62/42	59/31	202/98	81/53	121/45*	86/55	116/43*
Age (year)	36.65 ± 11.73	36.58 ± 11.88	33.23 ± 9.95	44.88 ± 12.22***	34.05 ± 10.31	48.65 ± 11.51***	47.13 ± 9.95	47.96 ± 13.28	41.55 ± 12.83	53.14 ± 11.26***	41.21 ± 12.72	53.95 ± 10.67***
BMI (kg/m ²)	23.08 ± 3.16	22.28 ± 3.18***	22.02 ± 3.26	22.91 ± 2.91**	22.15 ± 3.22	22.87 ± 2.97*	22.35 ± 1.83	23.18 ± 3.13*	23.28 ± 2.48	23.1 ± 3.61	23.21 ± 2.49	23.15 ± 3.65
APRI	0.09 ± 0.04	0.79 ± 1.33***	0.6 ± 0.81	1.27 ± 2.08***	0.63 ± 0.92	1.55 ± 2.4**	0.65 ± 0.7	0.65 ± 0.7	0.61 ± 0.78	0.69 ± 0.57**	0.6 ± 0.76	0.72 ± 0.58***
AST/ALT	0.82 ± 0.32	0.69 ± 0.44***	0.6 ± 0.31	0.94 ± 0.59***	0.61 ± 0.33	1.1 ± 0.63***	1.09 ± 0.6	1.09 ± 0.6	0.98 ± 0.57	1.24 ± 0.61***	0.97 ± 0.57	1.27 ± 0.61***
FIB-4	0.62 ± 0.33	2.86 ± 7.21***	1.42 ± 1.48	6.57 ± 12.71***	1.56 ± 1.7	9.42 ± 15.8***	3.64 ± 3.64	3.64 ± 3.64	2.49 ± 2.58	5.26 ± 4.27***	2.45 ± 2.53	5.54 ± 4.3***
ALT (IU/L)	30.97 ± 15.63	176.49 ± 199.81***	195.3 ± 213.41	128.51 ± 150.32***	193.08 ± 208.27	94.18 ± 121.97***	17.92 ± 7.79	81.57 ± 117.77***	111.4 ± 141.39	57.2 ± 87.35***	108.45 ± 138.58	57.43 ± 89.06***
AST (IU/L)	21.81 ± 6.84	93.21 ± 99.71***	95.65 ± 100.23	86.98 ± 98.47	96.32 ± 102.01	77.75 ± 86.32	20.25 ± 4.37	68.73 ± 73.78***	85.6 ± 91.48	54.95 ± 51.62*	83.2 ± 89.82	55.74 ± 52.57
TBIL (μmol/L)	15.5 ± 4.84	27.73 ± 32.98***	21.75 ± 13.27	42.87 ± 55.62***	23.06 ± 20.64	50.58 ± 61.18***	13.98 ± 3.62	33.67 ± 47.77***	23.77 ± 38.4	41.77 ± 52.99***	23.54 ± 37.45	42.77 ± 53.94***
ALP (IU/L)	85.57 ± 19.18	89.72 ± 76.26	81.3 ± 65.24	111.05 ± 95.86**	82.37 ± 61.46	125.77 ± 119.98**	77.82 ± 19.21	93.68 ± 70.86*	70.57 ± 55.75	112.56 ± 76.25***	70.97 ± 54.47	114.08 ± 77.53***
GGT (IU/L)	17.12 ± 10.72	69.18 ± 97.06***	61.13 ± 69.75	89.73 ± 143.43*	66.72 ± 74.55	81.42 ± 169.46	26.26 ± 19.07	76.87 ± 82.16***	77.22 ± 83.58	76.59 ± 81.23	76.14 ± 81.77	77.53 ± 82.76
TP (g/L)	74.41 ± 4.76	73.47 ± 8.55*	75.68 ± 5.31	67.87 ± 12.03***	75.42 ± 5.24	63.82 ± 13.76***	73.65 ± 3.19	71.31 ± 17.94***	72.89 ± 5.84	69.97 ± 23.74***	73.07 ± 5.82	69.66 ± 24.23***
ALB (g/L)	49.23 ± 2.77	40.19 ± 5.91***	42.14 ± 3.37	35.23 ± 7.81***	41.75 ± 3.54	32.47 ± 8.67***	44.92 ± 2.19	37.86 ± 7.18***	41.44 ± 4.79	34.93 ± 7.48***	41.39 ± 4.71	34.69 ± 7.54***
TBA (μmol/L)	4.67 ± 3.18	28.47 ± 45.11***	20.33 ± 37.28	49.68 ± 55.81***	21.3 ± 36.44	65.41 ± 63.96***	3.6 ± 2.63	43.41 ± 55.16***	25.05 ± 37.86	58.89 ± 62.37***	24.33 ± 37.04	61.11 ± 62.91***
PLT (10 ⁹ /L)	261.02 ± 65.25	164.52 ± 61.7***	184.2 ± 47.74	114.68 ± 65.05***	179.06 ± 49.86	93.23 ± 64.82***	132.44 ± 62.97	132.44 ± 62.97	163.58 ± 51	106.03 ± 60.14***	161.54 ± 51.18	105.27 ± 60.9***
Collagen proportionate area	7.46 ± 4.01	1.96 ± 4.01	1.96 ± 4.01	9.95 ± 6.03***	2.71 ± 2.45	15.17 ± 7.11***						
HBV-DNA (log10)	6.25 ± 2.42	6.25 ± 2.42	6.32 ± 2.44	5.99 ± 2.34	6.33 ± 2.38	5.39 ± 2.66						
Negative HbeAg, n	191	115	115	76	142	49						
Negative HbeAb, n	223	153	153	70	175	48						
Negative HbsAg, n	29	24	24	5	26	3						

Values are expressed as mean ± SD

ALT alanine transaminase, AST aspartate transaminase, TBIL total bilirubin, ALP alkaline phosphatase, GGT gamma-glutamyl transferase, ALB albumin, TBA total bile acid, PLT platelet
*p < 0.05, **p < 0.01, ***p < 0.001, by Student's t test, CLD vs. NC, S3-4 vs. S0-2, cirrhosis vs. fibrosis

length of 1.5 cm of the liver biopsy and at least six portal tracts were required for diagnosis. Histological grading of necro-inflammation (G0 to G4) and staging of liver fibrosis (S0 to S4) were carried out according to Scheuer's classification [20]. All samples were independently assessed by three pathologists from Shanghai Medical College of Fudan University, Shanghai, China, who were blinded to the sample ID. Specimens with discrepant assessments were re-examined until a consensus was reached. The final assessments of the three pathologists were further processed using the kappa concordance test.

Histological assessment of liver injury

The obtained liver tissues via liver biopsy were fixed in 10% formalin (Sigma), processed using established protocols, and embedded in paraffin. Sections (5 μm) of each sample were cut and stained with hematoxylin and eosin (H&E) for histopathological analysis. All sections were examined using a light microscope. Based on the H&E staining results, the necro-inflammation activity of chronic hepatitis was determined as G0 to G4 according to Scheuer's classification as G0 (absent), G1 (portal inflammation only), G2 (mild interface hepatitis), G3 (moderate interface hepatitis), and G4 (severe interface hepatitis) (Additional file 2: Figure S1).

Collagen proportionate area using digital image analysis

The obtained liver tissues via liver biopsy were fixed in 10% formalin (Sigma). Tissue samples were embedded in paraffin blocks and then sliced into 5- μm -thick sections. Sections were processed and stained with Masson's trichrome as reported [21]. Masson staining kits were from Abcam Co., Ltd. (Trichrome Stain, ab150686). Collagen stained blue (Additional file 2: Figure S2). In order to characterize collagen area, Masson's trichrome-stained slides were scanned with a Leica SCN400 scanner (Leica Microsystems) at $\times 40$ magnification and measured using Aperio ImageScope (v12.3.2.5030, Aperio Technologies). The images were saved as ".scn" format files. The Color Deconvolution algorithm (v9, Aperio Technologies) was used to isolate individual stains for semi-quantification. The percent total positive, total stained area (mm^2), and total analysis area (mm^2) in each visual field were measured and recorded. The analytical data were saved as ".xls" format files. CPA = percent total positive \times total stained area/total analysis area.

Serum sample collection

Overnight fasting (12 h) blood samples were collected from all subjects, and sera were delivered to our laboratories on ice within 2 h of collection. Samples were aliquoted and stored at -80°C until analysis.

Blood clinical marker measurement

Hematological and standard biochemical tests were performed using an LH750 Hematology Analyzer and a Synchron DXC800 Clinical System (Beckman Coulter, USA) according to the manufacturer's protocol. The coagulation function was measured using an automatic coagulation analyzer (STAGO Compact, Diagnostica Stago, France). The serum HBV-DNA level was quantified using a real-time polymerase chain reaction (PCR) system (LightCycler 480, Roche, USA).

Metabolomics analysis

Samples in cohort 1 were analyzed at the Center for Translational Medicine, Shanghai Jiao Tong University Affiliated Sixth People's Hospital. Samples in cohort 2 were analyzed at the Metabo-Profile Biotechnology (Shanghai) Co., Ltd. BAs and AAs were quantified using ultra-performance liquid chromatography (UPLC)-triple quadrupole mass spectrometry (Waters XEVO TQ-S, Milford, MA), and FFAs were quantified using UPLC quadrupole time-of-flight mass spectrometry (Waters XEVO G2S, Milford, MA), according to our previously reported protocol [22–25].

The detailed procedure and analysis were performed as described in Additional file 1.

Classification performance evaluation

ROC curve is a plot of the true positive rate (sensitivity/recall) against the false positive rate ($1 - \text{specificity}$) at different cutoffs of a binary classifier. AUROC measures the area under the ROC curves, and a higher value of AUROC suggests better classification performances while an AUROC of 0.5 represents the random guess. The PR curve demonstrates the relationship between positive predictive values (precision) and true positive rate (sensitivity/recall), and a higher value of AUPR indicates better diagnostic capacity of the model. PR curves are usually preferable for evaluating unbalanced data compared to ROC curves. NRI and IDI were also used for the evaluation of prediction improvement. We compared RF models to existing clinical indices by splitting the continuous risk scores into ten equal risk intervals (default). We used the R software version 3.2.3 for data analysis and the "PRROC" R package for binary ROC and PR curves [26], the "pROC" package for calculating the specificities and sensitivities of classifiers [27], and the "PredictABEL" package for NRI and IDI calculation [28].

Feature selection and method comparison

Quantitative variables were expressed as mean \pm SD for clinical parameters and median (25% quantile, 75% quantile) of log₁₀ transformed concentration for metabolites. Categorical variables were expressed as percentages. The univariate analysis (Wilcoxon's rank-sum test)

was carried out to identify the variables that were significantly different between CLD patients and NC, between fibrosis and cirrhosis (S0–3 vs. S4), and among CLD patients at different fibrotic stages (early stage fibrosis (S0–2) vs. advanced stage fibrosis (S3–4)).

For differential metabolites with $p < 0.001$ across, all univariate analyses were used in two machine learning methods, LASSO [29] and RF [30], to further select markers for the three classifications listed above. Data were log and z-score transformed before being fed into LASSO to ensure that the coefficients were comparable with each other. The regularization parameter lambda of LASSO was determined using 10-fold cross-validation (CV). The RF model used 500 decision trees. We ranked the metabolites according to their LASSO non-zero coefficients and RF mean decrease of accuracy, and kept the intersection of top 5 LASSO and RF metabolites in the three classifications. Considering the overlaps of the second and the third classification tasks, we further selected the intersecting variables of these two situations and then merged with variables selected from the first situation to construct our final metabolite markers (Fig. 2).

To identify an appropriate classification method, we introduced two linear models, i.e., logistic regression (LR) and linear discriminant analysis (LDA), and one decision tree-based ensemble model, i.e., RF, for the classifier construction for the markers we selected. For RF, we used 500 decision trees and two candidate variables at each split. For LDA, the tolerance parameter was set to $1.0E-4$ (default). We applied 10-fold CV on the training set (cohort 1) to compare the classification performances of these four models and three established fibrosis markers, i.e., APRI, AST/ALT ratio, and FIB-4. AUROC and AUPR were recorded at each internal validation set in CV. We used R packages “randomForest,” “glmnet,” and “MASS” for RF, LASSO, and LDA constructions, respectively [31, 32].

Predictive model construction and validation

We trained the final RF models for different classification objectives using the training data (cohort 1), with model 1 differentiating CLD and NC, model 2 differentiating fibrosis and cirrhosis, and model 3 differentiating early and advanced stages of liver fibrosis. A total of 500 decision trees were included in a single RF model with two variables randomly sampled as candidates at each split. We re-balanced the sample size for different groups at each bootstrap resampling step for models 2 and 3 considering the unbalanced samples [33].

In RF, each decision tree was fitted on the bootstrap samples and tested on the untouched OOB samples. Thus, the OOB predictions provided unbiased estimates of how the RF model performed on the training data and were used for the evaluation on cohort 1. We

further validated our mark panel-based RF models in the independent validation datasets from cohort 2, and compared results with the established fibrosis markers, AST/ALT ratio, APRI, and FIB-4. ROC and PR curves were drawn, and AUROC and AUPR values, respectively, were calculated to evaluate their diagnostic performances. Optimal cutoffs were selected to maximize the sum of sensitivity and specificity for the RF model. For APRI, FIB-4, and AST/ALT, predefined cutoffs were used (1.0 and 2.0 for APRI to distinguish fibrosis and cirrhosis [6], 1.45 and 3.25 for FIB-4 to distinguish S0–2 and S3–4 [7], and 0.8 and 1.0 for AST/ALT to distinguish S0–2 and S3–4 [5, 34]). Bootstrap resampling (1000 times) was conducted to calculate 95% confidence intervals (CIs) of AUCs for all binary classifiers. A comparison of the AUROC of our biomarker panel vs. FIB-4, AST/ALT, or APRI was performed using DeLong’s test. The significance level was adjusted for multiple testing according to the Benjamini and Hochberg procedure [35]. Log and z-score transformed data were also used for constructing heatmaps. The R packages “ggplot2” and “cowplot” were used for data visualization and multiple plots arrangement.

We further derived an RF risk score for each participant based on the marker panel and logit function of the predicted probability (Prob.) of the RF model for corresponding classification objective:

$$\text{RF score} = \text{logit}(\text{Prob.}) = \log\left(\frac{\text{Prob.}}{1-\text{Prob.}}\right)$$

F1 scores were then calculated at the predefined cutoffs using following formula:

$$F1 = \frac{2}{\text{Precision}^{-1} + \text{Recall}^{-1}}$$

To determine whether the RF score could independently predict the fibrosis staging in the presence of other potential confounding factors, we applied logistic regression on the RF score to differentiate cirrhosis from fibrosis as well as discriminate early and advanced fibrosis while adjusting for HBV-DNA levels, the degree of necro-inflammation, HBeAb status, HBeAg status, liver function tests (i.e., PT, ALB, DBIL, IBIL), platelets, BMI, and medication (entecavir) use.

Multi-group classification of S0–2 vs. S3 vs. S4

We built a new RF model based on our metabolite marker panel and applied multinomial regressions to APRI, AST/ALT, and FIB-4 separately to differentiate S0–2 vs. S3 vs. S4 in cohort 1. Then, we compared and validated these multi-group classifiers on both cohort 1 and cohort 2 datasets using micro-average ROC and PR curves. Micro-average ROC and PR curves were

calculated by stacking binary classification results from each group together to generate a concatenated binary classification result [36]. We then calculated AUROC and AUPR with 95% CIs using 100 times bootstrap resampling. We used the “multiROC” R package for calculating the micro-average AUROC and AUPR as well as for plotting [37].

Results

Characteristics of the participants

Two independent cohorts were studied (Fig. 1). CLD groups were staged and assigned according to the results of their liver biopsy. Cohort 1 consisted of 1006 participants (502 NC and 504 biopsy-proven HBV-CLD patients (400 with liver fibrosis (S0–3) and 104 with cirrhosis (S4), or 349 with early stage fibrosis (S0–2) and 155 with advanced stage fibrosis (S3–4))). Cohort 2 consisted of 390 participants (90 NC and 300 biopsy-proven CLD patients comprising 141 with fibrosis and 159 with cirrhosis, or 134 with early stage fibrosis (S0–2) and 166 with advanced stage fibrosis (S3–4)). Models established

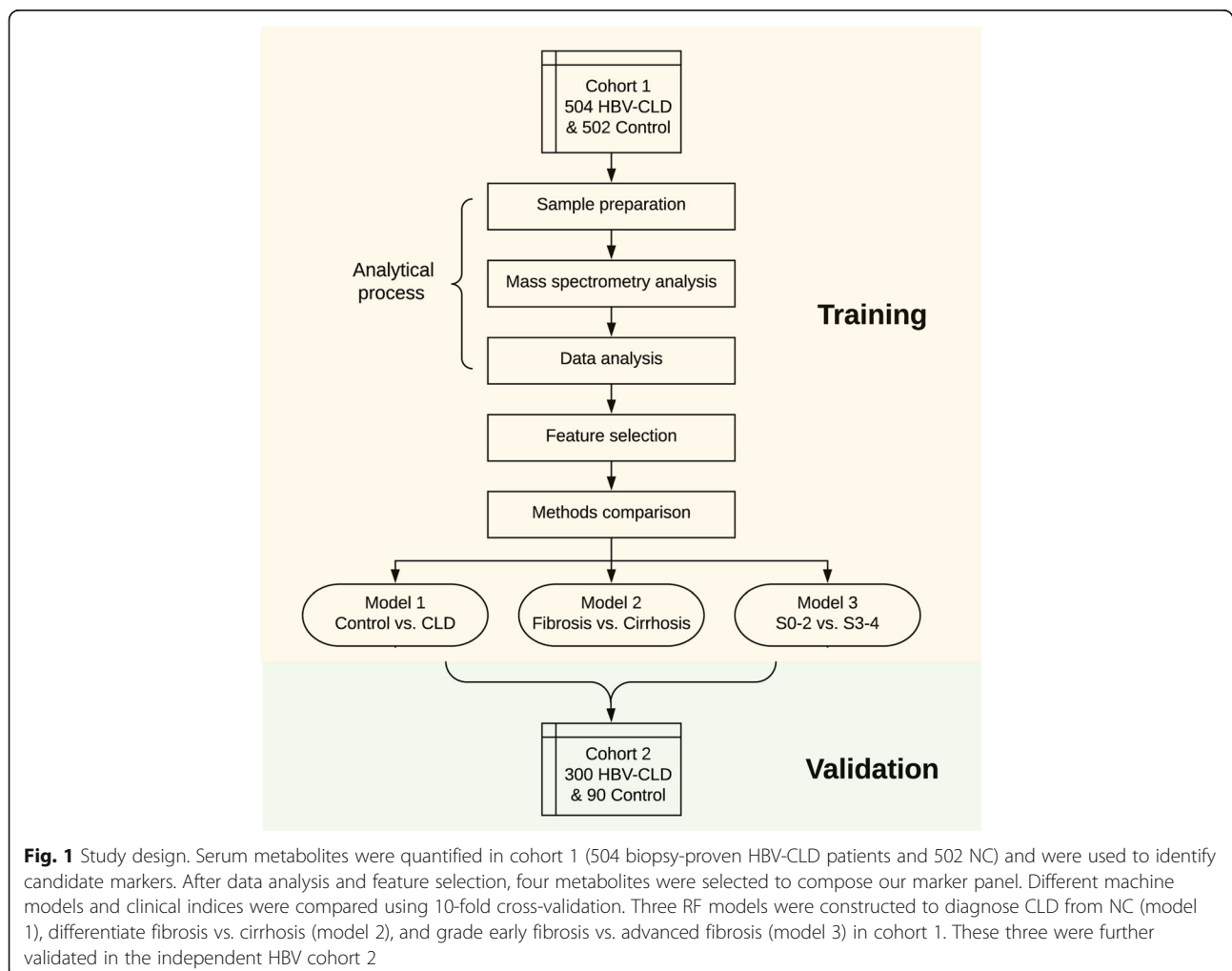
from cohort 1 were validated in cohort 2 (models 1, 2, and 3). The cohort stratification and major demographic and clinical characteristics are shown in Table 1. More detailed clinical data are provided in Additional file 3: Table S1.

Quantification of metabolites in serum

Using targeted metabolomic protocols established in our lab [22–24], we quantified the concentrations of 98 metabolites, including 24 BAs, 42 FFAs, and 32 AAs, in the sera of all participants (Additional file 3: Table S2). These metabolites were used for the subsequent metabolite marker selection.

Serum metabolite marker selection

From the 98 serum metabolites, we identified 26 differential metabolites in three classification situations (i.e., to diagnose CLD patients from NC, to differentiate fibrosis from cirrhosis, and to differentiate advanced fibrosis from early fibrosis) using univariate analysis (Wilcoxon’s rank-sum test, $p < 0.001$). The 26



statistically significant metabolites were then entered into least absolute shrinkage and selection operator (LASSO) [29] and random forest (RF) [30]. According to the rank of LASSO non-zero coefficients and RF mean decrease of accuracy, four metabolite markers were selected, which included one FFA, linoelaidic acid (C18:2 n6t); one BA, taurocholate (TCA); and two AAs, tyrosine (Tyr) and valine (Val) (Fig. 2). The principal component analysis (PCA) of these four metabolite markers also showed a clear separation between CLD patients and NC (Additional file 2: Figure S3). We also derived one

ratio, the Tyr/Val ratio, to further improve the classification performances while also including one extra accessible risk factor, age, to our panel for the differentiation of fibrosis and cirrhosis and the staging of fibrosis. Correlations of the four metabolites with fibrosis stage, necro-inflammation, CPA, AST, ALT, AST/ALT ratio, PLT, FIB-4, and APRI were assessed using Spearman's correlation analysis (Additional file 2: Figure S4). The four metabolite markers (including the Tyr/Val ratio) all significantly correlated with fibrosis stage ($\rho = 0.38$ for TCA, $\rho = 0.50$ for Tyr, $\rho = 0.53$ for Tyr/Val ratio, and

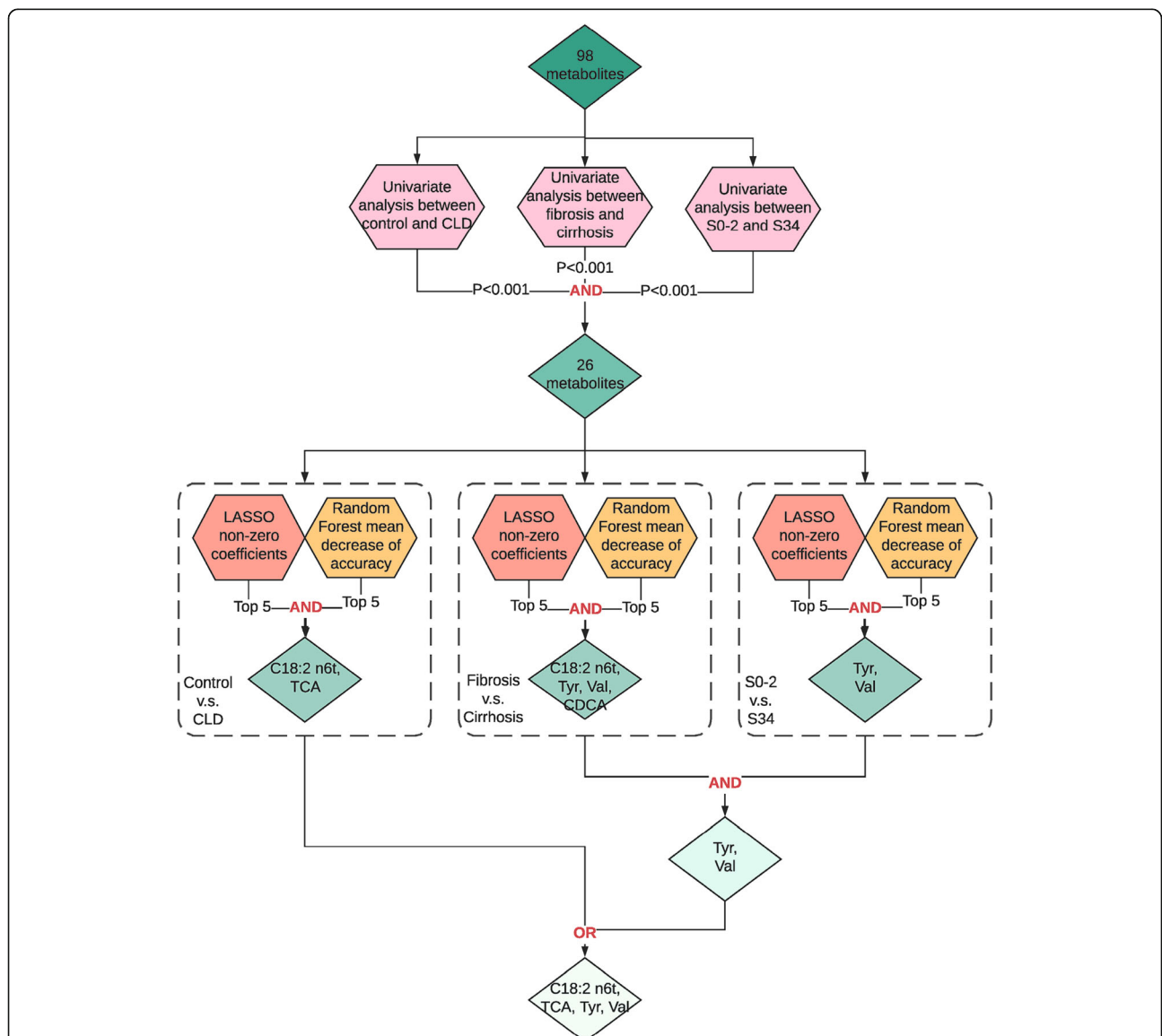


Fig. 2 Workflow chart of feature selection. For a total of 98 metabolites (including AAs, BAs, and FFAs), univariate analyses (Wilcoxon's rank-sum test) were employed for three clinical aims (aim 1: CLD vs. NC, aim 2: cirrhosis vs. fibrosis, aim 3: early fibrosis vs. advanced fibrosis). Twenty-six metabolites with $p < 0.001$ in all three clinical aims were selected and fed into LASSO and random forest algorithms for three aims. The overlap of top 5 LASSO non-zero coefficients and top 5 important variables from random forest (by mean decrease of accuracy) was selected. For aim 2 and aim 3, we selected the overlapped variables and combined with variables selected from aim 1 to yield the final panel four metabolites. "OR" means the union of two sets, and "AND" means the intersection of two or more sets

$\rho = 0.23$ for C18:2 n6t) using Spearman's correlation analysis. In addition, we found our metabolite markers showed stronger associations with the fibrosis stage than the previously used clinical indices.

To determine an appropriate classification model, we applied 10-fold CV on cohort 1 to compare the classification performances of RF models and two linear models (i.e., logistic regression (LR), linear discriminant analysis (LDA)) as well as the clinical indices, APRI, AST/ALT ratio, and FIB-4. The CV-area under the receiver operating characteristic curve (CV-AUROC) and CV-area under the precision-recall curve (CV-AUPR) were employed as the evaluation metrics. We found that to differentiate CLD from control, APRI, LR, LDA, and RF had the highest AUROCs and AUPR, while RF demonstrated the most robust classification performance (Additional file 2: Figure S5a). For the differentiation of fibrosis and cirrhosis and S0–2 vs. S3–4, RF outperformed other methods with the highest CV-AUROC and CV-AUPR overall (Additional file 2: Figure S5b, c). PCA score plot showed linearly separable discrimination between the most CLD and control subjects (Additional file 2: Figure S3); thus, linear models could achieve good classification performances. However, for a situation where there is more extensive overlapping of groups (Additional file 2: Figure S6), the decision tree-based ensemble learning algorithm, RF, achieved improved classification performances compared to other methods (Additional file 2: Figure S5).

Model 1: Differentiating CLD patients from NC

The concentration of linoelaidic acid (C18:2 n6t) was significantly higher in the control group than in the CLD group, and conversely, the levels of TCA, Tyr, and Tyr/Val ratio were higher in the CLD group than in the control group (Fig. 3a, b).

Model 1 was constructed using an RF model that utilized these four metabolite markers, to differentiate CLD patients from NC in cohort 1. Out-of-bag (OOB) estimates were employed for the RF model evaluations. Model 1 showed an AUROC of 0.997 (0.993–1.000) and AUPR of 0.994 (0.998–1.000) (Fig. 3c, d) which was significantly higher than the APRI (AUROC = 0.973, $p < 0.001$), FIB-4 (AUROC = 0.848, $p < 0.001$), and AST/ALT ratio (AUROC = 0.665, $p < 0.001$) (Table 2). An example decision tree from the RF model is shown in Additional file 2: Figure S7a, where we observed that the lower concentration of C18:2 n6t and the higher concentration of TCA would lead to higher risk of CLD.

Based on the OOB predicted probabilities, we calculated a diagnostic RF score for model 1 using the logit function. The waterfall plot showed a clear ascending trend of RF scores from NC (lower RF scores) to CLD patients (higher RF scores) along with the differentiation

trend shown in the heatmap of the four markers (Fig. 3b). We observed significant differences in the RF score between both groups in cohort 1 ($p < 0.001$, Fig. 3g), yielding a sensitivity of 98.4% and specificity of 99% for CLD patients in the training set at a cutoff value of 0.434 (Table 2). The sensitivity and specificity of our RF model were superior to those of AST/ALT ratio, APRI, and FIB-4 for differentiating CLD patients from NC using the optimal cutoffs generated in cohort 1 using the Youden index (Additional file 3: Table S3).

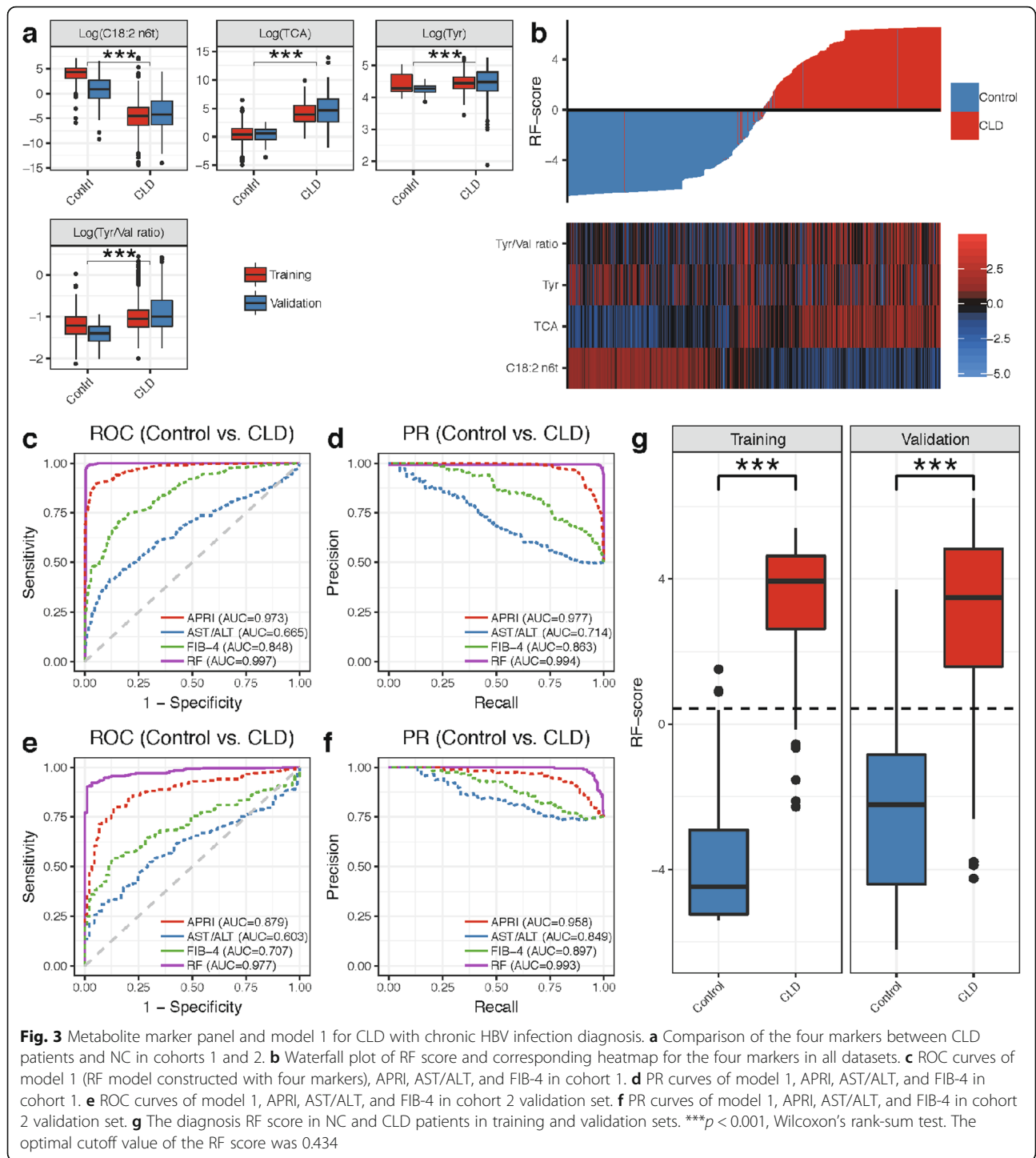
Model 2: Differentiating cirrhosis from fibrosis among CLD patients

The discriminant prediction model was constructed using an RF model employing the four metabolite markers along with age to differentiate CLD patients with cirrhosis from those without cirrhosis in cohort 1. This model demonstrated an AUROC of 0.941 (0.914–0.964) and AUPR of 0.87 (0.824–0.913) (Fig. 4a, b) based on OOB predictions. These results were better than those of the APRI (AUROC = 0.698, $p < 0.001$), AST/ALT (AUROC = 0.815, $p < 0.001$), and FIB-4 (AUROC = 0.869, $p < 0.001$) (Table 2). We showed an example decision tree for model 2 in Additional file 2: Figure S7b, and we found that the higher the Tyr/Val ratio, Tyr, and C18:2 n6t, the higher the risk of CLD with cirrhosis.

The model 2 RF score differentiated CLD patients with cirrhosis from fibrosis in cohort 1 ($p < 0.001$) (Fig. 4e). The constructed model yielded a sensitivity of 87.0% and specificity of 90.4% in the cohort 1 dataset at a cutoff value of 0.01 (Table 2). The RF scores remained significant with a coefficient of 0.755 ($p < 0.001$) after adjusting for HBV-DNA levels, degree of necro-inflammation, HBeAb status, HBeAg status, body mass index (BMI), platelets (PLT), liver function tests (i.e., prothrombin time (PT), albumin (ALB), direct bilirubin (DBIL), indirect bilirubin (IBIL)), and medication (entecavir) (Additional file 3: Table S4). The accuracy of our RF model was superior to those of AST/ALT ratio, APRI, and FIB-4 (Table 2 and Additional file 3: Table S3).

Model 3: Differentiating advanced fibrosis from early fibrosis among CLD patients

In this study, fibrosis stages 0–2 were defined as early fibrosis, and stages 3–4 were defined as advanced fibrosis. Model 3 was established based on age and the four metabolite markers selected from cohort 1 data using the RF model. It was then shown to successfully separate CLD patients with early fibrosis from those with advanced fibrosis in cohort 1 with AUROC of 0.918 (0.889–0.946) and AUPR = 0.892 (0.854–0.925) (Fig. 4f, g). Model 3 results demonstrated better classification performances than those of APRI (AUROC = 0.647, $p < 0.001$), AST/ALT (AUROC =



0.714, $p < 0.001$), and FIB-4 (AUROC = 0.802, $p < 0.001$) in predicting liver fibrosis stages (Table 2). An example decision tree from model 3 showed that the higher Tyr/Val ratio, Tyr, age, and TCA indicated a higher risk of CLD with advanced fibrosis (Additional file 2: Figure S7c).

A logit diagnostic RF score for model 3 differentiated CLD patients with early stage fibrosis from those with advanced fibrosis in cohort 1 (Fig. 4j). The model yielded a sensitivity of 86.7% and specificity of 90.5% in cohort 1 at a cutoff value of -0.115 (Table 2). After adjusting for HBV-DNA levels, degree

Table 2 Results for measurement of the metabolite marker panel, APRI, FIB-4, and AST/ALT ratio in the prediction of liver fibrosis

	Cohort 1 training set			Cohort 2 validation set		
	CLD vs. controls	Fibrosis vs. cirrhosis	S0-2 vs. S3-4	CLD vs. controls	Fibrosis vs. cirrhosis	S0-2 vs. S3-4
Metabolite marker panel						
AUROC (95% CI) [#]	0.997 (0.993-1)	0.941 (0.914-0.964)	0.918 (0.889-0.946)	0.977 (0.963-0.988)	0.844 (0.797-0.884)	0.807 (0.756-0.852)
AUPR (95% CI)	0.994 (0.986-1)	0.87 (0.824-0.913)	0.892 (0.854-0.925)	0.993 (0.989-0.997)	0.827 (0.761-0.884)	0.817 (0.764-0.866)
Cutoff value (sensitivity (%)/specificity (%)/F1 (%)) [†]	0.434 (98.4/99/98.7)	0.01 (87/90.4/78.4)	-0.115 (86.7/90.5/84.6)	0.434 (92.2/94.4/95.2)	0.01 (71.8/81.6/73.3)	-0.115 (72.9/76.1/71.8)
FIB-4						
AUROC (95% CI)	0.848 (0.823-0.87)	0.869 (0.829-0.906)	0.802 (0.762-0.844)	0.707 (0.652-0.762)	0.758 (0.692-0.815)	0.739 (0.68-0.798)
AUPR (95% CI)	0.863 (0.84-0.883)	0.725 (0.657-0.79)	0.707 (0.651-0.761)	0.897 (0.873-0.918)	0.726 (0.657-0.794)	0.726 (0.66-0.795)
Cutoff value 1 (sensitivity (%)/specificity (%)/F1 (%)) [*]			1.45 (68/73.6/62)			1.45 (81.5/42.5/68.1)
Cutoff value 2 (sensitivity (%)/specificity (%)/F1 (%)) [*]			3.25 (44.2/93.4/56.3)			3.25 (57/81.3/65.5)
APRI						
AUROC (95% CI)	0.973 (0.965-0.981)	0.698 (0.644-0.752)	0.647 (0.595-0.698)	0.879 (0.841-0.915)	0.608 (0.534-0.671)	0.595 (0.529-0.669)
AUPR (95% CI)	0.977 (0.969-0.983)	0.416 (0.345-0.497)	0.492 (0.434-0.554)	0.958 (0.942-0.972)	0.53 (0.474-0.605)	0.542 (0.488-0.614)
Cutoff value 1 (sensitivity (%)/specificity (%)/F1 (%)) ^{**}		1 (33.9/86.2/37.1)			1 (22.7/84.4/32.4)	
Cutoff value 2 (sensitivity (%)/specificity (%)/F1 (%)) ^{**}		2 (18.3/94.6/26.6)			2 (3.9/94.3/8.5)	
AST/ALT						
AUROC (95% CI)	0.665 (0.631-0.697)	0.815 (0.766-0.862)	0.714 (0.668-0.759)	0.603 (0.54-0.657)	0.684 (0.619-0.75)	0.667 (0.597-0.728)
AUPR (95% CI)	0.714 (0.685-0.747)	0.579 (0.496-0.674)	0.582 (0.516-0.654)	0.849 (0.819-0.875)	0.641 (0.573-0.721)	0.648 (0.583-0.727)
Cutoff value 1 (sensitivity (%)/specificity (%)/F1 (%)) ^{***}			0.8 (48.1/84.8/54.2)			0.8 (78.5/42.5/66.7)
Cutoff value 2 (sensitivity (%)/specificity (%)/F1 (%)) ^{***}			1 (33.1/92.8/45.1)			1 (65.9/59/63.8)
Comparison of AUROC						
Metabolite marker panel versus FIB-4 ^{****}	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	0.01
Metabolite marker panel versus APRI ^{****}	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
Metabolite marker panel versus AST/ALT ^{****}	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

[#]95% CI was calculated using 1000 times bootstrap resampling on ROC and PR curves. CI confidence interval

[†]Cutoff values were determined to maximize the sum of sensitivity and specificity for the cohort 1 training dataset

^{*}Pre-determined cutoff values of FIB-4 were used (1.45 and 3.25 to distinguish extensive fibrosis)

^{**}Pre-determined cutoff values of APRI were used (1.0 and 2.0 to distinguish cirrhosis)

^{***}Pre-determined cutoff values of AST/ALT were used (0.8 and 1.0 to distinguish extensive fibrosis)

APRI AST-to-platelet ratio index, AST/ALT aspartate transaminase/alanine transaminase ratio, FIB-4 fibrosis-4 index, AUROC area under the receiver operating characteristic curve, AUPR area under the precision-recall (PR) curve

^{****}Comparisons of AUROC between biomarker panel vs. FIB-4, AST/ALT, or APRI were performed using DeLong's test

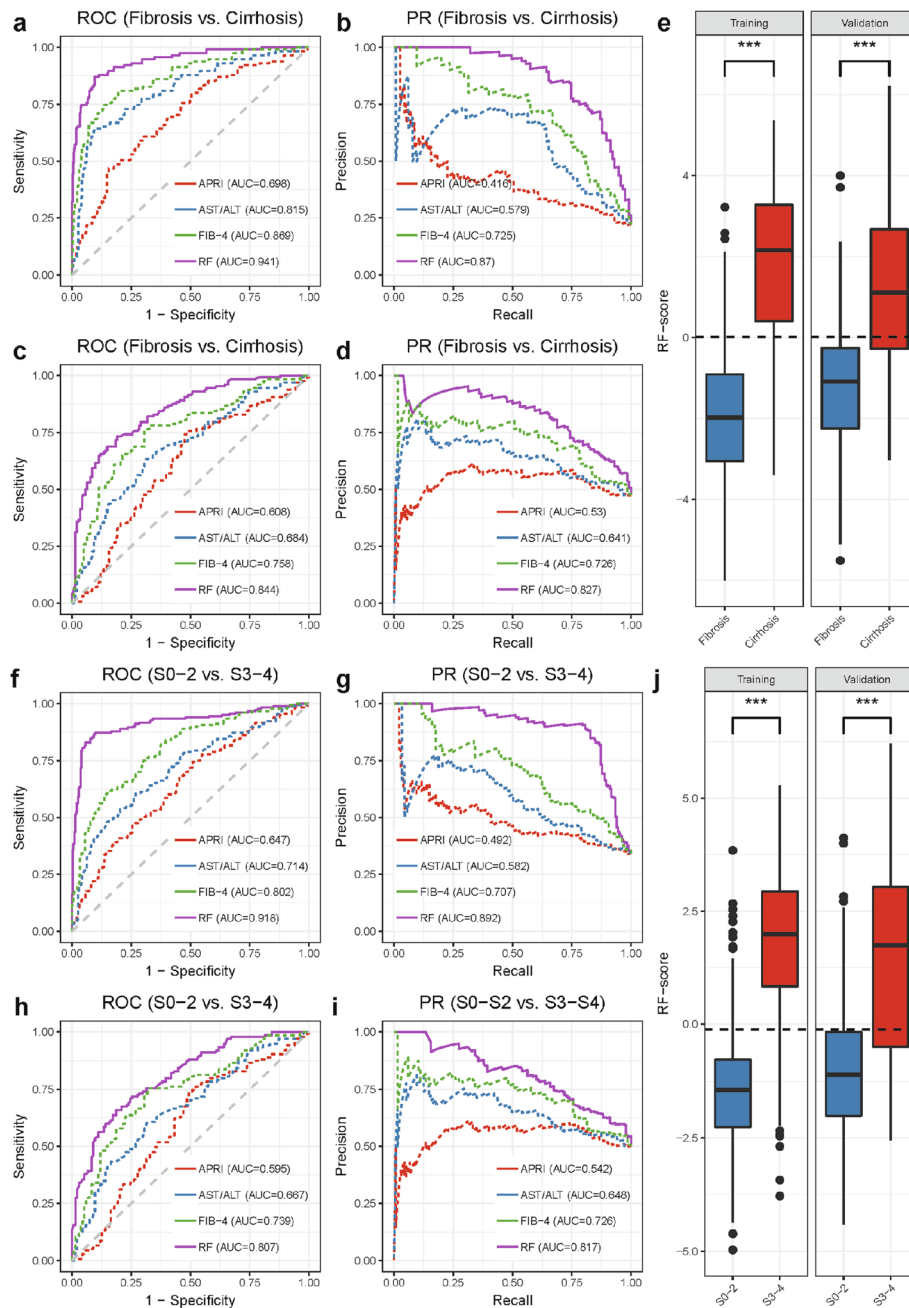


Fig. 4 Model 2 for differentiating fibrosis vs. cirrhosis and model 3 for differentiating early fibrosis vs. advanced fibrosis in CLD patients with chronic HBV infection. **a** ROC curves of model 2 (RF model constructed with four metabolite markers and age), APRI, AST/ALT, and FIB-4 in cohort 1. **b** PR curves of model 2, APRI, AST/ALT, and FIB-4 in cohort 1. **c** ROC curves for model 2, APRI, AST/ALT, and FIB-4 for the cohort 2 validation set. **d** PR curves for model 2, APRI, AST/ALT, and FIB-4 for the cohort 2 validation set. **e** The RF score in CLD patients with fibrosis and cirrhosis in the HBV training, validation sets. The optimal cutoff value of the RF score was 0.01. **f** ROC curves of model 3 (RF model constructed with four metabolite markers and age), APRI, AST/ALT, and FIB-4 in cohort 1. **g** PR curves of model 3, APRI, AST/ALT, and FIB-4 in cohort 1. **h** ROC curves for model 3, APRI, AST/ALT, and FIB-4 for the cohort 2 validation set. **i** PR curves for model 3, APRI, AST/ALT, and FIB-4 for the cohort 2 validation set. **j** The RF score in CLD patients with S0-2 and S3-4 in the HBV training, validation sets. The optimal cutoff value of the RF score was -0.115. *** $p < 0.001$, Wilcoxon's rank-sum test

of necro-inflammation, HBeAb status, HBeAg status, liver function tests (i.e., PT, ALB, DBIL, IBL), platelets, BMI, and medication (entecavir) use, RF scores

remained statistically significant with a coefficient of 0.805 ($p < 0.001$) (Additional file 3: Table S4). The accuracy of our RF model was superior to those of

AST/ALT ratio, APRI, and FIB-4 (Additional file 3: Table 2 and Table S3).

Validation of the predictive models in an independent HBV cohort (cohort 2)

The metabolite markers identified and related models obtained in cohort 1 were further validated for their liver fibrosis staging performance as well as for CLD diagnosis performance in cohort 2, and the results were similar to those obtained from cohort 1 (Table 2).

For the diagnosis of CLD patients, compared to APRI (AUROC = 0.879, AUPR = 0.958), AST/ALT (AUROC = 0.603, AUPR = 0.849), and FIB-4 (AUROC = 0.707, AUPR = 0.897), we again observed higher classification performances for model 1 with AUROC of 0.977 (0.963–0.988) and AUPR of 0.993 (0.989–0.997) in the validation set (Fig. 3e, f). In addition, the model 1 predicted RF score in cohort 2 differentiated CLD from NC with a sensitivity of 92.2% and specificity of 94.4% at the cutoff values determined for cohort 1 (Fig. 3g, Table 2).

Applying model 2 to cohort 2 successfully discriminated cirrhotic patients from fibrotic patients with an AUROC of 0.844 (0.797–0.884) and AUPR of 0.827 (0.761–0.884) (Fig. 4c, d) and outperformed those of the APRI (AUROC = 0.608, $p < 0.001$), AST/ALT (AUROC = 0.684, $p < 0.001$), and FIB-4 (AUROC = 0.758, $p < 0.001$) indices. The model 2 RF score in cohort 2 differentiated cirrhotic patients from fibrotic patients with a sensitivity of 71.8% and specificity of 81.6% at the same cutoff value used for the cohort 1 dataset (Fig. 4e, Table 2). Similarly, applying model 3 to grade fibrosis stage in cohort 2 resulted in greater performance with AUROC of 0.807 (0.756–0.852) and AUPR of 0.817 (0.764–0.866) (Fig. 4h, i) than those of APRI (AUROC = 0.595, $p < 0.001$), AST/ALT (AUROC = 0.667, $p < 0.001$), and FIB-4 (AUROC = 0.739, $p = 0.01$) indices. And the model 3 RF score also differentiated S0–2 fibrosis from S3–4 fibrosis with a sensitivity of 72.9% and specificity of 76.1% (Fig. 4j, Table 2).

We then introduced net reclassification improvement (NRI) and integral discriminant improvement (IDI) to quantify the improvement of our model to existing clinical indices. For different classification aims (control vs. CLD, fibrosis vs. cirrhosis, S0–2 vs. S3–4) in an independent validation cohort (cohort 2), the categorical and the continuous NRI and IDI of the RF models all achieved positive values when compared to FIB-4, APRI, and AST/ALT, suggesting an augmentation of classification performances for our biomarker panel and RF models (Additional file 3: Table S5).

Classification of S0–2 vs. S3 vs. S4

In addition to the binary classifications that we have performed, we further determined whether our biomarker panel could classify multiple groups among CLD

patients. We trained a new RF model with the marker panel and applied multinomial regression to APRI, AST/ALT, and FIB-4 respectively for the discrimination of S0–2 vs. S3 vs. S4 using cohort 1. We compared their performances on cohort 1 (OOB predictions of RF model) and cohort 2 using micro-average AUROC and AUPR, and we found that our marker panel-based multi-group classifier outperformed other methods. In the cohort 1 data, our classifier showed higher AUROC of 0.944 (0.928–0.963) and AUPR of 0.908 (0.883–0.938) compared to APRI (AUROC = 0.79, AUPR = 0.658), AST/ALT (AUROC = 0.817, AUPR = 0.688), and FIB-4 (AUROC = 0.858, AUPR = 0.774) (Additional file 2: Figure S8a, b). In the cohort 2 validation data, our marker panel classifier consistently displayed higher AUROC of 0.841 (0.799–0.885) and AUPR of 0.748 (0.674–0.81) compared to APRI (AUROC = 0.790, AUPR = 0.608), AST/ALT (AUROC = 0.772, AUPR = 0.597), and FIB-4 (AUROC = 0.816, AUPR = 0.699) (Additional file 2: Figure S8c, d).

Discussion

As the prevalence of CLD rises worldwide, accurate and reliable assessments for the severity of this disease are increasingly important for treatment selection and longitudinal monitoring [13]. Attempts to develop noninvasive tools for staging CLD have yielded multiple scores, indices, and imaging modalities [4–7, 10] that might be used in lieu of liver biopsy, with the AST/ALT ratio, APRI, and FIB-4 as examples [5–7]. Current noninvasive assessments have the advantage of allowing repeated applications and are well-received by the patients. In this study, we identified a panel of metabolite markers that consisted of C18:2 n6t, TCA, Tyr, and a Tyr/Val ratio that was highly correlated with discrete stages of CLD progression in patients with HBV infection.

Histologic staging of CLD by liver biopsy provided a reference standard for our study. In the Scheuer system, one of the most clinically validated systems for staging liver fibrosis, S0 is defined as no fibrosis, S1 as portal fibrosis, S2 as periportal fibrosis, S3 as septal fibrosis, and S4 as cirrhosis [20]. The clinically overt stage of cirrhosis includes compensated cirrhosis with/without portal hypertension and decompensated cirrhosis [38]. In this study, we first identified candidate markers that significantly differed between NC and patients with CLD that correlated well with fibrotic stage and necro-inflammation based on univariate, LASSO, and RF analyses. We then constructed diagnostic models to discriminate CLD patients from NC, and to discriminate CLD patients at different fibrosis stages, i.e., early vs. advanced fibrosis (S0–2 vs. S3–4) and fibrosis vs. cirrhosis (S0–3 vs. S4). This resulted in three optimized marker panel-based RF predictive models for staging liver fibrosis that, upon validation, showed acceptable

performance across independent cohort. Although the AUROCs of our models in the validation set were not as high as in the training set, we still achieved relatively good AUROC (all > 0.8) considering a rough guide for classifying the accuracy of a diagnostic test in the traditional academic point system [39]. A decreased external validation/testing accuracy was a common fact when applied machine learning in biomedical studies [40]. The AUROC and AUPR of our biomarker panel were significantly greater than those of the AST/ALT ratio, APRI, and FIB-4, suggesting superior predictive value for this metabolite marker panel.

Altered BA profile and BA synthesis are associated with various hepatic diseases, such as chronic hepatitis B, primary biliary cirrhosis, chronic hepatitis C, and NAFLD. Circulating BAs are commonly used in clinical practice to assist evaluation of the severity of CLD [41]. Several studies, including our previous work, on cirrhosis and HCC have shown dramatically increased levels of GCA, GCDCA, TCA, and TCDCA in the circulation of patients with NAFLD [42], NASH [42], HBV [43], cirrhosis [44], and HCC [44]. The liver also plays a major role in lipid metabolism by taking up FFAs and manufacturing, storing, and transporting lipid metabolites [45]. A characteristic pattern of plasma amino acids has been described in cirrhotic subjects [42, 46, 47], and in samples collected in England and the USA, metabolic and biochemical differences have been shown between stable and unstable cirrhotics [42, 46]. Advanced liver fibrosis, especially cirrhosis, was also associated with altered plasma AA patterns, including decreased levels of branched chain amino acids (leucine, isoleucine, valine) and increased concentrations of the aromatic amino acids phenylalanine and tyrosine [19]. An index based on AA concentration has already been proposed for diagnosing liver fibrosis [17]. In patients admitted to either the Veterans Administration Hospital or the Yale-New Haven Medical Center between 1 January 1965 and 1 May 1966, fasting tyrosine levels tended to be slightly increased in patients with hepatitis and markedly increased in patients with cirrhosis [48]. The present study showed that a combined panel of FFA, BA, and AA was a strong predictor for CLD progress.

Linoelaidic acid is an isomer of linoleic acid. It has been reported that linoelaidic acid may inhibit the development of tumors through its antioxidant effects, has a role in the prevention of atherosclerosis, and modulates certain aspects of immune system [49]. The significantly decreased levels of linoelaidic acid may thus be an indication of a disease state. Further research on these findings and human epidemiological data is warranted to confirm this.

The major strengths of our study were the use of large sample sizes to construct and verify all models, and the

quantification of the metabolite markers (BA, FFA, and AA) using standardized protocols. Furthermore, participants in the validation set (cohort 2) were recruited independently from those in cohort 1, and this new set of patients confirmed the robustness of our marker panel and predictive models.

The limitations of our study included the following: (1) Use of medications was a confounding factor for our model, but key findings were not altered after correcting for medication use. Larger studies are needed to further evaluate the effect of these medications; (2) HBV infection was the only or major cause of CLD in this study, and the participants were all Chinese. Therefore, the results may not be extrapolated to CLD with other etiologies outside these diseases, or to other racial/ethnic groups. Future large-scale validation studies should include CLD with other etiologies and participants of other race/ethnicity, before implementing this 4-marker panel in clinical practice. (3) In addition to cross-sectional studies, longitudinal studies are needed to further validate the reproducibility of the current findings and the predictive values of the models, especially those used to differentiate early from advanced liver fibrosis, and (4) the cost of full spectrum metabolomic analysis is high. However, if the robustness of this 4-marker panel is proven in future validation studies, specific tests may be developed for only C18:2 n6t, TCA, Tyr, and Val to decrease the cost and to translate this marker panel to clinical practice.

Conclusions

In summary, using targeted metabolomics analyses, we identified four metabolite markers from serum that accurately differentiated CLD patients from NC, and differentiated varied stages of liver fibrosis, including S0–2 vs. S3–4, and S0–3 vs. S4. The diagnostic performance of this novel, noninvasive 4-marker panel was superior to FIB-4, AST/ALT ratio, and APRI. If validated in future studies, this 4-marker panel will be useful in reducing the need for liver biopsies in identifying patients with non-significant fibrosis, as well as aiding in the continued assessment of CLD in patients previously diagnosed with CLD.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12916-020-01595-w>.

Additional file 1. Fibrosis and Cirrhosis Patients with hepatitis B viral (HBV) Infection and Normal Controls (Cohort 1). Patients with CHB-induced Fibrosis and Cirrhosis and normal controls (Validation Set, Cohort 2). Inclusion and exclusion criteria for patients with chronic HBV infection. Exclusion criteria for normal controls. Medication the patients received at the time of sampling. Quality of care. Measurement of bile acids.

Measurement of Amino Acids. Measurement of FFAs. Quality Control Procedure.

Additional file 2: Figure S1. Representative H&E staining images of chronic liver disease patients with necro-inflammation activity at G0 (A), G1 (B), G2 (C), G3 (D) to G4 (E) according to the Scheuer's classification. Scale, 200 μ m. **Figure S2.** Representative Masson's trichrome staining, collagen stained blue. Collagen portionate area increased significantly along with the degree of liver fibrosis (from S0 to S4, Table 1). Scale, 200 μ m. **Figure S3.** PCA scores plot for CLD patients and normal controls using the identified four metabolite markers in training and validation sets. **Figure S4.** Correlation coefficient matrix among the four selected serum metabolites, previously proposed liver fibrosis markers, and clinical markers of chronic liver disease (fibrosis stages, necro-inflammation, and medication). **Figure S5.** 10-fold cross-validation AUROC and AUPR of machine learning methods and clinical indices. **Figure S6.** PCA scores plot for CLD patients of S0–2, S3 and S4 using the identified four metabolite markers in training and validation sets. **Figure S7.** Example decision trees from random forest models. (a) An example decision tree of Model 1. (b) An example decision tree of Model 2. (c) An example decision tree of Model 3. **Figure S8.** Micro-ROC and micro-PR of metabolite marker panel and clinical indicators in multi-group classification of S0–2 vs. S3 vs. S4. (a) micro-ROC and (b) micro-PR for the classification of S0–2 vs. S3 vs. S4 in Cohort 1. (c) micro-ROC and (d) micro-PR for the classification of S0–2 vs. S3 vs. S4 in Cohort 2.

Additional file 3 Table S1. Clinical data of patients with chronic liver disease (CLD) and normal controls in Cohorts 1 and 2. **Table S2.** Serum bile acid, free fatty acid, and amino acid concentrations in patients with chronic liver disease (CLD) and in normal controls in Cohorts 1 and 2. **Table S3.** Results for measurement of the metabolite marker panel, APRI, FIB-4, and ALT/AST ratio in the prediction of liver fibrosis using the optimal cut-off values generated using the cohort specific data from this study. **Table S4.** Logistic regression analysis of metabolite marker panel-based RF-score to discriminate patients with fibrosis from patients with cirrhosis and S0–2 with S3–4 adjusting with potential confounding variables. **Table S5.** Net reclassification improvement and integral discriminant improvement analyses comparing RF and other clinical indexes on validation sets.

Abbreviations

CLD: Chronic liver disease; ROC: Receiver operating characteristic; PR: Precision-recall; AUROC: Area under the ROC curve; BAs: Bile acids; FFAs: Free fatty acids; AAs: Amino acids; HBV: Hepatitis B virus; CIs: Confidence intervals; C18:2 n6: Linolelaidic acid; TCA: Taurocholate; Tyr: Tyrosine; Val: Valine; ALT: Alanine transaminase; AST: Aspartate transaminase; TBIL: Total bilirubin; ALP: Alkaline phosphatase; GGT: Gamma-glutamyl transferase; ALB: Albumin; PALB: Prealbumin; TBA: Total bile acid; CHE: Cholinesterase; CREA: Creatinine; BUN: Blood urea nitrogen; CHOL: Cholesterol; TG: Triglyceride; HDLC: High-density lipoprotein cholesterol; LDLC: Low-density lipoprotein cholesterol; ApoAI: Apolipoprotein A1; ApoB: Apolipoprotein B; PT: Prothrombin Time; Fib: Fibrinogen; GLU: Glucose; RBC: Red blood cell count; WBC: White blood cell; HCT: Hematocrit; HGB: Hemoglobin; MCH: Mean corpuscular hemoglobin; MCHC: Mean corpuscular hemoglobin concentration; MPV: Mean platelet volume; PLT: Platelet; GLB: Globin

Acknowledgements

We wish to thank the research coordinators of the participating hospitals for their assistance in collecting clinical data and samples.

Authors' contributions

WJ was the principal investigator of this study and designed the study. PL organized the patient recruitment and provided the biospecimens for this study. GXX drafted the manuscript. GXX, XNW, RMW, JYW, AHZ, TLC, LW, YPD, CR, SK, HB, and XG critically revised the manuscript. GXX, RMW, JYW, XNW, AHZ, and TLC performed the data analysis. XNW, YXW, HZ, ZX, and XZL were responsible for the clinical data collection. GXX, XNW, and AHZ performed the metabolomics analysis. All authors read and approved the final manuscript.

Funding

This study was financially supported by the National Key R&D Program of China (2017YFC0906800); the National Institutes of Health/National Cancer Institute Grant 1U01CA188387-01A1; the National Natural Science Foundation of China (81974073, 81774196); the China Postdoctoral Science Foundation funded project, China (2016 T90381); and the E-Institutes of Shanghai Municipal Education Commission, China (E03008). The funding sources did not have any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Availability of data and materials

All the data supporting the findings of this study are available within the article and its supplementary information files or from the corresponding author upon reasonable request.

Ethics approval and consent to participate

The study was approved by the institutional review board of hospitals (no. 2012-206-22-01). All participants provided written informed consent.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹E-Institute of Shanghai Municipal Education Committee, Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China. ²Human Metabolomics Institute, Inc., Shenzhen 518109, Guangdong, China. ³Key Laboratory of Liver and Kidney Diseases (Ministry of Education), Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China. ⁴University of Hawaii Cancer Center, Honolulu, HI 96813, USA. ⁵Shanghai Key Laboratory of Diabetes Mellitus and Center for Translational Medicine, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai 200233, China. ⁶Department of Endocrinology and Metabolism, Zhongshan Hospital, Fudan University, Shanghai 200032, China. ⁷Institute of Liver Diseases, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, 528 Zhangheng Road, Shanghai 201203, China. ⁸School of Chinese Medicine, Hong Kong Baptist University, Kowloon Tong, Hong Kong, China.

Received: 16 February 2020 Accepted: 16 April 2020

Published online: 05 June 2020

References

- Chang TT, Liaw YF, Wu SS, Schiff E, Han KH, Lai CL, Safadi R, Lee SS, Halota W, Goodman Z, et al. Long-term entecavir therapy results in the reversal of fibrosis/cirrhosis and continued histological improvement in patients with chronic hepatitis B. *Hepatology*. 2010;52(3):886–93.
- Battaller R, Brenner DA. Liver fibrosis. *J Clin Invest*. 2005;115(2):209–18.
- Carey E, Carey WD. Noninvasive tests for liver disease, fibrosis, and cirrhosis: is liver biopsy obsolete? *Cleve Clin J Med*. 2010;77(8):519–27.
- Imbert-Bismut F, Ratziu V, Pieroni L, Charlotte F, Benhamou Y, Poinard T. Biochemical markers of liver fibrosis in patients with hepatitis C virus infection: a prospective study. *Lancet*. 2001;357(9262):1069–75.
- Park SY, Kang KH, Park JH, Lee JH, Cho CM, Tak WY, Kweon YO, Kim SK, Choi YH. Clinical efficacy of AST/ALT ratio and platelet counts as predictors of degree of fibrosis in HBV infected patients without clinically evident liver cirrhosis. *Korean J Gastroenterol*. 2004;43(4):246–51.
- Wai CT, Greenon JK, Fontana RJ, Kalbfleisch JD, Marrero JA, Conjeevaram HS, Lok ASF. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology*. 2003;38(2):518–26.
- Terling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, M SS, Torriani FJ, Dieterich DT, Thomas DL, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology*. 2006;43(6):1317–25.
- Toshima T, Shirabe K, Ikegami T, Yoshizumi T, Kuno A, Togayachi A, Gotoh M, Narimatsu H, Korenaga M, Mizokami M, et al. A novel serum marker,

- glycosylated Wisteria floribunda agglutinin-positive Mac-2 binding protein (WFA(+)-M2BP), for assessing liver fibrosis. *J Gastroenterol.* 2015;50(1):76–84.
9. Wei R, Wang J, Wang X, Xie G, Wang Y, Zhang H, Peng CY, Rajani C, Kwee S, Liu P, et al. Clinical prediction of HBV and HCV related hepatic fibrosis using machine learning. *EBioMedicine.* 2018;35:124–32.
 10. Brancatelli G, Federle MP, Ambrosini R, Lagalla R, Carriero A, Midiri M, Vilgrain V. Cirrhosis: CT and MR imaging evaluation. *Eur J Radiol.* 2007;61(1):57–69.
 11. Meng F, Zheng Y, Zhang Q, Mu X, Xu X, Zhang H, Ding L. Noninvasive evaluation of liver fibrosis using real-time tissue elastography and transient elastography (FibroScan). *J Ultrasound Med.* 2015;34(3):403–10.
 12. Morikawa H. Real-time tissue elastography and transient elastography for evaluation of hepatic fibrosis; 2012.
 13. Chen T, Xie G, Wang X, Fan J, Qiu Y, Zheng X, Qi X, Cao Y, Su M, Xu LX, et al. Serum and urine metabolite profiling reveals potential biomarkers of human hepatocellular carcinoma. *Mol Cell Proteomics.* 2011;10(7):M110004945.
 14. Shlomai A, Halfon P, Goldiner I, Zelber-Sagi S, Halpern Z, Oren R, Bruck R. Serum bile acid levels as a predictor for the severity of liver fibrosis in patients with chronic hepatitis C. *J Viral Hepat.* 2013;20(2):95–102.
 15. Wang X, Xie G, Zhao A, Zheng X, Huang F, Wang Y, Yao C, Jia W, Liu P. Serum bile acids are associated with pathological progression of hepatitis B-induced cirrhosis. *J Proteome Res.* 2016;15(4):1126–34.
 16. Zhang JW, Zhao Y, Xu CF, Hong YN, Lu HL, Wu JP, Chen Y. Association between serum free fatty acid levels and nonalcoholic fatty liver disease: a cross-sectional study. *Sci Rep.* 2014;4:6.
 17. Zhang Q, Takahashi M, Noguchi Y, Sugimoto T, Kimura T, Okumura A, Ishikawa T, Kakumu S. Plasma amino acid profiles applied for diagnosis of advanced liver fibrosis in patients with chronic hepatitis C infection. *Hepatol Res.* 2006;34(3):170–7.
 18. Chinese Society of Hepatology and Chinese Society of Infectious Diseases, Chinese Medical Association. The guideline of prevention and treatment for chronic hepatitis B (2010 version). *Chin J Hepatol* 2011, 19(1):13–24.
 19. Campolito O, Sprengers D, McIntyre N. The BCAA/AAA ratio of plasma amino acids in three different groups of cirrhotics. *Rev Invest Clin.* 1992;44(4):513–8.
 20. Scheuer PJ, Standish RA, Dhillon AP. Scoring of chronic hepatitis. *Clin Liver Dis.* 2002;6(2):335–47 v-vi.
 21. M DG, M KM, M E-BH, M FA-M, SHARAF E-DOA. Digital quantification of fibrosis in liver biopsy sections: description of a new method by Photoshop software. *J Gastroenterol Hepatol.* 2004;19(1):78–85.
 22. Xie G, Wang Y, Wang X, Zhao A, Chen T, Ni Y, Wong L, Zhang H, Zhang J, Liu C, et al. Profiling of serum bile acids in a healthy Chinese population using UPLC-MS/MS. *J Proteome Res.* 2015;14(2):850–9.
 23. Ni Y, Zhao L, Yu H, Ma X, Bao Y, Rajani C, Loo LW, Shvetsov YB, Yu H, Chen T, et al. Circulating unsaturated fatty acids delineate the metabolic status of obese individuals. *EBioMedicine.* 2015;2(10):1513–22.
 24. Chen T, Ni Y, Ma X, Bao Y, Liu J, Huang F, Hu C, Xie G, Zhao A, Jia W, et al. Branched-chain and aromatic amino acid profiles and diabetes risk in Chinese populations. *Sci Rep.* 2016;6:20594.
 25. Xie G, Zhong W, Li H, Li Q, Qiu Y, Zheng X, Chen H, Zhao X, Zhang S, Zhou Z, et al. Alteration of bile acid metabolism in the rat induced by chronic ethanol consumption. *FASEB J.* 2013;27(9):3583–93.
 26. Grau J, Grosse I, Keilwagen J. PPRoc: computing and visualizing precision-recall and receiver operating characteristic curves in R. *Bioinformatics.* 2015;31(15):2595–7.
 27. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. pRoc: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011;12:77.
 28. Kundu S, Aulchenko YS, van Duijn CM, Janssens ACJW. PredictABEL: an R package for the assessment of risk prediction models. *Eur J Epidemiol.* 2011;26(4):261–4.
 29. Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B.* 1996;58:267–88.
 30. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32.
 31. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33(1):1–22.
 32. Liaw A, Wiener M. Classification and regression by randomForest. *R News.* 2002;2(3):18–22.
 33. Breiman L. Bagging predictors. *Mach Learn.* 1996;24(2):123–40.
 34. McPherson S, Stewart SF, Henderson E, Burt AD, Day CP. Simple non-invasive fibrosis scoring systems can reliably exclude advanced fibrosis in patients with non-alcoholic fatty liver disease. *Gut.* 2010;59(9):1265–9.
 35. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol.* 1995;57(1):289–300.
 36. Asch W: Macro-and micro-averaged evaluation measures [[BASIC DRAFT]]. In: 2013; 2013.
 37. Wei RM, Wang JY, Jia W: multiROC: calculating and visualizing ROC and PR curves across multi-class classifications. In., 2018-06-26 edn; 2018.
 38. D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. *J Hepatol.* 2006;44(1):217–31.
 39. Mehdi T, Ahmadi B. Kernel smoothing for ROC curve and estimation for thyroid stimulating hormone. *Int J Public Health Res.* 2011;Special Issue: 239–42.
 40. Way GP, Sanchez-Vega F, La K, Armenia J, Chatila WK, Luna A, Sander C, Cherniack AD, Mina M, Ciriello G, et al. Machine learning detects pan-cancer Ras pathway activation in the cancer genome atlas. *Cell Rep.* 2018;23(1):172–80 e173.
 41. Asgharpour A, Kumar D, Sanyal A. Bile acids: emerging role in management of liver diseases. *Hepatol Int.* 2015;9(4):527–33.
 42. Kalhan SC, Guo L, Edmison J, Dasarthy S, McCullough AJ, Hanson RW, Milburn M. Plasma metabolomic profile in nonalcoholic fatty liver disease. *Metabolism.* 2011;60(3):404–13.
 43. Lian JS, Liu W, Hao SR, Guo YZ, Huang HJ, Chen DY, Xie Q, Pan XP, Xu W, Yuan WX, et al. A serum metabolomic study on the difference between alcohol- and HBV-induced liver cirrhosis by ultra-performance liquid chromatography coupled to mass spectrometry plus quadrupole time-of-flight mass spectrometry. *Chin Med J.* 2011;124(9):1367–73.
 44. Yin PY, Wan DF, Zhao CX, Chen J, Zhao XJ, Wang WZ, Lu X, Yang SL, Gu JR, Xu GW. A metabolomic study of hepatitis B-induced liver cirrhosis and hepatocellular carcinoma by using RP-LC and HILIC coupled with mass spectrometry. *Mol BioSyst.* 2009;5(8):868–76.
 45. Berlanga A, Guiu-Jurado E, Porras JA, Auguet T. Molecular pathways in non-alcoholic fatty liver disease. *Clin Exp Gastroenterol.* 2014;7:221–39.
 46. Marchesini G, Bianchi GP, Vilstrup H, Checchia GA, Patrono D, Zoli M. Plasma clearances of branched-chain amino acids in control subjects and in patients with cirrhosis. *J Hepatol.* 1987;4(1):108–17.
 47. Dam G, Sorensen M, Buhl M, Sandahl TD, Moller N, Ott P, Vilstrup H. Muscle metabolism and whole blood amino acid profile in patients with liver disease. *Scand J Clin Lab Invest.* 2015;75(8):674–80.
 48. Levine RJ, Conn HO. Tyrosine metabolism in patients with liver disease. *J Clin Invest.* 1967;46(12):2012–20.
 49. MacDonald HB. Conjugated linoleic acid and disease prevention: a review of current knowledge. *J Am Coll Nutr.* 2000;19(2 Suppl):111s–8s.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

