Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Non-Invasive Evaluation of MAFLD and the Contribution of Genes: An MRI-PDFF-Based Cross-Sectional Study

Aruhan Yang
Jilin University
Xiaoxue Zhu
Jilin University
Lei Zhang
Jilin University
Yingwen Zhang
Jilin University
Dezhi Zhang
Jilin University
Junqi Niu
Jilin University
Huimao Zhang
Jilin University
Yanhua Ding (dingyanhua2003@126.com)
Jilin University https://orcid.org/0000-0002-0498-8939
Guoyue Lv
Jilin University

Research Article

Keywords: Non-alcoholic fatty liver disease (NAFLD), Metabolic fatty liver disease (MAFLD), MRI-proton-density-fat-fraction (PDFF), Liver fat Content, Single-nucleotide polymorphisms (SNP), Odds Ratios, Nonalcoholic Steatohepatitis, Liver Steatosis, Liver Dysfunction, Digestive System Disease, Hepatobiliary Disorders

Posted Date: March 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1428925/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Objective

To investigate the clinical, laboratory and genetic features of MAFLD patients based on MRI-PDFF and liver biopsy in China.

Design:

Patients with high ALT and with a diagnosis of fatty liver were included in this cross-sectional study. Fasting blood was collected to test biomarkers and SNPs. A total of 266 patients underwent MRI-PDFF and FibroScan examinations, and 38 underwent liver biopsy. Diagnostic models (decision tree, LASSO, and elastic net) were developed based on diagnosis from MRI-PDFF reports.

Results

Approximately 1/3 were found to have NASH and fibrosis. After quantifying liver steatosis by MRI-PDFF (healthy: N = 47; mild MAFLD: N = 136; moderate/severe MAFLD: N = 83; liver fat content (LFC): 3.6% vs. 8.7% vs. 19.0%), most biomarkers showed significant differences among the three groups, and patients without obesity were found to have a similar LFC as those with obesity (11.1% vs. 12.3%). Variant alleles of PNPLA3, HSD17B13 and MBOAT7 were identified as genetic risk factors causing higher LFC (8.7% vs. 12.3%; 11.0% vs. 14.5%; 8.5% vs. 10.2%, $P < 0.05$); those with the UQCC1 rs878639 variant allele showed lower LFC (10.4% vs. 8.4%; OR = 0.58, $P < 0.05$). Patients with more risk alleles had higher LFCs (8.1% vs. 10.7% vs. 11.6% vs. 14.5%). Models including biomarkers showed strong diagnostic ability (accuracy: 0.80–0.91).

Conclusions

When the effects of genes on liver steatosis were first quantified by MRI-PDFF, the UQCC1 rs878639 G allele was identified as a protective factor, and the MBOAT7 T allele was identified as a risk only among nonobese individuals. Machine learning-based models can accurately predict disease status.

Key Messages

- **What is already known on this topic** – *Compared with other imaging modalities, MRI-PDFF has the highest diagnostic accuracy for the quantification of liver steatosis in MAFLD.*
- **What this study adds** – *When the effects of gene polymorphisms on liver steatosis were first quantified using MRI-PDFF, the UQCC1 rs878639 G allele was identified as a protective factor, and the*

MBOAT7 T allele was identified as a risk factor only among nonobese individuals; machine learning-based models could more accurately predict the disease status.

- **How this study might affect research, practice or policy** – *The* UQCC1 *rs878639 G allele may be a new therapeutic target; tree models will be able to handily predict the disease status in the clinic more genetic risk factors and relevant serum biomarkers will be identified in MAFLD based on MRI-PDFF.*

1 Introduction

With the global obesity epidemic and the unprecedented rates of metabolic abnormalities, nonalcoholic fatty liver disease (NAFLD) has become a major cause of chronic liver disease worldwide, especially in China, and the incidence of fatty liver disease is highest in China, surpassing the incidence in European countries and the USA.[1] NAFLD has a wide histological spectrum, from simple steatosis and nonalcoholic steatohepatitis (NASH) to cirrhosis and hepatocellular carcinoma (HCC) [2]. An international panel of experts and the Chinese Society of Hepatology have endorsed a proposal to change the term NAFLD to metabolic-associated fatty liver disease (MAFLD) to improve health care and disease awareness.[3, 4] Fatty liver can also be found in lean patients with insulin resistance (IR) and atherogenic dyslipidaemia, approximately 44.3% of MAFLD patients in China are nonobese. [5] Due to the differences in race, economy, and lifestyle between China and other countries, clinical, genetic and biomarker features in Chinese MAFLD patients need to be further investigated.

Many studies have demonstrated that the risk of liver-related mortality increases exponentially with an increase in fibrosis stage.[6, 7] However, a recent study has demonstrated that all NAFLD histological stages, including steatosis, are associated with significantly increased overall mortality.[8] Liver biopsy remains the diagnostic gold standard for MAFLD, but the fat content cannot be quantified.[9] Some noninvasive imaging diagnosis methods can assess the liver fat content (LFC), such as ultrasonography (US), computed tomography (CT), controlled attenuation parameter (CAP) and magnetic resonance imaging-derived proton density fat fraction (MRI-PDFF). Compared with other imaging modalities, MRI-PDFF has the highest diagnostic accuracy for the quantification of LFC and is commonly used in NASH trials.[10] Few studies have discussed the clinical characteristics of adult MAFLD based on MRI-PDFF, [11–13] and limited biomarkers have been investigated.[14–17]

Serum biomarkers such as cytokeratin 18 (CK18), fibroblast growth factor 19 (FGF19), and tissue inhibitor matrix metalloproteinase 1 (TIMP1) showed a strong correlation with disease severity.[18, 19] Some molecular prediction models based on serum biomarkers have been constructed and used to predict the developmental risk of MAFLD and predict the hepatic fat content[20, 21]. However, the evaluation ability of most algorithms has not been verified in Chinese patients based on MRI-PDFF. Furthermore, to develop new models, we used machine learning methods (LASSO, Elastic Net and Decision Tree) as alternative approaches to logistic regression to handle high-dimensionality datasets obtained from demographic, clinical, and laboratory data.

MAFLD has a heritable component, with genetic differences between individuals influencing disease risk estimates by 20–70%.[22] Thus, potential risk single nucleotide polymorphisms (SNPs) need to be further investigated. Several SNPs were identified in Chinese individuals, such as PNPLA3, TM6SF2, and KLB, [23–25] while the diagnostic criterion used in those studies was US, whose accuracy is limited by both interobserver and intraobserver variability;[18] thus, a more reliable assessment method is needed as a diagnostic criterion.

The aims of this study were to investigate the clinical, laboratory and genetic features of MAFLD patients based on MRI-PDFF and liver biopsy from a Chinese population and to differentiate MAFLD patients by serum biomarkers through machine learning methods.

2 Patients And Methods

2.1 Patients

From January 2020 to March 2021, a total of 301 participants were recruited, Participants with missing data were excluded. This study was conducted in strict compliance with the Declaration of Helsinki. This study was approved by the Ethics Committee of the First Hospital of Jilin University (Ethical Approval Number: 19K096001), and written informed consent was obtained from all participants. Participants with previous ALT elevation and fatty liver indicated by ultrasound were included. Subjects with the following conditions were excluded: (1) excess alcohol consumption (males > 210 g/w, females > 140 g/w) or cigarettes; (2) accompanying conditions of viral hepatitis, drug-induced hepatitis, autoimmune hepatitis, or other factors inducing chronic liver diseases; and (3) suffering from tumours or surgery in the past 2 years.

2.2 Clinical parameter collection

The basic information of each participant, such as their name, age, and major medical histories, was collected by a standard questionnaire. Subjects with excess alcohol intake and other liver-related diseases were excluded from this study. Blood samples were taken to assess liver and metabolism status and to exclude patients with viral hepatitis and other liver diseases. Each participant was subjected to an overnight fast, and venous blood was sampled. Each participant was subjected to MRI-PDFF, FibroScan and liver biopsy examination (Supplementary Materials).

2.7 Serum Biomarker Tests and Genotyping

The values of the FLI, HSI, FIB-4 and FAST models were calculated based on serum and anthropometric results. Serum FGF19, FGF21, insulin-like growth factor-1 (IGF-1), gamma interferon-inducible protein 10 (IP-10), interleukin-6 (IL-6), TIMP1 and CK18 were quantified by ELISA (R&D Systems or Peviva) following the manufacturer's instructions. Nine SNPs were analysed, and their loci are shown in Supplementary Table 1. All SNPs except one were in Hardy-Weinberg equilibrium[26] (PNPLA3 rs738409, Hardy-Weinberg equilibrium, $P < 0.05$, Supplementary materials).

2.8 Metabolic status definition

Metabolic syndrome was defined according to the updated criteria of the Adult Treatment Panel III and International Diabetes Federation (IDF).[27] Diagnosed diabetes was defined as a self-reported diagnosis that was determined previously by a health care professional. Undiagnosed diabetes was defined according to the American Diabetes Association 2010 criteria.[28]

2.9 Statistical analysis

According to the MRI-PDFF results, volunteers were categorized into three groups: the healthy group, mild MAFLD group and moderate/severe MAFLD group. Qualitative data composition ratios were compared using chi-square analysis. One-way ANOVA was used for pairwise comparisons. Through machine learning, the least absolute shrinkage and selection operator (LASSO) and elastic net method were used to build the model.[29, 30] They were built and tested based on a training dataset (73% of the whole data) and a testing dataset (27% of the whole data), and their performance was compared by the area under the receiver operating characteristic curve (AUC). Optimal cut-off values were determined and assessed by the sensitivity, specificity. A decision tree visualizes the classification or regression results by a tree structure. Twenty-seven variables were incorporated into the decision tree analysis. Spearman's rank correlation test was used to assess the correlations. Logistic regression was applied to calculate the odds ratios. All statistical analyses were performed in R 4.0.5. Methods of developing models through machine learning (LASSO, Elastic Net and Decision tree) were introduced in Supplementary Methods.

3 Results

3.1 Histological results of 38 patients

In total, 38 volunteers underwent liver biopsy (Supplementary Table 2), all of whom were diagnosed with MAFLD; 13 (34%) were diagnosed with NASH, whose NASs were over 5. When they were grouped by NASH status (Table 1), CK18, CAP and LFC were higher in the NASH group (164.1 U/L vs. 377.1 U/L; 286 dB/m vs. 346 dB/m; 8.7% vs. 15.8%; all $P < 0.05$). ALT and AST also showed increasing trends in NASH. When patients were grouped by steatosis degree, a significant difference in LFC by MRI-PDFF was found among the three groups (7.2% vs. 15.1% vs. 27.1%). With the aggravation of inflammation and steatosis, higher CK18, ALT, and AST levels were found in the more severe groups.

Table 1
Clinical and laboratory results of patients grouped by NASH or non-NASH.

Variable	Non-NASH (n = 25)	NASH (n = 12)	p value
Basic Information			
Age	46 (10)	36 (19)	0.1826
BMI	28.9 (3.8)	29.45 (4.95)	0.6149
WC	97 (8.7)	96.45 (8.1)	0.9353
AC	97.9 (9.8)	100.15 (7.1)	0.5163
HC	101.4 (10.3)	102.35 (5.22)	0.7827
SBP	127 (16)	123.5 (15)	0.3547
DBP	84 (13)	80.5 (15.25)	0.4652
Clinical Data			
ALT	43 (51.5)	65 (32.85)	0.0556
AST	31 (19.3)	40.5 (25.5)	0.0717
UA	332 (158)	446 (106)	0.1534
TG	2.07 (1.23)	2.08 (3.59)	0.4265
TC	5.5 (1.4)	5.35 (0.94)	0.3988
LDL	3.3 (1.57)	2.6 (0.99)	0.178
HDL	1.17 (0.2)	1.1 (0.26)	0.2608
TB	14.1 (7.6)	16.5 (7.3)	0.592
TBA	2.5 (1.6)	2.55 (0.88)	0.4549
Ferritin	149.8 (168.9)	237.2 (230.8)	0.4754
Insulin	89.76 (113.1)	108.65 (53.29)	0.8711
Serum Biomarkers			
Adiponectin	42.07 (37.17)	29.18 (23.43)	0.173
IP-10	126 (49.7)	140.75 (59.88)	0.3551
leptin	115.4 (158.29)	78.9 (41.91)	0.0577
CK18*	164.1 (311.41)	377.05 (359.48)	0.0179

Caption: The table compared the difference between the NASH and no-NASH group in the clinical, laboratory indicators, *P < 0.05.

Variable	Non-NASH (n = 25)	NASH (n = 12)	p value
FGF21	352.1 (195.9)	387.8 (326.63)	0.697
FGF19	143 (85.7)	111.65 (24.25)	0.1356
FibroScan, MRI-PDFF			
CAP*	286 (57)	346.5 (44.75)	0.0496
LSM	5.5 (1.7)	5.35 (3)	0.9482
LFCmean*	8.7 (9.1)	15.75 (9.43)	0.0067
Caption: The table compared the difference between the NASH and no-NASH group in the clinical, laboratory indicators, *P < 0.05.			

Based on the steatosis degree (0–3) of liver biopsy, ordinal logistic regression analysis showed that the LFC (beta = 0.414, P = 0.006) assessed by MRI-PDFF can better predict the fat content of the liver than the CAP (beta=-0.003, P = 0.895). Furthermore, Spearman's correlation (Supplementary Fig. 2.) showed a stronger correlation of the LFC with steatosis and NAS score (R = 0.76, R = 0.66, respectively; both P < 0.05) than that of the CAP (R = 0.56, P = 0.0022; R = 0.39, P = 0.014, respectively).

3.2 Noninvasive assessment of subjects grouped by LFC

3.2.1 Clinical features of 266 subjects grouped by LFC

A total of 266 participants with MAFLD with previous ALT elevation and a diagnosis of fatty liver were included. According to the MRI-PDFF results, participants were divided into three groups: healthy group (n = 47, 18%; LFC median = 3.6%), mild MAFLD group (n = 136, 51%; LFC median = 8.7%) and moderate/severe MAFLD group (n = 83, 31%; LFC median = 19.0%). The characteristics and disease composition of the three groups are shown in Table 2. The proportion of male participants was higher than that of female participants in our whole cohort (190 vs. 76). No significant differences were found in age among the three groups.

Table 2
Clinical and laboratory results of 266 patients grouped by MRI-PDFF.

Variable	Healthy (n = 47)	Mild (n = 136)	Mod-Sev (n = 83)	p value
Basic Information				
T2DM, %	10.6	18.4	19.3	0.021
MetS, %	51.1	80.1	80.3	0.1123
age	42 (10.5)	40 (15)	36 (19.5)	0.0299
WC	91 (12.45)	96.9 (13.35)	97.5 (10.7)	0.0005
AC	94.3 (9.35)	97.6 (10.75)	99 (11.4)	0.0009
HC	98.5 (7.05)	101.85 (9.82)	104.3 (8.15)	0.0023
BMI	27.3 (4.05)	28 (4.72)	28.3 (4.1)	0.007
Clinical Data				
ALT	28 (14.25)	49.65 (34.22)	68 (40.55)	< 0.0001
AST	25 (8.55)	33 (14.8)	41 (21.3)	< 0.0001
GGT	34 (30.5)	48.35 (36.12)	54.3 (44.15)	< 0.0001
LDL	3.02 (1.38)	3.29 (1.2)	3.05 (1.15)	0.1122
HDL	1.1 (0.3)	1.08 (0.24)	1.07 (0.27)	0.2001
TC	4.7 (1.28)	5.22 (1.33)	5.18 (1.12)	0.0871
TG	1.55 (1.01)	2.18 (1.33)	2.52 (2.3)	< 0.0001
UA	368 (124.5)	422 (129.25)	423 (116.5)	0.0046
Insulin	64.11 (50.37)	105.85 (70.04)	134.6 (101.62)	< 0.0001
Ferritin	153.5 (90.5)	215.55 (278.7)	290.7 (222.85)	< 0.0001
TB	13 (5.65)	15 (6.07)	17 (9.6)	0.0037
TBA	2.3 (2.35)	2.45 (1.8)	2.1 (1.65)	0.3031
Serum Biomarkers				
Adiponectin	30.76 (33.17)	28.63 (25.14)	24.17 (20.06)	0.2114
IP-10	84 (38.68)	119.35 (65.04)	125.8 (34)	< 0.0001
IL-6	1.12 (0.89)	1.56 (1.21)	1.81 (1.47)	0.0009

Caption: The table compares the characteristics between three groups when subjects are categorised by MRI-PDFF results.

Variable	Healthy (n = 47)	Mild (n = 136)	Mod-Sev (n = 83)	p value
TIMP1	155 (66.69)	176.75 (66.28)	190 (67.55)	0.0026
Leptin	68.86 (119.13)	81.16 (115.05)	91.12 (124.8)	0.3909
CK18	112.5 (129.9)	191.9 (180.75)	337.1 (272.4)	< 0.0001
FGF21	255.8 (164.05)	320 (237.25)	430.3 (302.6)	< 0.0001
FGF19	158.1 (128.52)	130.35 (84.83)	122.4 (94.88)	0.0492
FLI	8.29 (8.25)	21.24 (33.16)	30.39 (35.03)	< 0.0001
HSI	37.33 (6.87)	40.97 (6.38)	42.39 (7.11)	< 0.0001
FAST	0.11 (0.07)	0.25 (0.24)	0.44 (0.24)	< 0.0001
APRI	0.19 (0.07)	0.26 (0.11)	0.32 (0.16)	< 0.0001
FibroScan, MRI-PDFF				
CAP	260 (44)	285.5 (53)	341 (49)	< 0.0001
LSM	4.7 (1.2)	5.6 (2)	6.4 (2.1)	< 0.0001
LFCmean	3.6 (1.4)	8.65 (3.95)	19 (8.35)	< 0.0001
Caption: The table compares the characteristics between three groups when subjects are categorised by MRI-PDFF results.				

BMI was significantly lower in the healthy group. The incidence of T2DM in the healthy group (10.6%) was lower than those in the mild MAFLD group (18.4%) and the moderate/severe MAFLD group (19.3%), and the incidence of metabolic syndrome also showed a similar result (51.1% vs. 80.1% vs. 80.3%). Markers associated with glucose, lipid, purine and iron metabolism were significantly different among the healthy and MAFLD groups (insulin: 64.1 pmol/ml vs. 105.9 pmol/ml vs. 134.6 pmol/ml; TG: 1.55 mmol/L vs. 2.18 mmol/L vs. 2.52 mmol/L; UA: 368 μmol/L vs. 422 μmol/L vs. 423 μmol/L; and ferritin 153.5 ng/ml vs. 215.5 ng/ml vs. 290.7 ng/ml; all $P<0.05$), and even between the mild and moderate/severe groups, indicating a close relationship between MAFLD and metabolism.

In addition, markers of liver injury, such as ALT (healthy: 33 U/L, mild MAFLD: 49 U/L, and moderate/severe MAFLD: 70 U/L; $P<0.001$), AST (healthy: 26 U/L, mild MAFLD: 31 U/L, and moderate/severe MAFLD: 45 U/L; $P<0.001$) and GGT (healthy: 34 U/L, mild MAFLD: 46 U/L, and moderate/severe MAFLD: 42 U/L; $P<0.001$), were higher in the MAFLD group, especially in the moderate/severe fatty liver group.

3.2.2 Serum biomarkers in different MAFLD groups

Among the serum biomarkers that were explored and showed significant differences in the three groups, CK18, an indicator of hepatocyte apoptosis, was the highest in the moderate/severe group (112.5 pg/ml vs. 191.9 pg/ml vs. 337.1 pg/ml, P<0.001). FGF21 showed a significant difference among the three groups (healthy: 255.8 pg/ml, mild MAFLD: 320.0 pg/ml, and moderate/severe MAFLD: 430.3 pg/ml; P<0.05). Adiponectin, leptin and FGF19, which are protective factors, showed no significant differences among the three groups. Markers of inflammation (IL-6, IP-10) showed higher serum concentrations in the MAFLD group and remained significant between the mild and moderate/severe MAFLD groups. Predicting models (FAST, HSI, and FLI) that were used to predict liver steatosis were significantly different among the three groups. The FAST (healthy: 0.11, mild MAFLD: 0.25, and moderate/severe MAFLD: 0.44; P<0.001), HSI (healthy: 37.3, mild MAFLD: 41.0, and moderate/severe MAFLD: 42.4; P<0.05), FLI (healthy: 8.3, mild MAFLD: 21.2, and moderate/severe MAFLD: 30.4; P<0.001) all performed well.

3.2.3 Features of nonobese MAFLD and obese MAFLD

To investigate the clinical characteristics between obese and nonobese MAFLD groups, 219 MAFLD patients were categorized according to BMI (< 28 kg/m^2, nonobese MAFLD; ≥ 28 kg/m^2, obese MAFLD, Table 3). Significantly higher WC, HC, AC and BMI were found in the obese group, but the LFC and CK18 and ALT levels were found to be similar. A lower FGF19 level was associated with higher insulin and leptin concentrations in the obese group (144.2 pg/ml vs. 121.9 pg/ml, P = 0.023 for FGF19; 86.5 pmol/L vs. 140.5 pmol/L for insulin; and 58.9 pg/ml vs. 128.1 pg/ml for leptin; all P < 0.001), indicating a worse metabolic status in obese patients. Perhaps, more severe inflammation existed in the obese group, in which IP-10 and IL-6 levels were slightly higher. In the nonobese MAFLD group (N = 101), 96 (95%) patients were metabolically unhealthy, suggesting the role of metabolic disorders in the pathogenesis of MAFLD even if patients are not obese.

Table 3
Clinical and laboratory results of MAFLD patients grouped by BMI.

Variable	Nonobese (n = 101)	Obese (n = 118)	p value
Basic Information			
Age	38 (16)	39 (17.75)	0.864
WC	91.7 (6.5)	102.7 (10.38)	< 0.0001
AC	93.2 (7.5)	103.25 (8.98)	< 0.0001
HC	97.9 (5.8)	106.55 (8)	< 0.0001
BMI	26 (2.2)	30.6 (3.25)	< 0.0001
Clinical Data			
ALT	53 (37)	56 (40.75)	0.8852
AST	34 (18)	36.4 (17.75)	0.463
GGT	48.4 (42)	52 (46.62)	0.6617
LDL	3.14 (1.26)	3.3 (1.15)	0.352
HDL	1.07 (0.22)	1.09 (0.29)	0.5917
TC	5.39 (1.37)	5.14 (1.19)	0.3542
TG	2.4 (1.69)	2.22 (1.51)	0.2896
UA	425 (119)	412.5 (136)	0.6333
Insulin	93.8 (61.24)	141.15 (102.2)	< 0.0001
Ferritin	247.2 (269.3)	256.1 (254.18)	0.7982
TB	16 (6)	15.5 (9.23)	0.9821
TBA	2.4 (1.6)	2.2 (1.77)	0.3307
Serum Biomarkers			
Adiponectin	26.42 (20.65)	28.56 (24.06)	0.7467
IP-10	115.1 (47.9)	127.4 (60.25)	0.0557
IL-6	1.4 (1.01)	1.83 (1.35)	0.0001
TIMP1	169.2 (62.73)	190.15 (69.07)	0.0091
Leptin	58.9 (58.1)	128.1 (138.66)	< 0.0001
CK18	234.4 (238.8)	246.65 (234.6)	0.7499

Caption: The table compares the characteristics between non-obese and obese MAFLD

Variable	Nonobese (n = 101)	Obese (n = 118)	p value
FGF21	376 (275.3)	350.2 (252.77)	0.8406
FGF19	143.8 (111.1)	118.7 (78.89)	0.0028
FLI	15.34 (19.46)	37.92 (38.46)	< 0.0001
HSI	39.51 (4.93)	44.09 (6.01)	< 0.0001
FAST	0.27 (0.24)	0.39 (0.29)	0.0015
APRI	0.26 (0.14)	0.28 (0.14)	0.4634
FibroScan, MRI-PDFF			
CAP	295 (73)	315 (65.5)	0.0202
LSM	5.1 (1.9)	6.45 (2.57)	< 0.0001
LFCmean	11.1 (8)	12.3 (8.7)	0.294
Caption: The table compares the characteristics between non-obese and obese MAFLD			

3.3 Correlation analysis

As demonstrated in Fig. 1A and B, among the clinical parameters, ALT (r = 0.52) showed the strongest correlation with the LFC, followed by insulin and AST, with correlation coefficients of 0.42 and 0.47, respectively (P < 0.05). Among the serum biomarkers, CK18 was most strongly associated with LFC (r = 0.57, P < 0.05). FGF21 was next with a correlation of 0.39 (P < 0.05), whereas FGF19 was not strongly correlated (r=-0.16, P > 0.05). The FAST, CAP, HSI, and FLI models, which could predict liver steatosis, showed it strong positive correlations with the LFC (r = 0.64, r = 0.63, r = 0.39, and r = 0.39, respectively; all P < 0.05).

3.4 Classification tree to distinguish different degrees of liver steatosis

The classification tree in Fig. 4A and B indicated the probability of different degrees of MAFLD in our cohort. We used a classification tree algorithm to identify biomarkers that could be used to predict MAFLD. The importance of variables in the four classification tree models is shown, and only the top five most important variables are shown (Supplementary Fig. 3A and B). After testing, ALT and TC were more important variables for differentiating the MAFLD group from the healthy group (accuracy: 81.1%). When ALT and TC levels were both lower than their thresholds (N = 17), 87.5% of the subjects (N = 14) had MAFLD and an LFC over 5%. Among the patients with MAFLD, CAP combined with AST performed better in distinguishing patients with mild MAFLD from those with moderate/severe MAFLD (accuracy: 80.3%).

When the CAP value and AST level were both higher than their thresholds (306.5 dB/m and 46.5 U/L, respectively, N = 24), 91.7% (N = 22) of the patients had moderate/severe MAFLD with an LFC > 14%.

As approximately 50% of patients with MAFLD are nonobese in China, it is important to filter patients with MAFLD who are not obese. To this end, we selected subjects with a BMI < 28 kg/m^2 and constructed a classification tree (Fig. 5A and B; the importance of variables is shown in Supplementary Fig. 4A and B). After testing, ALT and insulin could be used to distinguish healthy subjects and patients with MAFLD (accuracy: 84.2%). When ALT and insulin levels were both under their thresholds (47.5 U/L and 40.88 pmol/L, respectively, N = 12), 91.7% of the subjects (N = 11) were healthy. To discriminate the MAFLD status in nonobese patients with MAFLD, CAP was combined with CK18 and could distinguish mild and moderate/severe MAFLD in nonobese patients (accuracy: 80.4%); when the CAP value and CK18 level were both higher than their thresholds (N = 9), all patients had moderate/severe MAFLD.

3.5 Regression tree to predict LFC

The LFC values were used as response variables to construct a regression tree (Fig. 6A and B). After identifying the best values of the parameters of the tree model by 10-fold cross-validation, a number of variables were selected, and their importance values are plotted in Supplementary Fig. 5A and B. Finally, CAP, insulin and AST were found to be significant influential variables for predicting the LFC (mean squared error of model: 21.51). The median LFC in the group in which the CAP value and insulin level were low (N = 85) was 6.8%. The patients with both high CAP (> 301.5 dB/m) values and high AST (> 48 U/L) levels had more severe liver steatosis (N = 21, median LFC: 22.2%). In the nonobese group, CAP, ALT, insulin and CK18 were significant influential variables in predicting the LFC (mean squared error of the model: 39.21); the median LFC in the group with low CAP (< 300.5 dB/m) values and ALT (< 76.65 U/L) and insulin (< 68.1 pmol/ml) levels was 4.7%.

3.6 Characteristics of the models

To investigate the accuracy of the FLI, HSI, APRI and FAST and new models on the diagnosis of MAFLD, the receiver operating characteristic (ROC) curve was plotted based on the algorithm score and MRI-PDFF diagnosis results (Fig. 2A and B and Table 4). Among the published algorithms, FAST (AUC = 0.84), APRI (AUC = 0.74) and FLI (AUC = 0.71) have a better ability to diagnose MAFLD than the HSI (AUC = 0.69). After including general clinical and imaging information and serum biomarkers, three new models were constructed (logistic, lasso and Elastic Net), which showed better performances than those of previously published models (the AUCs of the three models were 0.91, 0.90, and 0.89, respectively).

Table 4
Best cut-off and their sensitivity, specificity and AUC of 7 models.

Model	Cut-off	Sensitivity	Specificity	AUC	AUC 95%cCI
HSI	45.05	0.19	1	0.691	[0.55–0.84]
FLI	38.03	0.431	1	0.711	[0.59–0.84]
APRI	0.27	0.466	0.933	0.728	[0.61–0.88]
FAST	0.289	0.534	1.0	0.837	[0.74–0.94]
Elastic net	0.808	0.759	0.933	0.891	[0.81–0.98]
Lasso	0.712	0.914	0.800	0.903	[0.81–0.99]
Logistic	0.777	0.879	0.800	0.906	[0.83–0.98]

Caption: Performance of four reported models (HSI, FLI, APRI and FAST) and 3 newly-developed models (Elastic Net, Lasso, Logistic)

3.7 SNP results

Samples from 230 subjects had genotyping data. The allele frequencies in the healthy and MAFLD groups are shown in Table 5. Heterozygous or homozygous subjects for the rs738409 G allele showed higher ALT (48 U/L vs. 55 U/L), AST (32 U/L vs. 35 U/L), UA (405 U/L vs. 438 U/L), TB (15.0 μmol/L vs. 16.7 μmol/L), LDL (3.02 mmol/L vs. 3.31 mmol/L), IP-10 (109 pg/ml vs. 125 pg/ml), CK18 (182 U/L vs. 245 U/L), FGF21 (314 pg/ml vs. 367 pg/ml), LSM (5.4 vs. 6.0) and LFC (8.7% vs. 12.3%, Fig. 3A), and the P values were all under 0.05 (Table 6). Logistic regression analysis (Table 7) showed that the PNPLA3 G allele was a significant genetic risk factor (OR = 1.81, $P < 0.001$) between the healthy (LFC < 5%) and MAFLD (LFC ≥ 5%) groups. In addition, among the MAFLD patient groups, it remained significant (OR = 1.84, $P < 0.001$) between the mild MAFLD (with LFC 5% − 13.9%) and moderate/severe MAFLD groups (with LFC > 14%).

Table 5
Allele frequency in healthy and MAFLD groups.

	Healthy(41)	MAFLD(189)
PNPLA3 rs738409		
CC	16(39)	60(31.7)
CG	18(43.9)	66(34.9)
GG	7(17.1)	63(33.3)
HSD17B13 rs72613567		
AA	2(4.9)	14(7.4)
DA	18(43.9)	70(37)
DD	21(51.2)	105(55.6)
RASGRP1 rs7403531		
TT	14(34.1)	83(43.9)
CT	25(61)	86(45.5)
CC	2(4.9)	20(10.6)
ANH6NW6 rs9991328		
CC	4(9.8)	18(9.5)
CT	19(46.3)	74(39.2)
TT	18(43.9)	97(51.3)
UQCC1 rs878639		
AA	18(43.9)	97(51.3)
AG	19(46.3)	73(38.6)
GG	4(9.8)	19(10.1)
MBOAT7 rs641738		
CC	29(70.7)	118(62.4)
CT	10(24.4)	66(34.9)
TT	2(4.9)	5(2.6)
ADIPOQ rs1501299		

Altered allele of 9 SNPs are located at the bottom of every three rows (D denotes deletion)

	Healthy(41)	MAFLD(189)
GG	19(46.3)	108(57.1)
GT	18(43.9)	69(36.5)
TT	4(9.8)	12(6.3)
NOS3 rs2070744		
CC	2(4.9)	2(1.1)
CT	6(14.6)	27(14.3)
TT	33(80.5)	160(84.7)
APLNR rs948847		
GG	2(4.9)	15(7.9)
GT	15(36.6)	68(36)
TT	24(58.5)	106(56.1)
Altered allele of 9 SNPs are located at the bottom of every three rows (D denotes deletion)		

Table 6
Comparison of features of clinical, laboratory and imaging in variant and common alleles of PNPLA3 rs738409.

Var	C (n = 302)	G (n = 158)	P value
Clinical Data			
ALT	48 (37)	55.5 (39)	0.0074
AST	32 (16.3)	35.55 (20.4)	0.0005
TB	15 (7.6)	16.65 (5.4)	0.001
LDL	3.02 (1.25)	3.31 (1.37)	0.0082
Serum Biomarkers			
IP10	109.95 (55.52)	125.15 (62.22)	< 0.0001
Leptin	75.75 (111.02)	86.22 (128.13)	0.146
CK18	182.7 (200.95)	245.7 (263.92)	0.0004
FGF21	314.1 (215.78)	367.3 (353.5)	0.0045
FGF19	126.3 (97.4)	126.2 (86.21)	0.6867
Fibroscan, MRI-PDFF			
CAP	287.5 (65)	292 (60)	0.6689
LSM	5.4 (2)	5.95 (2.5)	0.0041
LFCmean	8.65 (8.4)	12.3 (7.55)	< 0.0001

Table 7
Logistic Regression analysis of genetic risk factors.

MAFLD vs. Healthy				
Gene	OR	2.50%	97.50%	P
PNPLA3	1.81	1.36	4.53	0.02
HSD17B13	0.93	0.51	1.73	0.82
UQCC1	0.78	0.44	1.39	0.39
MBOAT7	1.2	0.62	2.43	0.6
Moderate/severe MAFLD vs. Healthy				
Gene	OR	2.50%	97.50%	P
PNPLA3	1.84	1.37	3.33	< 0.001
HSD17B13	2.54	1.68	4.63	< 0.001
UQCC1	0.58	0.34	0.95	0.04
MBOAT7	1.37	0.78	2.36	0.27
Nonobese population				
MAFLD vs. Healthy				
PNPLA3	1.469	0.748	2.993	0.274
HSD17B13	0.889	0.414	1.962	0.766
UQCC1	0.814	0.378	1.79	0.601
MBOAT7	4.467	1.492	20.034	0.018
Moderate MAFLD vs. Mild MAFLD				
PNPLA3	1.29	1.13	3.999	0.444
HSD17B13	2.26	1.092	4.483	0.029
UQCC1	0.813	0.377	1.7	0.588
MBOAT7	1.903	0.836	4.255	0.119
Caption: Adjusted ORs calculated by multiple logistic regression (Age, Sex, BMI and SNPs).				

Moreover, HSD17B13 (rs72613567 A allele, OR=2.54, P<0.001) and UQCC1 (rs878639 G allele, OR=0.58, P=0.04) were also significant factors for moderate/severe MAFLD, and higher LFCs were found in persons with the variant allele of HSD17B13 (11.0% vs. 14.5%, P=0.004, Fig. 3B). Patients with the variant

allele of UQCC1 showed lower LFC (10.4% vs. 8.4%, P=0.015, Fig. 3C) and UA (421 μmol/L vs. 391 μmol/L, P=0.018); beyond this, FGF21 and insulin both showed higher trends with the wild-type allele (P>0.05), while inflammation and apoptosis markers were at almost the same level between the two polymorphisms.

To investigate the genetic risk factors among the nonobese group, we performed regression analysis among patients with BMI ≤28 kg/m2. MBOAT7 (rs641738 T allele) was a significant risk factor (OR=4.47, P=0.018), and patients with the T allele showed a higher LFC than patients with the wild-type allele (C allele 8.3% vs. T allele 10.4%, P=0.0544, Fig. 3D). Among MAFLD patients (LFC >5.0%), HSD17B13 rs72613567 was a significant risk factor for moderate/severe MAFLD (OR=2.26, P=0.029), and patients with the risk allele showed a significantly higher LFC (no insertion 9.7% vs. A insertion 12.3%, P=0.028).

To further investigate genetic risk factors in MAFLD, subjects were grouped by GRS (0: N=159; 1: N=190; 2: N=89; and 3: N=22). Patients with higher GRSs showed higher LFC (8.1% vs. 10.7% vs. 11.6% vs. 14.5%, P<0.05, Table 8), AST (31.0 U/L vs. 34.1 U/L vs. 34.4 U/L vs. 31.9 U/L, P<0.05), TB (14.0 μmol/L vs. 16.0 μmol/L vs. 16.2 μmol/L vs. 17 μmol/L, P<0.05), UA (395 μmol/L vs. 424 μmol/L vs. 418 μmol/L vs. 424 μmol/L, P<0.05), and CK18 (174 U/L vs. 221 U/L vs. 228 U/L vs. 180 U/L). Regression analysis showed that the GRS was a significant risk factor for MAFLD (OR: 1.59, P=0.01) after adjusting for the other SNPs. In addition, among nonobese subjects, the GRS remained a significant risk factor for nonobese MAFLD (OR: 1.87, P<0.009).

Table 8
Patients grouped by GRS scores.

Variable	0 (n = 159)	1 (n = 190)	2 (n = 89)	3 (n = 22)	p value
Basic Information					
Age	41 (13)	39 (14)	40 (19)	38.5 (17.5)	0.4638
WC	96.6 (14.55)	96 (11.3)	96.1 (13.7)	94.35 (10.5)	0.6055
HC	101.1 (10.05)	100.7 (9.5)	102 (9.3)	102.05 (7.35)	0.6147
AC	97.4 (10.85)	97.4 (11.22)	97.6 (9.7)	98.9 (12.55)	0.5277
BMI	27.9 (4.95)	27.55 (4.2)	27.8 (4)	27.75 (4.12)	0.5014
Clinical Data					
ALT	47 (36)	53.5 (40)	50 (41)	54.5 (36.67)	0.1907
AST	31 (16)	34.1 (19.65)	34.4 (24.9)	31.9 (22.82)	0.0453
UA	395 (146.5)	424.5 (130.25)	418 (116)	423.5 (110)	0.0248
TG	2.17 (1.44)	2.2 (1.45)	2.19 (1.57)	2.09 (2.26)	0.7312
Insulin	97.61 (71.19)	105.5 (73.64)	108.1 (84.22)	98.25 (73.08)	0.6809
Ferritin	187 (278.15)	219.8 (197.2)	231.2 (272)	259.8 (268.15)	0.4253
TB	14 (6)	16 (6.65)	16.2 (9)	17 (8.75)	0.0378
TBA	2.5 (1.85)	2.3 (2.4)	2.2 (1.6)	2.15 (1.05)	0.5322
TC	5.1 (1.22)	5.2 (1.4)	5.3 (1.2)	4.85 (0.74)	0.6203
LDL	3.02 (1.22)	3.18 (1.42)	3.29 (1.22)	2.98 (0.87)	0.4987
HDL	1.03 (0.28)	1.1 (0.28)	1.1 (0.2)	1.05 (0.24)	0.6557
Serum Biomarkers					
Adiponectin	28.02 (24.35)	29.06 (26.26)	26.42 (19.18)	29.82 (21.17)	0.4701
IP-10	111.4 (60.6)	116 (55.29)	115.8 (52.52)	119.6 (67.89)	0.4819
Leptin	77.35 (115.57)	82.42 (94.19)	86.5 (130.12)	64.08 (43.78)	0.3843

GRS scores: 0 -no risk allele, 1- one risk allele, 2- two risk alleles, 3- three risk alleles; risk allele: PNPLA3 rs738409 G allele, MBOAT7 rs641738 T allele, HSD17B13 deletion allele.

Variable	0 (n = 159)	1 (n = 190)	2 (n = 89)	3 (n = 22)	p value
CK18	174.9 (192.1)	221.7 (231.4)	228.6 (256.3)	179.9 (386.37)	0.0415
FGF21	312.2 (213.05)	343.4 (243.58)	319 (270)	295.3 (332.33)	0.5468
FGF19	131.2 (97.88)	125.1 (94.31)	126.6 (80.42)	134.45 (71.55)	0.4189
FibroScan, MRI-PDFF					
CAP	287 (58)	291 (61.75)	291 (74)	285 (98.5)	0.839
LSM	5.4 (2)	5.7 (2.08)	5.4 (2.1)	6.05 (2.12)	0.355
LFCmean	8.1 (6.6)	10.7 (8.8)	11.6 (9.7)	14.5 (16.77)	0.0012
GRS scores: 0 -no risk allele, 1- one risk allele, 2- two risk alleles, 3- three risk alleles; risk allele: PNPLA3 rs738409 G allele, MBOAT7 rs641738 T allele, HSD17B13 deletion allele.					

4 Discussion

To the best of our knowledge, this is the first study based on MRI-PDFF and liver biopsy to investigate the clinical, laboratory and genetic features of the MAFLD population. Due to the inadequate diagnosis of NASH, insufficient knowledge of MAFLD and greater genetic predisposition to MAFLD, these conditions need to be further assessed.[31] It has been reported that steatosis is associated with significantly increased overall mortality, and this risk progressively increases with worsening MAFLD histology.[8] MRI-PDFF can accurately assess the steatosis degree of the liver, which was also demonstrated in our study; thus, it is necessary to evaluate the status of NAFLD patients by MRI-PDFF. However, this technique has not been commonly applied in clinical services in China. Here, this cross-sectional study investigated the clinical characteristics and noninvasive serum biomarkers, including potential risk SNPs, of MAFLD patients based on MRI-PDFF and liver biopsy.

In this study, we found that the incidences of MS and DM were higher in the MAFLD group and specifically were highest in the moderate/severe MAFLD group; in addition, most of the metabolic-related factors (insulin, HOMA, TG, ferritin, and FGF21) were significantly higher in MAFLD. Even between the mild MAFLD and the healthy groups, whose distinction was smaller (healthy was defined as LFC ~ 5.0%, and mild MAFLD was defined as LFC 5.1%-14.1%), correlation analysis showed that insulin, TG and FGF21 had strong positive correlations with LFC, suggesting that the glucose and lipid metabolic pathways might be the focus of disease aggravation and emphasizing the need for renaming NAFLD as MAFLD. Moreover, CK18, which indicates the degree of hepatocyte apoptosis,[32] has the strongest correlation with LFC among the serum biomarkers ($R = 0.6$, $P < 0.05$), suggesting the need to evaluate CK18 in long-term liver status monitoring.

However, we did not find a significant difference in FGF19 among the three groups; even between the moderate/severe group and the healthy group, FGF19 remained similar between the NASH and non-NASH groups. As a marker related to bile acid metabolism,[33] FGF19 was found to be obviously lower in the NASH group in another study,[34] whereas in our study, FGF19 only showed a significant difference between non-obese and obese MAFLD. Adiponectin and leptin also exhibited similar results. Thus, these indicators similar between different degree of liver steatosis may not be an efficacy parameter to reflect treatment effects in NASH clinical trials in Chinese population. Inflammatory markers have been demonstrated to play a pivotal role in the pathogenesis of experimental steatohepatitis[35] and can also reflect the disease status. IP-10 and IL-6 levels showed differences between different MAFLD statuses, and a great difference was also found between the mild MAFLD and healthy groups, indicating that chronic low-grade inflammation exists in patients with mild MAFLD.

To verify the diagnostic value of the MAFLD molecular prediction model (FLI, HSI, APRI and FAST), MRI-PDFF serves as the main diagnostic standard in testing prediction ability. The predictive ability of FAST, which is calculated based on AST, CAP, and LSM, has not been identified in China based on MRI-PDFF since it has only been recently reported.[36] After testing in test subsets, FAST performed better in predicting steatosis among the four reported models. Machine learning-based approaches were used to address the issue of a high-dimensional small dataset.[29, 30, 37] When predictive abilities of the reported and developed models were tested using the same test subset, their performances were similar and better than that of the previous model. These models (LASSO, elastic net, and logistic) were capable of accurately establishing the relationships between our analysed features and MAFLD, but they still need further investigation in other MAFLD cohorts.

Decision tree-based models were used to identify predictors of MAFLD, not only to differentiate healthy (LFC < 5%) and MAFLD (LFC > 5%) but also to differentiate mild MAFLD (LFC ≤ 14%) and moderate/severe MAFLD (LFC > 14%). An increased LFC (MRI-PDFF ≥ 15%) is associated with increased odds of fibrosis progression in patients with NAFLD at an early stage of fibrosis.[10] In addition, regression tree could reveal strong predictors of LFC. Compared with generalized linear models, the decision tree model is easier to understand because the results exported in the decision tree model resemble clinical decision-making processes and the tree structures are more flexible for distributing the response variable without preassumptions.[38] Our tree models provided strong evidence that ALT, AST, TC, CAP, CK18, and insulin were important predictors of MAFLD.

It has been demonstrated that approximately 44.3% of MAFLD patients are nonobese in China, which is similar to our result (49%).[39] Wong et al.[40] found similar clinical characteristics between nonobese and obese MAFLD; likewise, in our study, LFCs were comparable between obese and nonobese patients. In addition, more severe tissue inflammation was found in obese Chinese patients, as indicated by higher IP-10 and IL-6 levels. A study using liver biopsies or FibroScan also showed that obese NAFLD patients are more likely to develop advanced liver complications than are nonobese patients in China (AF 31.6% versus 6.3% and HCC 0.9% versus 0%).[7] Further follow-up studies could concentrate on comparing the different outcomes of nonobese and obese patients.

Genetic factors may be important in the development of MAFLD in either obese or nonobese patients.[41] The SNP rs738409 of PNPLA3 has been most widely reported as a hereditary risk factor associated with MAFLD. Other SNPs also contribute to the development of NAFLD (such as HSD17B13 and MBOAT7).[42] However, as the gold standard for MAFLD, histology is limited as a quantitative analysis of liver steatosis. MRI-PDFF, which can quantify liver steatosis, uncovers minor changes in liver steatosis.

In this study, based on MRI-PDFF, a significantly higher LFC was found in patients with the PNPLA3 G allele; furthermore, subjects with the PNPLA3 G allele showed higher concentrations of biomarkers related to liver injury and tissue inflammation. Among nonobese subjects, people with the MBOAT7 T allele had a significantly higher LFC, and thus, the MBOAT7 T allele was identified as a risk factor in nonobese subjects. The HSD17B13 rs72613567 variant allele was associated with moderate/severe MAFLD in both the nonobese and whole cohorts.

UQCC1 plays a role in mitochondrial respiratory chain complex III protein expression and is structurally similar to the mouse Bfzb controlling mouse brown fat.[43, 44] Previous associations at the UQCC1 rs4911494 locus included arm fat reduction,[45] reduced height, reduced body weight, and increased WHRadjBMI.[46] In our study, UQCC1 rs878639 (A > G) was identified as having a protective role against MAFLD, and people with the variant allele showed lower LFC, CAP and UA, which has not been found in other studies, while apoptosis and inflammation markers were not found to be different. Higher GRSs showed more severe disease status, suggesting the role of genetic factors in MAFLD.

4.1 Strengths and limitations

The data were collected using standardized protocols, which add validity to our study's results. To the best of our knowledge, this is the first study to investigate the clinical characteristics and serum biomarkers (including gene polymorphisms) of MAFLD patients in a Chinese population based on MRI-PDFF. As a foundation, this study first reported a MAFLD diagnostic model that uses a machine learning method based on MRI-PDFF. The small sample size is a main limitation in the study. The cross-sectional design of the study may represent a limitation. In cross-sectional analyses, reverse causation cannot be excluded. Further investigation needs to be conducted in follow-up studies.

5 Conclusions

Approximately 1/3 of patients had moderate/severe MAFLD and liver fibrosis in our study, suggesting the severity of fatty liver in China. Obese patients with MAFLD and the nonobese group showed a similar degree of hepatic steatosis. Models and decision trees could preliminarily evaluate diseases and differentiate moderate/severe MAFLD. The UQCC1 rs878639 G allele was a protective genetic factor, and the MBOAT7 T allele may be a risk factor among the nonobese population leading to MAFLD. The difference in the LFC between variants and common alleles could be quantified by MRI-PDFF, and more risk alleles showed more severe disease status, suggesting a very important role of genetic factors in MAFLD.

Abbreviations

AC	Abdominal circumference
ALT	alanine transaminase
AST	aspartate aminotransferase
AUC	area under curve
BMI	body mass index
CAP	controlled attenuation parameter
CK18	cytokeratin 18
DBP	diastolic blood pressure
FAST	Fibroscan-AST
FGF21	Fibroblast growth factor 21
FLI	fatty liver index
GGT	Gamma-glutamyltransferase
HC	hip circumference
HCC	hepatocellular carcinoma
HDL-C	high-density lipoprotein cholesterol
HSI	hepatocyte steatosis index
IGF1	insulin-like growth factor-1
IL6	interleukin-6
IP10	interferon-inducible protein 10
IR	insulin resistance
LDL-C	low-density lipoprotein cholesterol
LFC	Liver fat content
LSM	Liver stiffness measurement
MAFLD	metabolic associated fatty liver disease
MetS	metabolic syndrome
MRI-PDFF	magnetic resonance imaging-derived proton density fat fraction
NAFLD	non-alcoholic fatty liver disease
NASH	nonalcoholic steatohepatitis
OR	odd ratio

SBP	systolic blood pressure
T2DM	Diabetes Mellitus type 2
TB	total bilirubin
TBA	total bile acid
TC	total cholesterol
TG	triglyceride
TIMP1	tissue inhibitor matrix metalloproteinase 1
UA	uric acid
WC	waist circumference

Declarations

DATA AVAILABILITY

The data used to support the findings of this study are available from the corresponding author upon request.

ANIMAL RESEARCH (ETHICS)

NA

CONSENT TO PARTICIPATE (ETHICS)

This study was approved by the Ethics Committee of the First Hospital of Jilin University (Ethical Approval Number: 19K096001), and written informed consent was obtained from all participants.

CONSENT TO PUBLISH (ETHICS)

Every author consented to publish.

Clinical Trials Registration

NA

AUTHORS' CONTRIBUTIONS

YA, DY, LG, JN designed research. LZ, XZ, YZ, HUZ, DZ, QD, HJ, HC, LX, HOZ, MJ, LC, ZW, DW performed imaging or biopsy examination. YA performed experiments and analysed data. YA, DY, wrote the manuscript.

Conflict of Interest INTERESTS

None declared

FUNDING

This study was supported by the National Natural Science Foundation of China

ACKNOWLEDGEMENTS

We would like to show our deepest gratitude to every authors.

References

1. Paik JM, Golabi P, Younossi Y, et al. Changes in the global burden of chronic liver diseases from 2012 to 2017: the growing impact of NAFLD. Hepatology. 2020;72:1605–16.
2. Eslam M, Fan JG, Mendez-Sanchez N. Non-alcoholic fatty liver disease in non-obese individuals: the impact of metabolic health. Lancet Gastroenterol Hepatol. 2020;5:713–5.
3. Eslam M, Sanyal AJ, George J, et al. MAFLD: a consensus-driven proposed nomenclature for metabolic associated fatty liver disease. Gastroenterology 2020;158:1999–2014.e1.
4. Nan Y, An J, Bao J, et al. The Chinese society of hepatology position statement on the redefinition of fatty liver disease. J Hepatol. 2021;75:454–61.
5. Bessone F, Razori MV, Roma MG. Molecular pathways of nonalcoholic fatty liver disease development and progression. Cell Mol Life Sci. 2019;76:99–128.
6. Ekstedt M, Hagstrom H, Nasr P, et al. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. Hepatology. 2015;61:1547–54.
7. Leung JC, Loong TC, Wei JL, et al. Histological severity and clinical outcomes of nonalcoholic fatty liver disease in nonobese patients. Hepatology. 2017;65:54–64.
8. Simon TG, Roelstraete B, Khalili H, et al. Mortality in biopsy-confirmed nonalcoholic fatty liver disease: results from a nationwide cohort. Gut. 2021;70:1375–82.
9. Davison BA, Harrison SA, Cotter G, et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. J Hepatol. 2020;73:1322–32.
10. Tamaki N, Ajmera V, Loomba R. Non-invasive methods for imaging hepatic steatosis and their clinical importance in NAFLD. Nat Rev Endocrinol. 2022;18:55–66.
11. Chen J, Duan S, Ma J, et al. MRI-determined liver fat correlates with risk of metabolic syndrome in patients with nonalcoholic fatty liver disease. Eur J Gastroenterol Hepatol. 2020;32:754–61.
12. Shao C, Ye J, Li F, et al. Different predictors of steatosis and fibrosis severity among lean, overweight and obese patients with nonalcoholic fatty liver disease. Dig Liver Dis. 2019;51:1392–9.
13. Chen H, Zeng WK, Shi GZ, et al. Liver fat accumulation measured by high-speed T2-corrected multi-echo magnetic resonance spectroscopy can predict risk of cholelithiasis. World J Gastroenterol. 2020;26:4996–5007.

14. Lee SW, Lee TY, Yang SS, et al. The association of non-alcoholic fatty liver disease and metabolic syndrome in a Chinese population. Hepatobiliary Pancreat Dis Int. 2017;16:176–80.
15. Zheng X, Gong L, Luo R, et al. Serum uric acid and non-alcoholic fatty liver disease in non-obesity Chinese adults. Lipids Health Dis. 2017;16:202.
16. Sun L, Wang Q, Liu M, et al. Albumin binding function is a novel biomarker for early liver damage and disease progression in non-alcoholic fatty liver disease. Endocrine. 2020;69:294–302.
17. Dai J, Yi J, Zhang S, et al. Serum 3-carboxy-4-methyl-5-propyl-2-furanpropanoic acid is associated with lipid profiles and might protect against non-alcoholic fatty liver disease in Chinese individuals. J Diabetes Investig. 2019;10:793–800.
18. Castera L, Friedrich-Rust M, Loomba R. Noninvasive assessment of liver disease in patients with nonalcoholic fatty liver disease. Gastroenterology. 2019;156:1264–81.e4.
19. Alvarez-Sola G, Uriarte I, Latasa MU, et al. Fibroblast growth factor 15/19 (FGF15/19) protects from diet-induced hepatic steatosis: development of an FGF19-based chimeric molecule to promote fatty liver regeneration. Gut. 2017;66:1818–28.
20. Bedogni G, Bellentani S, Miglioli L, et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. BMC Gastroenterol. 2006;6:33.
21. Lee JH, Kim D, Kim HJ, et al. Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease. Dig Liver Dis. 2010;42:503–8.
22. Eslam M, George J. Genetic contributions to NAFLD: leveraging shared genetics to uncover systems biology. Nat Rev Gastroenterol Hepatol. 2020;17:40–52.
23. Ji F, Liu Y, Hao JG, et al. KLB gene polymorphism is associated with obesity and non-alcoholic fatty liver disease in the Han Chinese. Aging. 2019;11:7847–58.
24. Liu Q, Liu SS, Zhao ZZ, et al. TRIB1 rs17321515 gene polymorphism increases the risk of coronary heart disease in general population and non-alcoholic fatty liver disease patients in Chinese Han population. Lipids Health Dis. 2019;18:165.
25. Wang X, Liu Z, Peng Z, et al. The TM6SF2 rs58542926 T allele is significantly associated with non-alcoholic fatty liver disease in Chinese. J Hepatol. 2015;62:1438–9.
26. Meisner J, Albrechtsen A. Testing for Hardy-Weinberg equilibrium in structured populations using genotype or low-depth next generation sequencing data. Mol Ecol Resour. 2019;19:1144–52.
27. Kouvari M, Chrysohoou C, Skoumas J, et al. The presence of NAFLD influences the transition of metabolically healthy to metabolically unhealthy obesity and the ten-year cardiovascular disease risk: a population-based cohort study. Metabolism. 2021;128:154893.
28. Zhang J, Xu Q, Lai F, et al. Joint associations of metabolically healthy abdominal obesity and non-alcoholic fatty liver disease with prediabetes and diabetes in Chinese adults. BMJ Open Diabetes Res Care. 2021;9:e002362.
29. Garcia-Carretero R, Vigil-Medina L, Barquero-Perez O, et al. Logistic LASSO and elastic net to characterize vitamin D deficiency in a hypertensive obese population. Metab Syndr Relat Disord.

2020;18:79–85.

30. McEligot AJ, Poynor V, Sharma R, et al. Logistic LASSO regression for dietary intakes and breast cancer. Nutrients. 2020;12:2652.
31. Zhou F, Zhou J, Wang W, et al. Unexpected rapid increase in the burden of NAFLD in China from 2008 to 2018: a systematic review and meta-analysis. Hepatology. 2019;70:1119–33.
32. Wieckowska A, Zein NN, Yerian LM, et al. In vivo assessment of liver cell apoptosis as a novel biomarker of disease severity in nonalcoholic fatty liver disease. Hepatology. 2006;44:27–33.
33. Kliewer SA, Mangelsdorf DJ. Bile acids as hormones: the FXR-FGF15/19 pathway. Dig Dis. 2015;33:327–31.
34. Wojcik M, Janus D, Dolezal-Oltarzewska K, et al. A decrease in fasting FGF19 levels is associated with the development of non-alcoholic fatty liver disease in obese adolescents. J Pediatr Endocrinol Metab. 2012;25:1089–93.
35. Schuster S, Cabrera D, Arrese M, et al. Triggering and resolution of inflammation in NASH. Nat Rev Gastroenterol Hepatol. 2018;15:349–64.
36. Newsome PN, Sasso M, Deeks JJ, et al. FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study. Lancet Gastroenterol Hepatol. 2020;5:362–73.
37. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. BMJ. 2015;351:h3868.
38. Mi WF, Chen XM, Fan TT, et al. Identifying Modifiable Risk Factors for Relapse in Patients With Schizophrenia in China. Front Psychiatry. 2020;11:574763.
39. Ye Q, Zou B, Yeo YH, et al. Global prevalence, incidence, and outcomes of non-obese or lean non-alcoholic fatty liver disease: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol. 2020;5:739–52.
40. Wong VW, Wong GL, Chan RS, et al. Beneficial effects of lifestyle intervention in non-obese patients with non-alcoholic fatty liver disease. J Hepatol. 2018;69:1349–56.
41. Eslam M, Valenti L, Romeo S. Genetics and epigenetics of NAFLD and NASH: clinical impact. J Hepatol. 2018;68:268–79.
42. Trepo E, Valenti L. Update on NAFLD genetics: from new variants to the clinic. J Hepatol. 2020;72:1196–209.
43. Tucker EJ, Wanschers BF, Szklarczyk R, et al. Mutations in the UQCC1-interacting protein, UQCC2, cause human complex III deficiency associated with perturbed cytochrome b protein expression. PLoS Genet. 2013;9:e1004034.
44. Vetter K, Wurst W. Expression of a novel mouse gene 'mbFZb' in distinct regions of the developing nervous system and the adult brain. Mech Dev. 2001;100:123–5.
45. Neville MJ, Wittemans LBL, Pinnick KE, et al. Regional fat depot masses are influenced by protein-coding gene variants. PLoS ONE. 2019;14:e0217644.

46. Sanna S, Jackson AU, Nagaraja R, et al. Common variants in the GDF5-UQCC region are associated with variation in human height. Nat Genet. 2008;40:198–203.

Figures

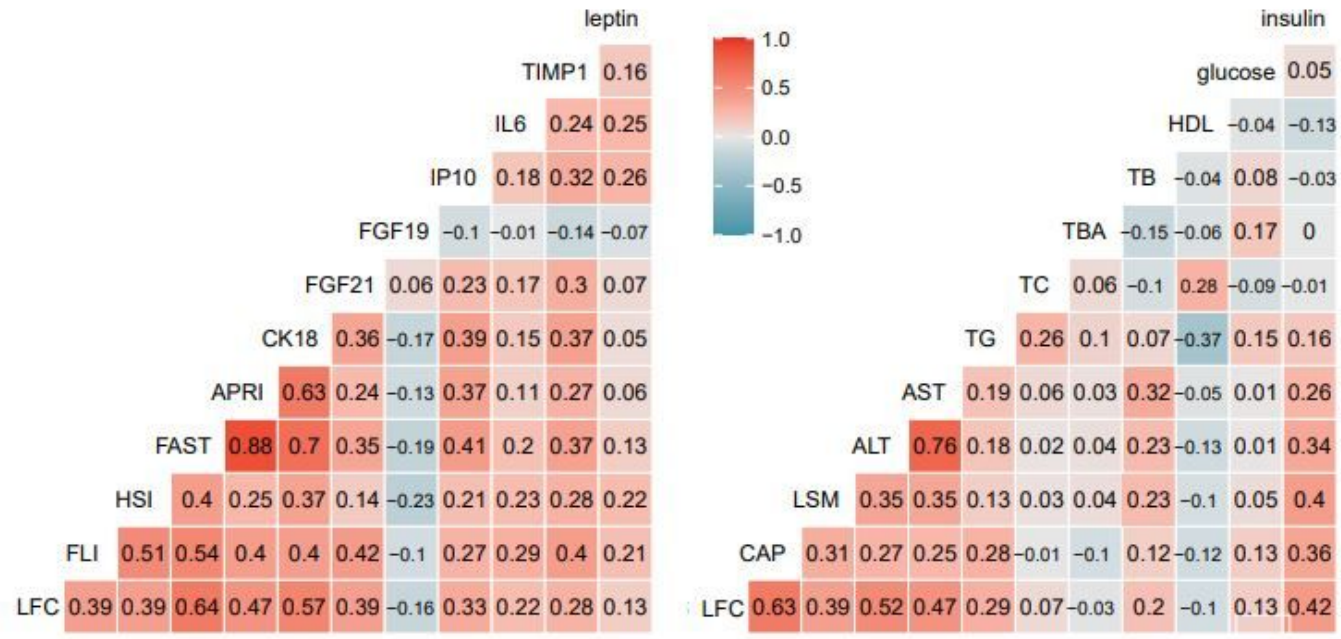

Figure 1

(A) Correlation analysis between the biomarkers, models and LFC.

(B) Correlation analysis between the FibroScan, biochemical and LFC results.

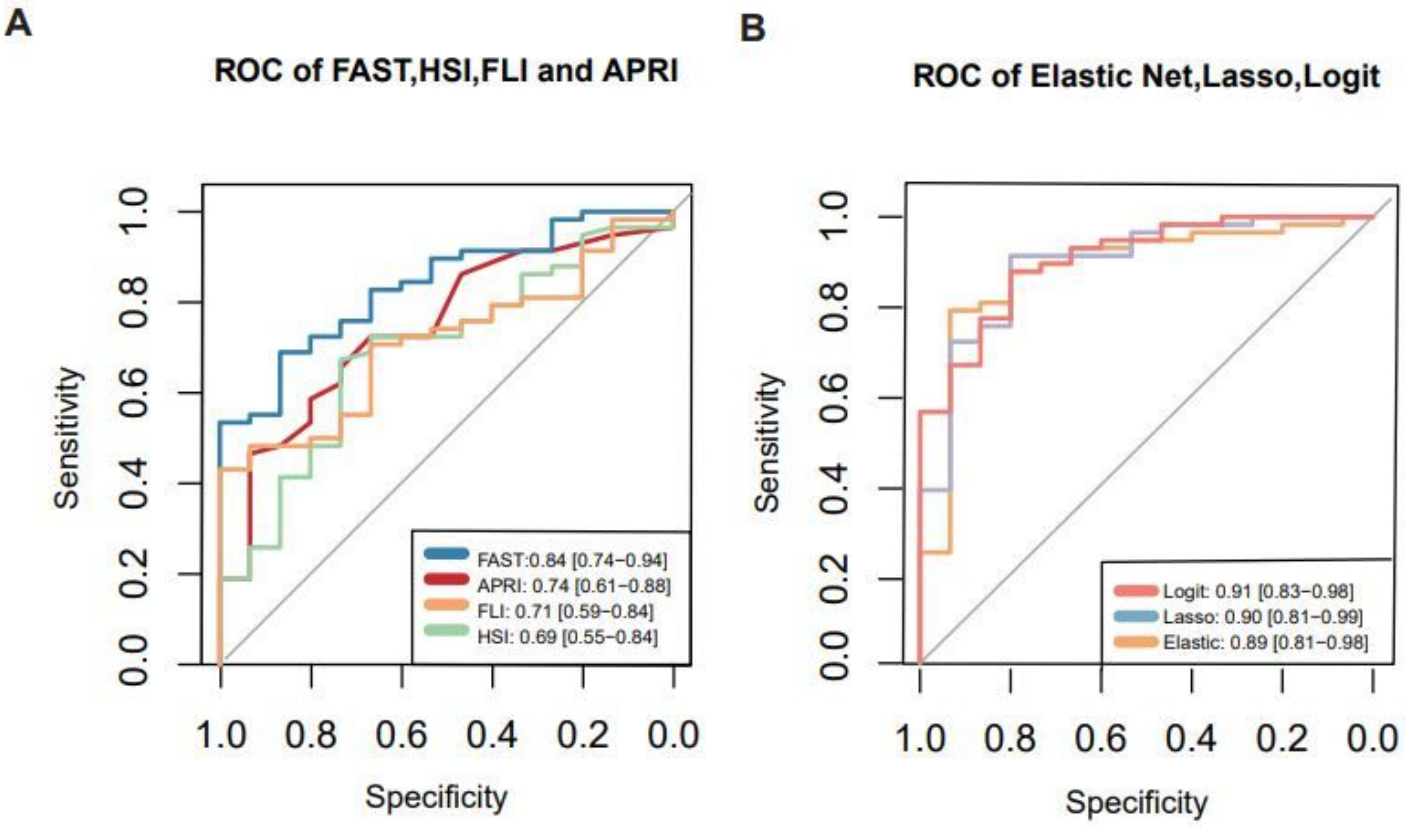

Figure 2

(A) Comparison of the AUCs of four reported models.

(B) Comparison of the AUCs of three newly constructed models.

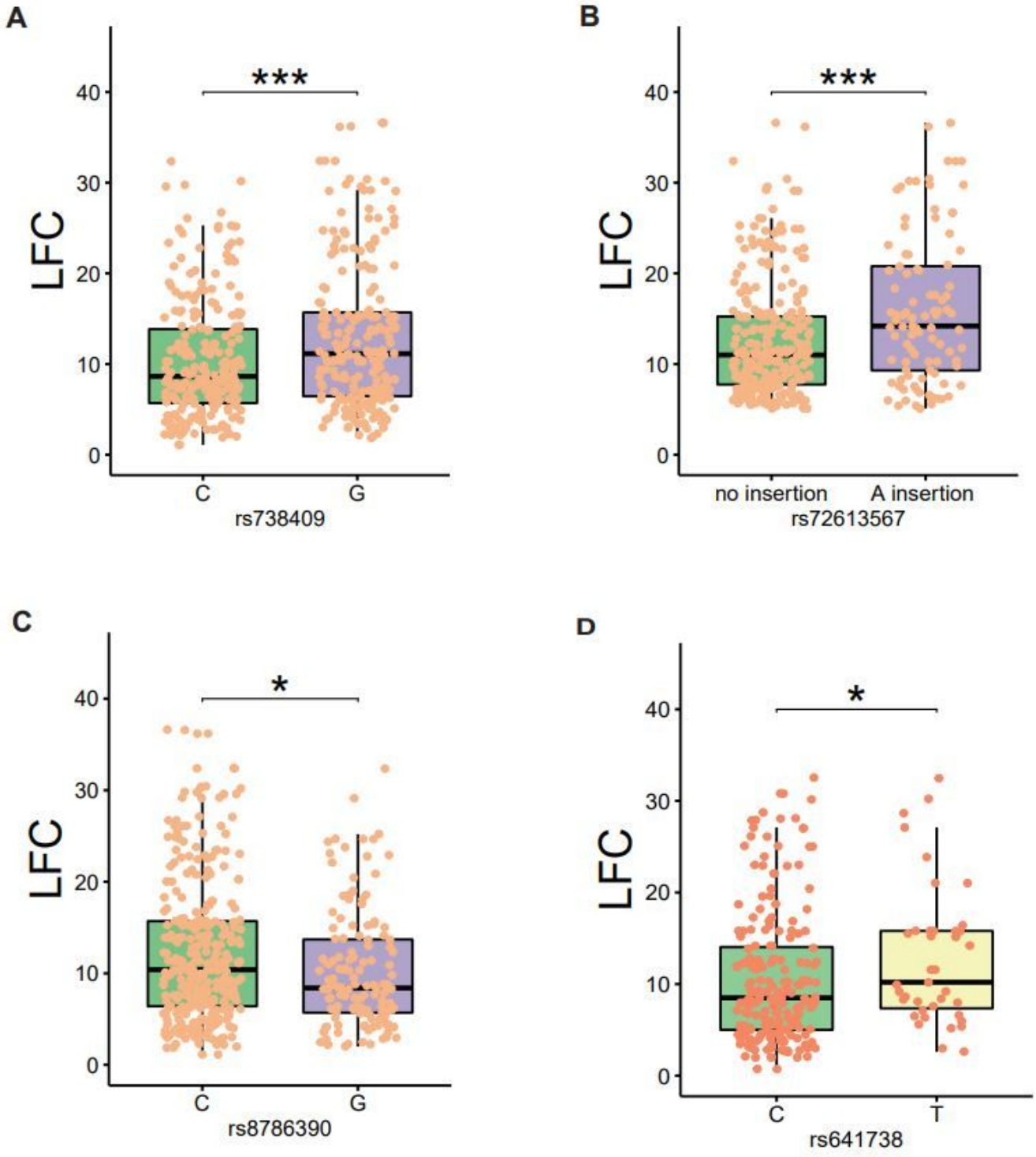

Figure 3

(A) LFC of patients with variant and common alleles of PNPLA3 rs738409.

(B) LFC of patients with variant and common alleles of HSD17B13.

(C) LFC of patients with variant and common alleles of UQCC1 rs878639.

(D) LFC of patients with variant and common alleles of MBOAT7 rs641738.

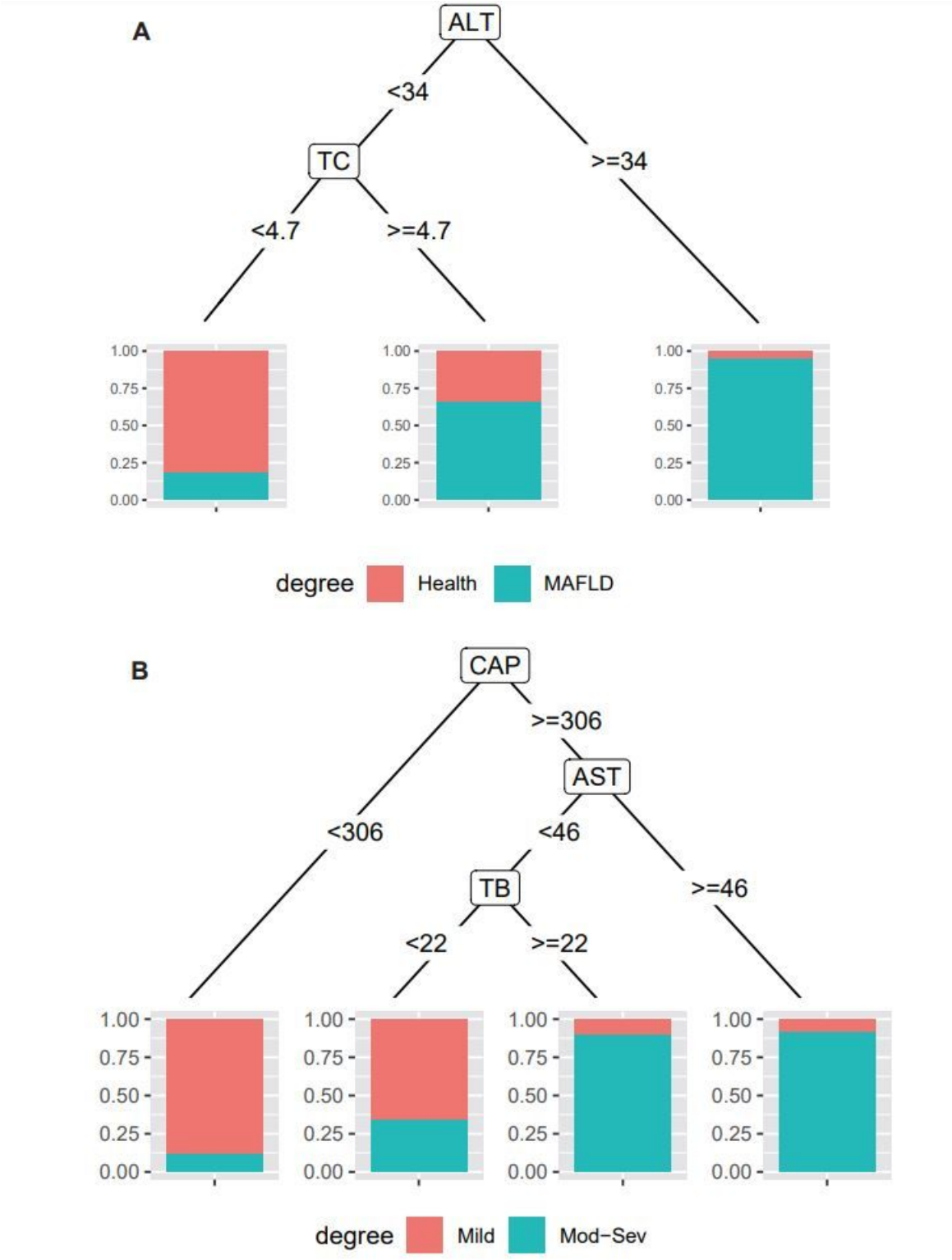

Figure 4

(A) Classification tree to differentiate healthy subjects and those with MAFLD.

(B) Classification tree to differentiate mild MAFLD and moderate/severe MAFLD.

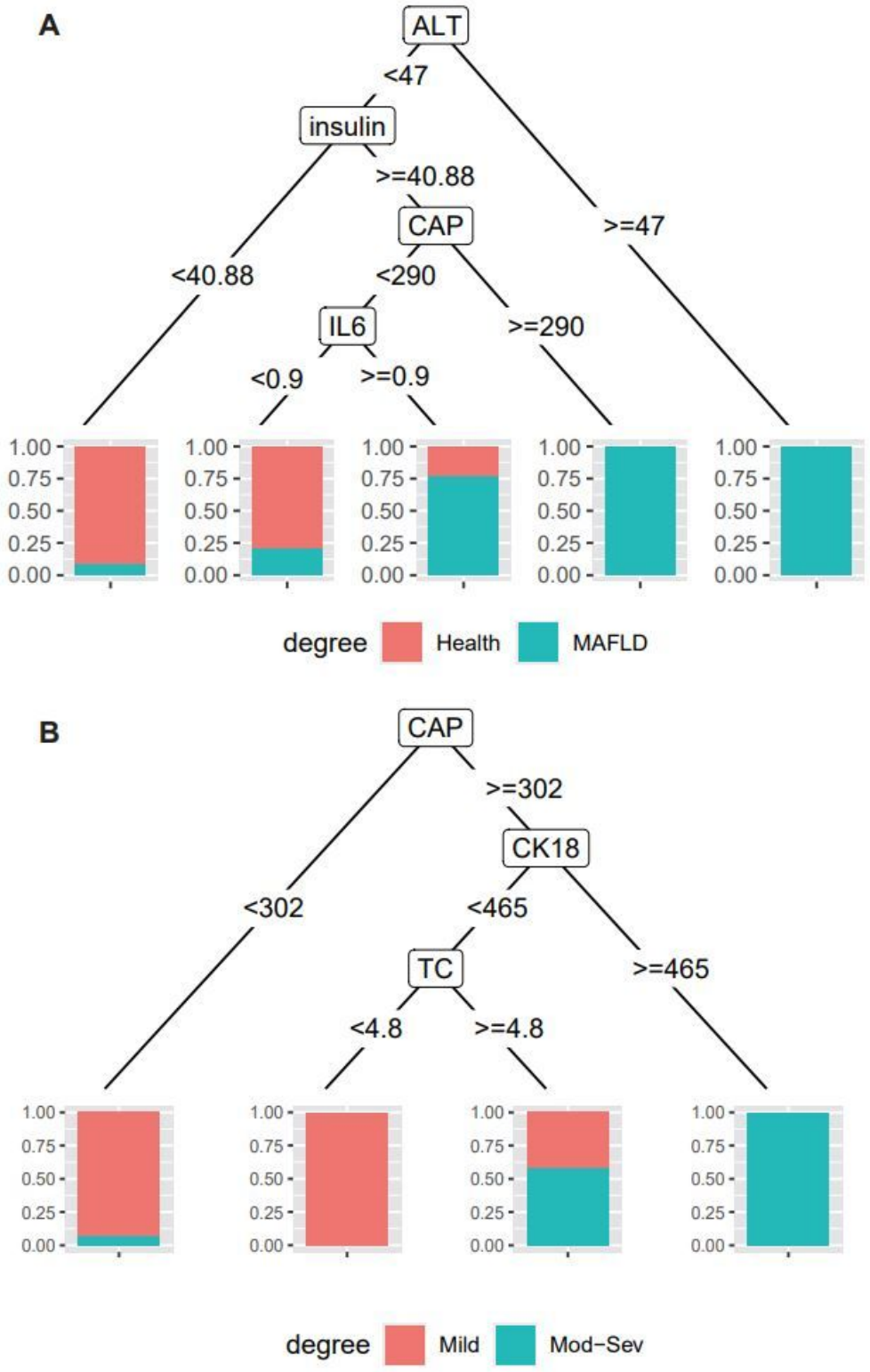

Figure 5

(A) Classification tree to differentiate healthy subjects and those with MAFLD among the nonobese population.

(B) Classification tree to differentiate mild MAFLD and moderate/severe MAFLD among the nonobese population.

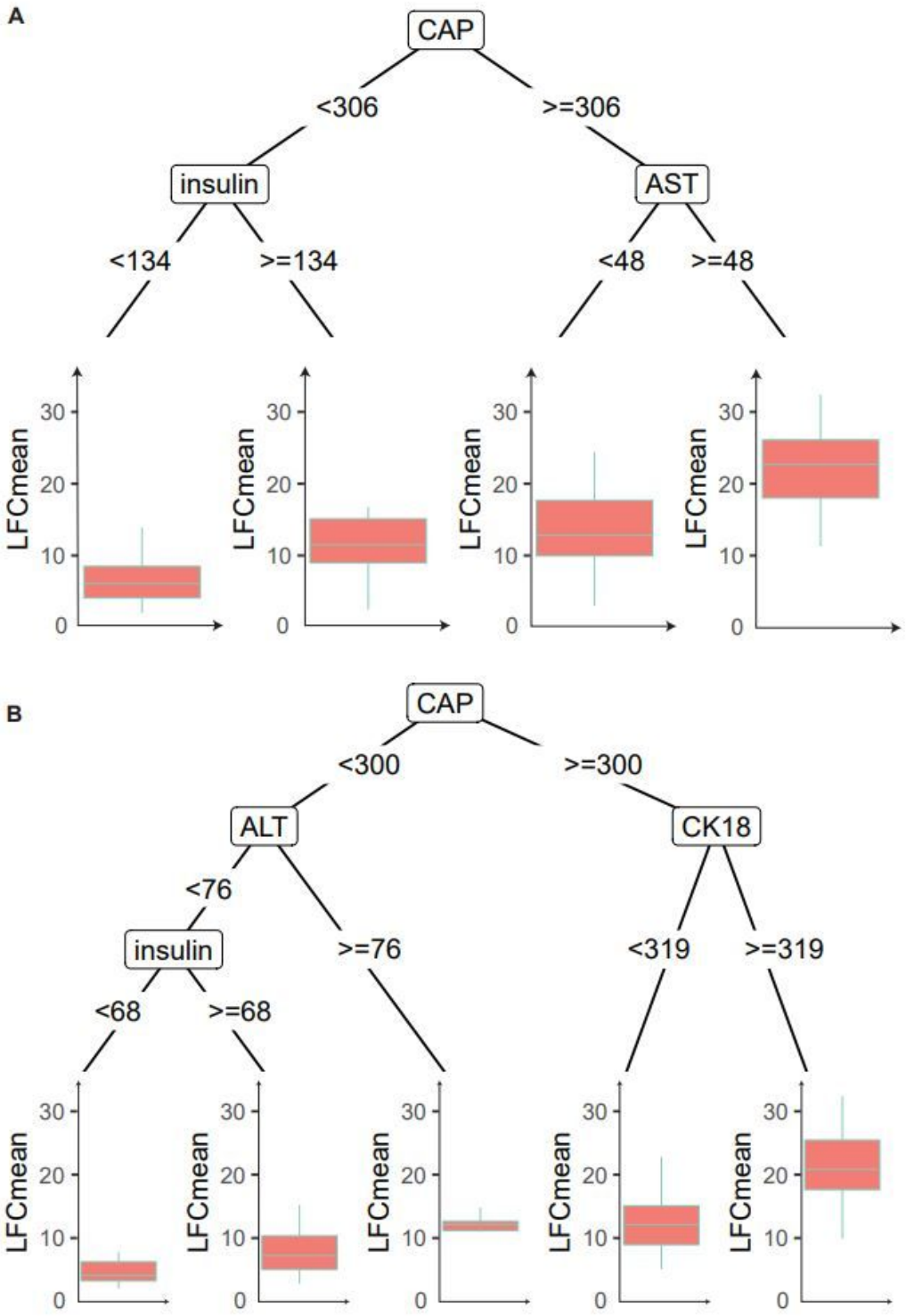

Figure 6

(A) Regression tree to predict LFC.

(B) Regression tree to predict LFC among the nonobese population.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryFigure1.pdf
- supplementaryFigure2.pdf
- supplementaryFigure3AB.pdf
- supplementaryFigure4AB.pdf
- supplementaryFigure5AB.pdf