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.

The interaction of SAMM50-rs738491, PARVB-rs5764455 and PNPLA3-rs738409 increases the susceptibility to nonalcoholic steatohepatitis

Ke Xu
The First Affiliated Hospital of Wenzhou Medical University
Kenneth I. Zheng
The First Affiliated Hospital of Wenzhou Medical University
Pei-Wu Zhu
The First Affiliated Hospital of Wenzhou Medical University
Wen-Yue Liu
The First Affiliated Hospital of Wenzhou Medical University
Hong-Lei Ma
The First Affiliated Hospital of Wenzhou Medical University
Gang Li
The First Affiliated Hospital of Wenzhou Medical University
Liang-Jie Tang
The First Affiliated Hospital of Wenzhou Medical University
Rafael S. Rios
The First Affiliated Hospital of Wenzhou Medical University
Giovanni Targher
Azienda Ospedaliera Universitaria Integrata Verona
Christopher D. Byrne
Southampton General Hospital
Xiao-Dong Wang
The First Affiliated Hospital of Wenzhou Medical University
Yong-Ping Chen
The First Affiliated Hospital of Wenzhou Medical University
Ming-Hua Zheng (zhengmh@wmu.edu.cn)
the First Affiliated Hospital of Wenzhou Medical University

Research

Keywords: Gene polymorphisms, Gene-gene interaction, SAMM50, PARVB, PNPLA3, NASH

Posted Date: November 17th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-108430/v1

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

Abstract

Background: Previous studies have reported that single nucleotide polymorphisms (SNPs) in *SAMM50-rs738491, PARVB-rs5764455* and *PNPLA3-rs738409* are associated with the development of non-alcoholic fatty liver disease (NAFLD). However, no studies have examined the effect of interactions between these three genotypes to affect liver disease severity.

Objective: Our aim was to assess the effect of these three SNPs on nonalcoholic steatohepatitis (NASH) and to investigate gene-gene interactions in a Chinese cohort of patients with biopsy-proven NAFLD.

Methods: 415 adult patients with biopsy-proven NAFLD were recruited to the study. Multivariable logistic regression analysis was undertaken to test associations between NASH and SNPs in *SAMM50-rs738491, PARVB-rs5764455* and *PNPLA3-rs738409*. Gene-gene interactions were analyzed by performing a generalized multifactor dimensionality reduction (GMDR) analysis.

Results: The mean age of patients was 41.3±12.5 years and 75.9% of them were men. Patients with *SAMM50-rs738491* TT, *PARVB-rs5764455* AA or *PNPLA3-rs738409* GG genotypes had a higher risk of having NASH even after adjustment for age, sex and body mass index. Furthermore, GMDR analysis showed that the combination of all three SNPs was the best model to predict NASH. Additionally, the odds ratio of the haplotype A-G-T for predicting risk of NASH was nearly three times higher than that of the haplotype G-C-C.

Conclusions: Patients with NAFLD who have *SAMM50-rs738491* TT, *PARVB-rs5764455* AA or *PNPLA3-rs738409* GG genotypes are at higher risk of NASH. Moreover, these SNPs synergistically interact to increase the susceptibility to NASH.

Background

Nonalcoholic steatohepatitis (NASH), i.e. the subtype of nonalcoholic fatty liver disease (NAFLD) presents with lobular inflammation and hepatocyte injury, is affecting about 400 million people around the world.[(1)] It has been estimated that the numbers of patients with NASH will increase 63% from 16.5 million in 2015 to 27 million in 2030 in the United States.[(2)] NASH may progress to cirrhosis and hepatocellular carcinoma, and is also associated with the development of extra-hepatic diseases, including cardiovascular disease and type 2 diabetes (T2DM).[(3–5)]

Recently, studies have shown that NAFLD is not only highly prevalent in overweight-obese individuals, but may also occur in non-obese individuals in which genetic factors have been found to be one of the independent predictors.(6) Strong evidence indicates that genetic and epigenetic factors can affect the development and progression of NAFLD.(3) In recent years, genome-wide association studies (GWAS) have explored the genetic background of NAFLD in different ethnic populations.(7–9) The rs738409 polymorphism in the patatin like phospholipase containing protein-3 (*PNPLA3*) gene has demonstrated to have the strongest effect on the entire histopathological spectrum of NAFLD across different countries and ethnicities.(10) Apart from the *PNPLA3* gene, other genetic variants have also been implicated in the development and progression of NAFLD. In a GWAS study, Kitamoto et al. showed that polymorphisms in the *SAMM50* and *PARVB* genes in addition to those in the *PNPLA3* gene were associated with the development and progression of NAFLD in a Japanese population.(9) Additionally, others have reported a significant association between the *SAMM50* or *PARVB* genotypes and

progression of NAFLD.(11, 12) That said, the contribution of rs738491 in the *SAMM50* gene to the progression of NAFLD from simple fatty liver to NASH remains controversial.(9, 12)

To date, no study has tested the association of the *SAMM50* and *PARVB* genes with presence of NASH in Chinese individuals with biopsy-proven NAFLD. It is important to note that the *PNPLA3, SAMM50* and *PARVB* genes are all located on chromosome 22q13.(9) In addition, the *SAMM50-rs738491*, *PARVB-rs5764455* and *PNPLA3-rs738409* genetic variants are in the same linkage disequilibrium block and are associated with the severity of steatosis and NAFLD activity score (NAS).(9) Based on the theory that complex disease traits are affected by the inheritance of different numbers of variants and also gene-gene interactions,(13) we speculated that there is an interaction between these three genetic polymorphisms to affect liver disease severity in NASH. Currently, there are no studies that have tested the effects of interactions among these three genes to affect disease severity in NAFLD.

Thus, our principle aim was to evaluate whether *SAMM50-rs738491, PARVB-rs5764455* and *PNPLA3-rs738409* genetic variants were associated with NASH and our secondary aim was to investigate the gene-gene interactions and combination effects of these genetic variants on NASH susceptibility in a Chinese population with biopsy-proven NAFLD.

Results

Patient characteristics

A total of 591 patients with histologically proven NAFLD were initially included in the study. After exclusion of patients with alcoholic fatty liver (n = 79), autoimmune liver diseases (n = 1), drug-induced hepatitis (n = 1), hepatitis B or C virus-infected patients (n = 45) or patients with missing records of important laboratory parameters (n = 50), 415 patients with biopsy-proven NAFLD were identified for the final analysis with an average age of 41.3 ± 12.5 years (range 18–72 years), and 75.9% were men. Amongst these patients, 246 (59.3%) had definite NASH. Table 1 shows the demographic and clinical characteristics of participants stratified by presence or absence of NASH. Compared to patients in the no-NASH group, patients with NASH were younger and had a higher BMI, and higher values of homeostasis model assessment of insulin resistance (HOMA-IR), serum liver enzymes and uric acid.

Table 1
Baseline characteristics of participants stratified by NASH status.

	no-NASH	NASH	*P-value*
N (%)	169 (40.7)	246 (59.3)	
Female sex, *n* (%)	35 (20.7)	65 (26.4)	0.181
Age, years	45 (37–52)	39 (29–48)	**< 0.0001**
BMI, kg/m^2	25.4 (23.4–27.3)	27.36 (25-29.2)	**< 0.0001**
Hypertension, *n* (%)			0.727
Yes	59 (34.9)	90 (36.6)	
No	110 (65.1)	156 (63.4)	
Diabetes, *n* (%)			0.610
Yes	40 (23.7)	53 (21.5)	
No	129 (76.3)	193 (78.5)	
HOMA-IR	2.73 (1.79–4.44)	3.99 (2.83–6.35)	**< 0.0001**
ALT, U/L	37 (23–54)	68 (41–124)	**< 0.0001**
AST, U/L	27 (22-36.5)	43 (30-66.2)	**< 0.0001**
GGT, U/L	42 (26-69.5)	56 (38-91.2)	**< 0.0001**
TG, mmol/L	1.81 (1.30–2.59)	1.98 (1.45–2.92)	0.107
TC, mmol/L	4.91 ± 1.12	5.10 ± 1.12	0.094
HDL-C, mmol/L	0.97 (0.85–1.14)	0.98 (0.87–1.14)	0.320
LDL-C, mmol/L	2.95 ± 0.88	3.11 ± 0.91	0.088
UA, mmol/L	364 (310–438)	401 (340.7-481.2)	**0.001**

Sample size n = 415. Data are expressed as numbers (percentages) for categorical variables, as mean ± standard deviation for normally distributed continuous variables and median (inter-quartile range) for skewed distributed continuous variables. Differences between the groups were determined using the Student’s t-test, Mann-Whitney U test for continuous variables and Pearson χ^2, Fisher’s exact test for categorical variables.

Association between genotypes and alleles of rs738491, rs5764455 and rs738409 and NASH

Genotypes of these three SNPs were in Hardy-Weinberg equilibrium for both patient groups ($P_{no-NASH}$ = 0.582, 0.542, 0.745; P_{NASH} = 0.204, 0.885, 0.171). Table 2 shows that the distribution of genotypes and allele frequencies of *SAMM50-rs738491 PARVB-rs5764455* and *PNPLA3-rs738409* were significantly different between patients with, and without, NASH. In addition, Table 3 shows that genotypes and alleles in each SNP had a strong link with the presence of NASH. TT in *SAMM50-rs738491*, AA in *PARVB-rs5764455* and GG in

PNPLA3-rs738409 were significantly associated with an increased risk of having NASH, even after adjusting for age, sex and BMI (adjusted-OR 2.26, 95%CI 1.24 to 4.13; adjusted-OR 3.27, 95%CI 1.73 to 6.18; adjusted-OR 2.89, 95%CI 1.59 to 5.26, respectively). Besides, Table 3 also shows rs738491-T, rs5764455-A and rs738409-G as the risk alleles significantly associated with NASH (adjusted-OR 1.54, 95%CI 1.15 to 2.08; adjusted-OR 1.74, 95%CI 1.29 to 2.35; adjusted-OR 1.73, 95%CI 1.29 to 2.33, respectively).

Table 2

Distribution of genotypes and alleles frequencies of genetic variants in *SAMM50-rs738491*, *PARVB-rs5764455* and *PNPLA3-rs738409* stratified by NASH status.

Genotypes and Alleles	**no-NASH**	**NASH**	χ^2	***P-value***
SAMM50-rs738491				
Genotypes			12.62	**0.002**
CC	43 (25.4)	43 (17.5)		
CT	88 (52.1)	108 (43.9)		
TT	38 (22.5)	95 (38.6)		
Alleles			11.79	**0.001**
C	174 (51.5)	194 (39.4)		
T	164 (48.5)	298 (60.6)		
PARVB-rs5764455				
Genotypes			16.04	**< 0.0001**
GG	64 (37.9)	57 (23.2)		
GA	83 (49.1)	124 (50.4)		
AA	22 (13.0)	65 (26.4)		
Alleles			15.93	**< 0.0001**
G	211 (62.4)	238 (48.4)		
A	127 (37.6)	254 (51.6)		
PNPLA3-rs738409				
Genotypes			14.75	**0.001**
CC	65 (38.5)	61 (24.8)		
CG	78 (46.2)	112 (45.5)		
GG	26 (15.4)	73 (29.7)		
Alleles			15.72	**< 0.0001**
C	208 (61.5)	234 (47.6)		
G	130 (38.5)	258 (52.4)		

Data are expressed as No. (%) and tested by Pearson χ^2 test and Fisher's exact test.

Table 3
Odds ratios for NASH according to the genotypes and alleles of *SAMM50-rs738491*, *PARVB-rs5764455* and *PNPLA3-rs738409*.

Genotypes and Alleles	**Unadjusted odds ratio (95%CI)**	***P*-value**	**Adjusted odds ratio (95%CI)**†	***P*-value**
SAMM50-rs738491				
Genotypes				
CC	1		1	
CT	1.227 (0.739, 2.039)	0.429	1.148 (0.667, 1.975)	0.618
TT	2.500 (1.420, 4.402)	**0.002**	2.262 (1.240, 4.127)	**0.008**
Alleles				
C	1		1	
T	1.630 (1.232, 2.156)	**0.001**	1.544 (1.148, 2.078)	**0.004**
PARVB-rs5764455				
Genotypes				
GG	1		1	
GA	1.677 (1.067, 2.637)	**0.025**	1.610 (0.993, 2.611)	0.053
AA	3.317 (1.819, 6.050)	**< 0.0001**	3.265 (1.725, 6.177)	**< 0.0001**
Alleles				
G	1		1	
A	1.773 (1.337, 2.352)	**< 0.0001**	1.744 (1.293, 2.352)	**< 0.0001**
PNPLA3-rs738409				
Genotypes				
CC	1		1	
CG	1.530 (0.972, 2.408)	0.066	1.415 (0.872, 2.297)	0.16
GG	2.992 (1.696, 5.279)	**< 0.0001**	2.894 (1.593, 5.258)	**< 0.0001**
Alleles				
C	1		1	
G	1.764 (1.331, 2.338)	**< 0.0001**	1.732 (1.285, 2.333)	**< 0.0001**

†Multivariable logistic regression model adjusted for age, sex and BMI.

We also compared the genotype distribution of genetic variants between lean NASH patients (BMI < 25 kg/m^2) and overweight/obese NASH patients (BMI ≥ 25 kg/m^2). **Table S1** shows no significant difference in distribution of these variants between the two groups. In addition, **Table S2** also shows no significant difference in distribution of these variants between NASH patients with, and without T2DM.

Association between genotypes of rs738491, rs5764455 and rs738409 and clinical features

Due to the fact that TT in *SAMM50-rs738491*, AA in *PARVB-rs5764455* and GG in *PNPLA3-rs738409* significantly increased the susceptibility to NASH, we divided patients into a TT group and a CC + CT group according to rs738491; a AA group and a GG + GA group according to rs5764455; and a GG group and a CC + CG group according to rs738409, respectively. Table 4 shows that the proportion of women was greater in the risk genotype group of each SNP. For *SAMM50-rs738491*, the TT group had a higher plasma total cholesterol, LDL-C and HDL-C levels compared to the CC + CT group. As to *PARVB-rs5764455*, the AA group had a higher plasma LDL-C level compared to the GG + GA group. The histological severity of NAFLD (i.e. NAS ≥ 4) significantly differed between the two genotype groups in each SNP (Table 4). After adjusting for age, sex and BMI, all three aforementioned SNPs were strongly associated with NAS ≥ 4, especially with increasing steatosis grade (Table 5). Fibrosis stage, the values of HOMA-IR and the levels of serum liver enzymes did not show any difference between the two groups for each SNP.

Table 4
Baseline characteristics of participants stratified by *SAMM50-rs738491*, *PARVB-rs5764455* and *PNPLA3-rs738409* genotypes.

	SAMM50-rs738491			***PARVB-rs5764455***			***PNPLA3-rs738409***		
	CC + CT	**TT**	P	**GG + GA**	**AA**	P	**CC + CG**	**GG**	P
N**(%)**	282 (67.95)	133 (32.05)		328 (79.04)	87 (20.96)		316 (76.14)	99 (23.86)	
Female sex (%)	58 (20.6)	42 (31.6)	**0.014**	70 (21.30)	30 (34.50)	**0.011**	70 (22.2)	30 (30.3)	0.098
Age, years	43 (32–50)	41 (32–49)	0.655	42 (32–50)	42 (31–52)	0.841	42 (31–50)	42 (33–48)	0.977
BMI, kg/m^2	26.57 (24.22–28.68)	26.95 (24.63–28.80)	0.339	26.72 (24.22–28.67)	26.53 (24.65–28.73)	0.711	26.69 (24.22–28.83)	26.63 (24.62–28.67)	0.893
Hypertension			0.957			0.344			0.913
Yes	101 (35.8)	48 (36.1)		114 (34.8)	35 (40.2)		113 (35.8)	36 (36.4)	
No	181 (64.2)	85 (63.9)		214 (65.2)	52 (59.8)		203 (64.2)	63 (63.6)	
Diabetes			0.337			0.470			0.959
Yes	67 (23.8)	26 (19.5)		76 (23.2)	17 (19.5)		71 (22.5)	22 (22.2)	
No	215 (76.2)	107 (80.5)		252 (76.8)	70 (80.5)		245 (77.5)	77 (77.8)	
HOMA-IR	3.61 (2.43–5.60)	3.27 (2.19–5.15)	0.165	3.61 (2.42–5.44)	3.20 (2.11–5.29)	0.246	3.45 (2.35–5.43)	3.54 (2.22–5.42)	0.504
ALT, U/L	49 (29.7–88)	55 (36–99)	0.168	49 (30-88.7)	60 (37–101)	0.100	49.5 (30-89.7)	56 (36–91)	0.204
AST, U/L	33.5 (25–52)	36 (26–60)	0.202	32.5 (25–52)	38 (26–61)	0.051	33 (25.5–52)	37 (26–59)	0.165
GGT, U/L	51 (33-84.2)	52 (32-80.5)	0.750	52 (33-84.7)	51 (31–77)	0.383	51 (33–83)	51 (33–80)	0.709
TG, mmol/L	1.96 (1.33–2.92)	1.82 (1.44–2.54)	0.303	1.96 (1.40–2.91)	1.82 (1.33–2.50)	0.111	1.95 (1.35–2.92)	1.87 (1.45–2.41)	0.282

Sample size n = 415. Data are expressed as numbers (percentages) for categorical variables, as mean ± standard deviation for normally distributed continuous variables and median (inter-quartile range) for skewed distributed continuous variables. Differences between the groups were determined using the Student's t-test, Mann-Whitney U test for continuous variables and Pearson χ2, Fisher's exact test for categorical variables.

	SAMM50-rs738491			*PARVB-rs5764455*			*PNPLA3-rs738409*		
TC, mmol/L	4.94 ± 1.12	5.20 ± 1.12	**0.027**	4.97 ± 1.11	5.22 ± 1.15	0.059	4.99 ± 1.12	5.13 ± 1.12	0.263
HDL-C, mmol/L	0.96 (0.85–1.12)	1.01 (0.89–1.16)	**0.031**	0.96 (0.85–1.13)	1.03 (0.89–1.15)	0.070	0.96 (0.85–1.13)	1.02 (0.89–1.19)	0.107
LDL-C, mmol/L	2.96 ± 0.89	3.23 ± 0.89	**0.004**	2.98 ± 0.87	3.30 ± 0.96	**0.003**	3.00 ± 0.89	3.18 ± 0.92	0.074
UA, mmol/L	389 (325-470.2)	380 (327–452)	0.741	389 (325–466)	378 (334–456)	0.788	389.5 (325.5–469)	377 (325–441)	0.370
Liver histology									
NAS ≥ 4			**0.001**			**0.002**			**0.001**
Yes	154 (54.6)	95 (71.4)		184 (56.1)	65 (74.7)		176 (55.7)	73 (73.7)	
No	128 (45.4)	38 (28.6)		144 (43.9)	22 (25.3)		140 (44.3)	26 (26.3)	
Significant fibrosis (F ≥ 2)	61 (21.6)	28 (21.1)	0.893	71 (21.6)	18 (20.7)	0.847	69 (21.8)	20 (20.2)	0.730

Sample size n = 415. Data are expressed as numbers (percentages) for categorical variables, as mean ± standard deviation for normally distributed continuous variables and median (inter-quartile range) for skewed distributed continuous variables. Differences between the groups were determined using the Student’s t-test, Mann-Whitney U test for continuous variables and Pearson $\chi 2$, Fisher’s exact test for categorical variables.

Table 5
Associations between SNPs and individual histologic features of NAFLD.

	Steatosis grade†		Hepatocyte ballooning†		Lobular inflammation†		NAS ≥ 4‡		Significant fibrosis (F ≥ 2)‡	
	β (SE)	P-value	**β (SE)**	P-value	**β (SE)**	P-value	**β (SE)**	P-value	**β (SE)**	P-value
SAMM50-rs738491										
CC + CT	Ref.		Ref.		Ref.		Ref.		Ref.	
TT	0.896 (0.202)	**< 0.0001**	-0.113 (0.211)	0.591	0.327 (0.241)	0.175	0.676 (0.241)	**0.005**	-0.115 (0.262)	0.662
PARVB-rs5764455										
GG + GA	Ref.		Ref.		Ref.		Ref.		Ref.	
AA	0.952 (0.233)	**< 0.0001**	-0.064 (0.242)	0.792	0.152 (0.275)	0.581	0.859 (0.289)	**0.003**	-0.144 (0.302)	0.634
PNPLA3-rs738409										
CC + CG	Ref.		Ref.		Ref.		Ref.		Ref.	
GG	0.920 (0.220)	**< 0.0001**	-0.139 (0.230)	0.545	0.344 (0.261)	0.188	0.818 (0.268)	**0.002**	-0.144 (0.289)	0.618

†Multivariable linear regression model adjusted for age, sex and BMI.

‡Multivariable logistic regression model adjusted for age, sex and BMI.

Analysis of gene-gene interactions

A GMDR analysis was performed to examine the gene-gene interactions. This analysis showed that the *PARVB-rs5764455* was the best single-locus model that predicted the presence of NASH. The best two-locus model was the combination of the *SAMM50-rs738491* and *PNPLA3-rs738409*. The three-locus model had perfect Cross-Validation Consistency (CVC: 10/10), high testing accuracy (57.23%) and $P = 0.011$, which appeared to be the best predictive model (Table 6). All these models were adjusted by age, sex and BMI. When we analyzed the strength of the association between each genetic model and NASH after controlling for age, sex and BMI, we found that OR values increased progressively with the increasing number of loci included (adjusted-OR 2.431, 95%CI 1.384 to 4.269 for a one-locus model; adjusted-OR 2.453, 95%CI 1.413 to 4.259 for a two-locus model; and adjusted-OR 2.751, 95%CI, 1.452 to 5.211 for a three-locus model, respectively). Importantly, the OR of the

haplotype A-G-T for predicting risk of NASH compared to G-C-C was nearly three times higher compared to that of the haplotype G-C-C. Additionally, the OR of haplotype A-G-T to G-C-C was nearly twice compared to haplotypes G-T to C-C or A to G, suggesting that the combination of these three risk alleles may significantly increase the susceptibility to NASH (Table 7).

Table 6
Best models to predict the presence of NASH by generalized multifactor dimensionality reduction (GMDR) analysis†.

GMDR model	Training Accuracy (%)	Testing Accuracy (%)	Sign Test (*P*)	CVC
PARVB-rs5764455	57.71	52.05	8 (0.055)	4/10
PNPLA3-rs738409 SAMM50-rs738491	59.90	57.80	8 (0.055)	10/10
PARVB-rs5764455 PNPLA3-rs738409 SAMM50-rs738491	60.59	57.23	9 (0.011)	10/10

†Adjusted for age, sex and BMI.

CVC = Cross-Validation Consistency

Table 7
Interaction analysis between each best genetic model and risk of having NASH.

Genotypes and alleles	Number (%)		Adjusted OR	*P-value*
	no-NASH	NASH	(95%CI)[†]	
PARVB-rs5764455				
Genotypes				
GG + GA	147 (86.98)	181 (73.58)	1	
AA	22 (13.02)	65 (26.42)	2.431(1.384,4.269)	**0.002**
Alleles				
G	211 (62.43)	238 (48.37)	1	
A	127 (37.57)	254 (51.63)	1.744 (1.293,2.352)	**< 0.0001**
PNPLA3-rs738409 SAMM50-rs738491				
Genotypes				
CC + CG, CC + CT	129 (84.31)	147 (68.06)	1	
GG, TT	24 (15.69)	69 (31.94)	2.453 (1.413,4.259)	**0.001**
Alleles				
C-C	172 (57.33)	185 (42.63)	1	
G-T	128 (42.67)	249 (57.37)	1.733 (1.261,2.381)	**0.001**
PARVB-rs5764455 PNPLA3-rs738409 SAMM50-rs738491				
Genotypes				
GG + GA, CC + CG, CC + CT	129 (88.97)	144 (73.85)	1	
AA, GG, TT	16 (11.03)	51 (26.15)	2.751 (1.452,5.211)	**0.002**
Alleles				
G-C-C	105 (70.00)	94 (41.23)	1	
A-G-T	45 (30.00)	134 (58.77)	3.157 (1.988,5.012)	**< 0.0001**

[†]Multivariable logistic regression model was adjusted for age, sex and BMI.

Discussion

The main and novel results of this study are that genetic variants of *SAMM50-rs738491*, *PARVB-rs5764455* and *PNPLA3-rs738409* are independently associated with the presence of NASH in Chinese individuals with biopsy-proven NAFLD. Furthermore, we have shown for the first time that *SAMM50-rs738491*, *PARVB-rs5764455* and *PNPLA3-rs738409* genetic variants interact synergistically to increase the susceptibility to NASH.

Recent studies have shown that various genes and SNPs play important roles in the development and progression of NAFLD.(3) In addition, it is known that the magnitude of the NAFLD-related SNP effect varies markedly in different ethnic populations.(14) Based on GWAS results, polymorphisms in *SAMM50-rs738491, PARVB-rs5764455* and *PNPLA3-rs738409* genes were found to be associated with the development and progression of NAFLD, while the association between these three SNPs and risk of having NASH was not significant.(9) With regard to the Chinese population, two prior studies have compared genetic variants in the *SAMM50* and *PARVB* genes with the presence of NAFLD detected by liver ultrasonography.(11, 12) In our study, involving patients with biopsy-confirmed NAFLD, we found that patients with TT in *SAMM50-rs738491*, AA in *PARVB-rs5764455* or GG in *PNPLA3-rs738409* were more likely to have NASH, and rs738491-T, rs5764455-A and rs738409-G were the risk alleles that were strongly associated with risk of NASH, even after adjusting for age, sex and BMI. We undertook a subgroup analysis comparing variants in NASH patients, stratified by BMI or T2DM. These data indicated that these three genetic variants play an independent role in the development of NAFLD, regardless of BMI or T2DM. However, it is necessary to increase the sample size of this cohort to further validate this observation.

Previous studies have found that the *PNPLA3, SAMM50* and *PARVB* genes may be implicated in regulating lipid metabolism,(15) maintaining mitochondrial morphology(16) and activating Akt/protein kinase B (PKB) signaling pathways,(17) respectively. Specifically, the PNPLA3 protein is an enzyme with triacylglycerol lipase and acylglycerol O-acyltransferase activity, which promotes remodeling of lipid droplets in hepatocytes.(15) The *PNPLA3-rs738409* C > G genotype reduces this enzymatic activity leading to the entire histopathological spectrum of NAFLD from simple steatosis to cirrhosis and hepatocellular carcinoma.(3) In our study, the analysis of liver histology features showed that the GG genotype of rs738409 was associated with a higher proportion of NAS $\geq$ 4 (especially with a higher steatosis grade), but not with a more atherogenic lipid profile. This finding is also in line with data from Speliotes et al. showing that the *PNPLA3*-rs738409 genetic variant specifically conferred an increased risk of histologic fat accumulation but not NAFLD-related metabolic traits. (18) The *SAMM50* gene encodes the protein Sam50 that is part of the sorting and assembly machinery (SAM), which is necessary for assembling β-barrel proteins located in the outer membrane of mitochondria.(16) β-barrel proteins play a crucial role in maintaining mitochondrial shape, morphology of mitochondrial cristae, and in assembling the respiratory chain complex.(16, 19) Abnormalities of β-barrel proteins lead to mitochondrial dysfunction, which in turn reduces the removal of reactive oxygen species (ROS).(9) ROS accumulation will lead to the destruction of organelles oxidizing fatty acids and promote lipotoxicity, which can be deleterious for hepatocytes.(20) Based on the fact that loss of mitochondrial cristae has been found in the livers of NASH patients after depletion of the Sam50,(21) there exists an association between *SAMM50-rs738491* and increasing pro-inflammatory factors in hepatocytes. Parvin-β, encoded by the *PARVB* gene, takes part in forming integrin-linked kinase-pinch-parvin complex, which transmits signals from integrin to Akt/PKB.(17) Kinases of Akt/PKB modulate basic cell activation to produce proinflammatory mediators.(22) Over-expression of parvin-β will not only lead to a concomitant increase in lipogenic gene expression,(23) but also promote apoptosis.(24) Increased hepatic lipogenesis and apoptosis are considered part of the progressive disorder that occurs with

NAFLD progression from simple steatosis to NASH and fibrosis.(25) Previous studies have shown that the rs5764455 in the *PARVB* gene had the most robust association with NASH.(9) Similarly, for rs5764455 in the *PARVB* gene, the OR for NASH was also the highest amongst all the SNPs that we investigated, comparing the A allele to the G allele.

The progression from simple steatosis to NASH is recognized to have two consecutive steps which are fat accumulation followed by necro-inflammation in the liver.(26) Our study suggests that the *PNPLA3-rs738409* and *SAMM50-rs738491* genetic variants might be mainly involved into the first and second steps of NASH progression, respectively, while the *PARVB-rs5764455* variant might play a role in both of these two steps.

All the findings mentioned above have indicated an association between these three SNPs and NASH and also supported possible underlying mechanisms. However, it has been shown that SNPs identified by GWAS only explain a small part of disease etiology, because the relatedness between complex diseases and multiple genes and/or their interactions are ignored(27). Therefore, analyses emphasizing a gene-gene interaction have been one of the new ways to better understand the aetiology of common complex traits. The GMDR methodology is effective and is widely used in detecting gene-gene and gene-environment interactions in various diseases.(28) Our results, obtained using the GMDR method, have identified *PARVB-rs5764455* as the best single-locus model and the combination of *PNPLA3-rs738409* and *SAMM50-rs738491* as the best double-locus model for predicting NASH. These conclusions further confirmed the correctness of the hypothesis we mentioned above, i.e. the *PNPLA3* and *SAMM50* genes were, respectively, involved in the first and second stages of progression from simple steatosis to NASH, while the *PARVB* gene was implicated in both of these stages. Our study showed that the susceptibility to NASH in patients with haplotype A-G-T was three times higher than patients carrying the haplotype G-C-C. In addition, the OR of haplotype A-G-T to G-C-C for predicting risk of NASH was nearly twice that haplotype G-T to C-C or A to G, thereby indicating a synergistic interaction between these three SNPs.

Our study has some important limitations that should be considered. Firstly, our study did not include a cohort of control subjects without NAFLD to investigate the association of the studied genetic variants in subjects without NAFLD. Secondly, only one SNP in each of the *SAMM50, PARVB* and *PNPLA3* genes was chosen. The limited number of SNPs was therefore not likely to capture most of the genetic information conveyed by these three genes. Thirdly, our results supporting the existence of a synergistic interaction between these three SNPs needs further validation in larger cohorts of patients of different ethnicity. Lastly, the underlying mechanisms that explain the interaction linking these three SNPs to the pathogenesis of NASH need to be better elucidated.

Conclusions

In conclusion, our novel results show that genetic variants of *SAMM50-rs738491*, *PARVB-rs5764455* and *PNPLA3-rs738409* are independently associated with the presence of NASH in Chinese individuals with biopsy-proven NAFLD. Furthermore, we have also shown for the first time that *SAMM50-rs738491*, *PARVB-rs5764455* and *PNPLA3-rs738409* genetic variants interact synergistically to increase the susceptibility to NASH.

Materials And Methods

Study population

This study involved analysis of data from the well-characterized Prospective Epidemic Research Specifically OF NASH (PERSONS).(29, 30) Patients from 18–75 years old with suspected NAFLD (i.e. defined as fatty liver on imaging methods, and/or elevated serum liver enzyme levels) were consecutively recruited for liver biopsy examination at the First Affiliated Hospital of Wenzhou Medical University from December 2016 to November 2018. Participants were excluded for the following reasons: (1) excessive alcohol consumption (> 140 g/week for men and > 70 g/week for women); (2) viral hepatitis (based on serum viral B or C markers) and autoimmune hepatitis (based on serum autoantibodies and histology); (3) use of potentially hepatotoxic drugs (i.e. more than 6 months and serum liver enzyme levels more than three times of normal upper limit); (4) liver cancer (based on imaging or pathological data which suspected the possibility of liver cancer according to the Clinical Practice Guidelines for hepatocellular carcinoma published in 2012 by the European Society of Study of the Liver (EASL) (31)); (5) missing data on genetic polymorphisms or other important laboratory parameters; and (6) hepatic steatosis < 5% on liver histology. All patients signed a written informed consent, before participation in this study. The study was approved by the internal review board for ethics of the First Affiliated Hospital of Wenzhou Medical University; the study protocol was registered in the Chinese Clinical Trial Registry (ChiCTR-EOC-17013562).

Clinical and laboratory data

We measured baseline characteristics including demographics, anthropometry, clinical parameters and comorbidities from all participants. Body mass index (BMI) was calculated as weight in kilograms divided by height in meters squared. Blood pressure was measured by a standardized method.(30) Hypertension was defined as blood pressure ≥ 140/90 mmHg or anti-hypertensive drug use. Type 2 diabetes mellitus (T2DM) was diagnosed by fasting glucose level ≥ 7.0 mmol/L or history of prescribed hypoglycaemic agents. Blood biochemical parameters including alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyltransferase (GGT), triglycerides (TG), total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and uric acid (UA) were assessed by an automated analyzer (Abbott AxSYM, Park, IL) using standard laboratory methods.

Liver histology

Liver biopsy specimens were revised by a single liver pathologist (X.D. Wang), who was blinded to participants' clinical data. Histological parameters of NAFLD were scored based on the NASH Clinical Research Network classification.(32) NASH was diagnosed as presence of NAS ≥ 4 and ≥ 1 for each of the three components including hepatic steatosis, lobular inflammation, and ballooning.(33–35) Significant fibrosis was defined by fibrosis ≥ F2 on histology according to Brunt's criteria.(36)

Genetic analysis

20 ng of genomic DNA from white blood cells from each participant was extracted for genetic analysis of *SAMM50-rs738491, PARVB-rs5764455* and *PNPLA3-rs738409*. Genotyping of these three SNPs were evaluated by the MassARRAY System (Agena Bioscience, San Diego, CA, USA). DNA samples were firstly amplified through locus-specific polymerase chain reaction (PCR) according to Assay Design Suite software (Version 3.1). Detection of rs738491, rs5764455 and rs738409 genotypes was performed by Matrix-Assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF) mass spectrometry.

Statistical analysis

All statistical analyses were undertaken with SPSS/PC version 25.0 (Chicago, IL, USA) software. Continuous variables were compared by the Student's t-test or Mann-Whitney U test (if not normally distributed) and expressed as mean ± standard deviation (SD) or median (25th, 75th centiles), respectively. Categorical variables were analyzed by the χ^2 test or the Fisher's exact test and expressed as number (%). Hardy-Weinberg equilibrium for the three aforementioned SNPs in the no-NASH group and the NASH group were tested by the χ^2 test. The distribution of genotypes and alleles between the NASH group and no-NASH group were compared by the χ^2 test or Fisher's exact test. Logistic and linear regression analyses were performed to test the associations of genetic variants with individual histological features of NAFLD, after adjustment for age, sex and body mass index (BMI). The strength of associations is presented as odds ratios (OR) with 95% confidence intervals (CI). Generalized multifactor dimensionality reduction (GMDR)(37) analysis was used to evaluate gene-gene interactions, which provided cross-validation consistency, training accuracy, testing accuracy, and the sign test (P). The detailed analytical procedure and the definition of each output parameter have been described elsewhere.(38) The effect of interaction between genotypes on the risk of NASH was assessed by logistic regression. A two-sided level of $P < 0.05$ was considered statistically significant. Power calculations were completed using the CATS Genetic Power Calculator,(39) with settings of a multiplicative genetic model. The prevalence of NASH was approximately 0.025, as estimated in previous studies.(40) Assuming a minor allele frequency (MAF) for rs738491 of 0.5566 (the frequency found in our population), and an odds ratio (OR) of 1.6, the expected power for a one-stage study is 89.2% in our cohort. Using the same assumption for rs5764455 and rs738409, with MAF of 0.4590 and 0.4675 separately, the expeted power was 91.0% for both SNPs.

Abbreviations

BMI = body mass index, HOMA-IR = homeostasis model assessment of insulin resistance, ALT = alanine aminotransferase, AST = aspartate aminotransferase, GGT = gamma-glutamyltransferase, TG = triglycerides, TC = total cholesterol, HDL-C = high-density lipoprotein cholesterol, LDL-C = low-density lipoprotein cholesterol, UA = uric acid.

Declarations

Ethics approval statement

Ethical approval was obtained from the First Affiliated Hospital of Wenzhou Medical University Ethics Committee and the study protocol has been registered at the Chinese Clinical Trial Registry (ChiCTR-EOC-17013562). Each subject has provided written informed consent.

Consent for publication

All authors contributed to the manuscript for important intellectual content and approved the submission.

Availability of data and materials

Not applicable

Conflicts of interest

The authors declare no conflicting interest.

Funding

This work was supported by grants from the National Natural Science Foundation of China (81500665), High Level Creative Talents from Department of Public Health in Zhejiang Province and Project of New Century 551 Talent Nurturing in Wenzhou. G. T. is supported in part by grants from the School of Medicine, University of Verona, Verona, Italy. CDB. is supported in part by the Southampton NIHR Biomedical Research Centre (IS-BRC-20004), UK.

Authors' Contributions

Study concept and design: Ke Xu and Ming-Hua Zheng

Acquisition of data: Pei-Wu Zhu, Wen-Yue Liu, Hong-Lei Ma, Gang Li, Liang-Jie Tang, Rafael S. Rios

Pathology analysis: Xiao-Dong Wang

Drafting of the manuscript: Ke Xu and Kenneth I. Zheng

Critical revision: Giovanni Targher, Christopher D. Byrne

Statistical analysis: Ke Xu and Kenneth I. Zheng

Study supervision: Yong-Ping Chen and Ming-Hua Zheng

Acknowledgements

Not applicable

References

1. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology. 2016;64(1):73-84.
2. Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology. 2018;67(1):123-33.
3. Eslam M, George J. Genetic contributions to NAFLD: leveraging shared genetics to uncover systems biology. Nature reviews Gastroenterology & hepatology. 2020;17(1):40-52.
4. Angulo P. Long-term mortality in nonalcoholic fatty liver disease: is liver histology of any prognostic significance? Hepatology. 2010;51(2):373-5.
5. Zheng KI, Fan JG, Shi JP, Wong VW, Eslam M, George J, et al. From NAFLD to MAFLD: a "redefining" moment for fatty liver disease. Chin Med J (Engl). 2020.
6. Molina-Molina E, Krawczyk M, Stachowska E, Lammert F, Portincasa P. Non-Alcoholic Fatty Liver Disease in Non-Obese Individuals: Prevalence, Pathogenesis and Treatment. Clin Res Hepatol Gastroenterol. 2019;43(6):638-45.

7. Chalasani N, Guo X, Loomba R, Goodarzi MO, Haritunians T, Kwon S, et al. Genome-wide association study identifies variants associated with histologic features of nonalcoholic Fatty liver disease. Gastroenterology. 2010;139(5):1567-76, 76.e1-6.
8. Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R, Kim LJ, Palmer CD, et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. PLoS genetics. 2011;7(3):e1001324.
9. Kitamoto T, Kitamoto A, Yoneda M, Hyogo H, Ochi H, Nakamura T, et al. Genome-wide scan revealed that polymorphisms in the PNPLA3, SAMM50, and PARVB genes are associated with development and progression of nonalcoholic fatty liver disease in Japan. Human genetics. 2013;132(7):783-92.
10. Sookoian S, Pirola CJ. Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of nonalcoholic fatty liver disease. Hepatology. 2011;53(6):1883-94.
11. Wu G, Wang K, Xue Y, Song G, Wang Y, Sun X, et al. Association of rs5764455 and rs6006473 polymorphisms in PARVB with liver damage of nonalcoholic fatty liver disease in Han Chinese population. Gene. 2016;575(2 Pt 1):270-5.
12. Chen L, Lin Z, Jiang M, Lu L, Zhang H, Xin Y, et al. Genetic Variants in the SAMM50 Gene Create Susceptibility to Nonalcoholic Fatty Liver Disease in a Chinese Han Population. Hepatitis monthly. 2015;15(10):e31076.
13. Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB. Genetic architecture: the shape of the genetic contribution to human traits and disease. Nature reviews Genetics. 2018;19(2):110-24.
14. Li YY. Genetic and epigenetic variants influencing the development of nonalcoholic fatty liver disease. World journal of gastroenterology. 2012;18(45):6546-51.
15. Jenkins CM, Mancuso DJ, Yan W, Sims HF, Gibson B, Gross RW. Identification, cloning, expression, and purification of three novel human calcium-independent phospholipase A2 family members possessing triacylglycerol lipase and acylglycerol transacylase activities. The Journal of biological chemistry. 2004;279(47):48968-75.
16. Ott C, Ross K, Straub S, Thiede B, Gotz M, Goosmann C, et al. Sam50 functions in mitochondrial intermembrane space bridging and biogenesis of respiratory complexes. Molecular and cellular biology. 2012;32(6):1173-88.
17. Kimura M, Murakami T, Kizaka-Kondoh S, Itoh M, Yamamoto K, Hojo Y, et al. Functional molecular imaging of ILK-mediated Akt/PKB signaling cascades and the associated role of beta-parvin. Journal of cell science. 2010;123(Pt 5):747-55.
18. Speliotes EK, Butler JL, Palmer CD, Voight BF, Hirschhorn JN. PNPLA3 variants specifically confer increased risk for histologic nonalcoholic fatty liver disease but not metabolic disease. Hepatology. 2010;52(3):904-12.
19. Ott C, Dorsch E, Fraunholz M, Straub S, Kozjak-Pavlovic V. Detailed analysis of the human mitochondrial contact site complex indicate a hierarchy of subunits. PloS one. 2015;10(3):e0120213.
20. Suzuki A, Diehl AM. Nonalcoholic Steatohepatitis. Annual review of medicine. 2017;68:85-98.
21. Sanyal AJ, Campbell-Sargent C, Mirshahi F, Rizzo WB, Contos MJ, Sterling RK, et al. Nonalcoholic steatohepatitis: association of insulin resistance and mitochondrial abnormalities. Gastroenterology.

2001;120(5):1183-92.

22. Abraham E. Akt/protein kinase B. Critical care medicine. 2005;33(12 Suppl):S420-2.
23. Johnstone CN, Mongroo PS, Rich AS, Schupp M, Bowser MJ, Delemos AS, et al. Parvin-beta inhibits breast cancer tumorigenicity and promotes CDK9-mediated peroxisome proliferator-activated receptor gamma 1 phosphorylation. Molecular and cellular biology. 2008;28(2):687-704.
24. Zhang Y, Chen K, Tu Y, Wu C. Distinct roles of two structurally closely related focal adhesion proteins, alpha-parvins and beta-parvins, in regulation of cell morphology and survival. The Journal of biological chemistry. 2004;279(40):41695-705.
25. Fujii H, Kawada N. Inflammation and fibrogenesis in steatohepatitis. Journal of gastroenterology. 2012;47(3):215-25.
26. Day CP, James OF. Steatohepatitis: a tale of two "hits"? Gastroenterology. 1998;114(4):842-5.
27. Eslam M, Valenti L, Romeo S. Genetics and epigenetics of NAFLD and NASH: Clinical impact. Journal of hepatology. 2018;68(2):268-79.
28. Xu HM, Xu LF, Hou TT, Luo LF, Chen GB, Sun XW, et al. GMDR: Versatile Software for Detecting Gene-Gene and Gene-Environ- ment Interactions Underlying Complex Traits. Current genomics. 2016;17(5):396-402.
29. Liu WY, Zheng KI, Pan XY, Ma HL, Zhu PW, Wu XX, et al. Effect of PNPLA3 polymorphism on diagnostic performance of various noninvasive markers for diagnosing and staging nonalcoholic fatty liver disease. Journal of gastroenterology and hepatology. 2019.
30. Sun DQ, Zheng KI, Xu G, Ma HL, Zhang HY, Pan XY, et al. PNPLA3 rs738409 is associated with renal glomerular and tubular injury in NAFLD patients with persistently normal ALT levels. Liver international : official journal of the International Association for the Study of the Liver. 2020;40(1):107-19.
31. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. Eur J Cancer. 2012;48(5):599-641.
32. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology. 2005;41(6):1313-21.
33. Harrison SA, Bashir MR, Guy CD, Zhou R, Moylan CA, Frias JP, et al. Resmetirom (MGL-3196) for the treatment of non-alcoholic steatohepatitis: a multicentre, randomised, double-blind, placebo-controlled, phase 2 trial. Lancet (London, England). 2019;394(10213):2012-24.
34. Friedman SL, Ratziu V, Harrison SA, Abdelmalek MF, Aithal GP, Caballeria J, et al. A randomized, placebo-controlled trial of cenicriviroc for treatment of nonalcoholic steatohepatitis with fibrosis. Hepatology. 2018;67(5):1754-67.
35. Neuschwander-Tetri BA, Loomba R, Sanyal AJ, Lavine JE, Van Natta ML, Abdelmalek MF, et al. Farnesoid X nuclear receptor ligand obeticholic acid for non-cirrhotic, non-alcoholic steatohepatitis (FLINT): a multicentre, randomised, placebo-controlled trial. Lancet (London, England). 2015;385(9972):956-65.
36. Brunt EM, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. The American journal of gastroenterology. 1999;94(9):2467-74.
37. Lou XY, Chen GB, Yan L, Ma JZ, Zhu J, Elston RC, et al. A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. American journal of human genetics. 2007;80(6):1125-37.

38. Tunesi S, Ferrante D, Mirabelli D, Andorno S, Betti M, Fiorito G, et al. Gene-asbestos interaction in malignant pleural mesothelioma susceptibility. Carcinogenesis. 2015;36(10):1129-35.
39. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet. 2006;38(2):209-13.
40. Younossi Z, Tacke F, Arrese M, Chander Sharma B, Mostafa I, Bugianesi E, et al. Global Perspectives on Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis. Hepatology. 2019;69(6):2672-82.

Supplementary Files

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

- TableS1.docx
- TableS1.docx
- TableS1.docx
- TableS2.docx
- TableS2.docx
- TableS2.docx