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

Takuya Kitamoto[1], Aya Kitamoto[1], Masato Yoneda[2], Hideyuki Hyogo[3], Hidenori Ochi[3], Takahiro Nakamura[4], Hajime Teranishi[5], Seiho Mizusawa[5], Takato Ueno[6], Kazuaki Chayama[3], Atsushi Nakajima[2], Kazuwa Nakao[1, 7], Akihiro Sekine[1], and Kikuko Hotta[1]

[1]EBM Research Center, Kyoto University Graduate School of Medicine, Kyoto, Japan
[2]Division of Gastroenterology, Yokohama City University Graduate School of Medicine, Yokohama, Japan
[3]Department of Medicine and Molecular Science, Division of Frontier Medical Science, Programs for Biomedical Research, Graduate School of Biomedical Sciences, Hiroshima University, Hiroshima, Japan
[4] Laboratory for Mathematics, National Defense Medical College, Tokorozawa, Japan
[5]Center for Genomic Medicine, Unit of Genome Informatics, Kyoto University Graduate School of Medicine, Kyoto, Japan
[6]Research Center for Innovative Cancer Therapy, Kurume University, Kurume, Japan
Department of Medicine and Clinical Science, Kyoto University Graduate School of Medicine, Kyoto, Japan
[7]Department of Medicine and Clinical Science, Kyoto University Graduate School of Medicine, Kyoto, Japan

Corresponding author:
    Kikuko Hotta, MD, PhD
    Assistant Professor
    EBM Research Center, Kyoto University Graduate School of Medicine
    Yoshida-Konoecho, Sakyo-ku, Kyoto
     606-8501, Japan
    Phone: +81-75-751-2022
    Fax: +81-75-751-2082
    E-mail: kikukoh@kuhp.kyoto-u.ac.jp

Abstract

We examined the genetic background of nonalcoholic fatty liver disease (NAFLD) in the

Japanese population, by performing a genome-wide association study (GWAS).   For GWAS,

372 Japanese NAFLD subjects and 934 control individuals were analyzed.   For replication

studies, 172 NAFLD and 1012 control subjects were monitored.   After quality control, 261,540

single-nucleotide polymorphisms (SNPs) in autosomal chromosomes were analyzed by using a

trend test.   Association analysis was also performed by using multiple logistic regression

analysis using genotypes, age, gender and body mass index (BMI) as independent variables.

Multiple linear regression analyses were performed to evaluate allelic effect of significant SNPs

on biochemical traits and histological parameters adjusted by age, gender, and BMI.

Rs738409 in the *PNPLA3* gene was most strongly associated with NAFLD after adjustment (P

$= 6.8 \times 10^{-14}$, odds ratio = 2.05).   Rs2896019, and rs381062 in the *PNPLA3* gene, rs738491,

rs3761472, and rs2143571 in the *SAMM50* gene, rs6006473, rs5764455, and rs6006611 in the

PARVB gene had also significant P-values ($<2.0 \times 10^{-10}$) and high odds ratios (1.84 to 2.02).

These SNPs were found to be in the same linkage disequilibrium block and were associated

with decreased serum triglycerides and increased aspartate aminotransferase (AST) and alanine

aminotransferase (ALT) in NAFLD patients.   These SNPs were associated with steatosis grade

and NAFLD activity score (NAS).   Rs738409, rs2896019, rs738491, rs6006473, rs5764455,

and rs6006611 were associated with fibrosis.   Polymorphisms in the *SAMM50* and *PARVB*

genes in addition to those in the *PNPLA3* gene were observed to be associated with the

development and progression of NAFLD.

Introduction

Nonalcoholic fatty liver disease (NAFLD) is now recognized as an important health concern (Angulo 2002; Farrell 2003).   NAFLD has a broad spectrum of effects, including simple steatosis, nonalcoholic steatohepatitis (NASH), fibrosis/cirrhosis, and hepatocellular carcinoma.   Excess fat accumulation in the liver is observed in 20–30% of the population in American and European countries, where NASH is associated with approximately 1–3% of the population (Ludwig et al. 1980).   NAFLD is now considered to be a part of metabolic syndrome (Marchesini et al. 2001, Stefan N, et al. 2008).   Genetic as well as environmental factors are important in the development of NAFLD (Wilfred et al. 2008).

Single-nucleotide polymorphisms (SNPs) are useful tools for identifying genetic factors and have been intensively investigated for various common diseases.   We previously reported that variations in peroxisome proliferator-activated receptor γ coactivator 1α (*PPARGC1A*), angiotensin II type 1 receptor (*ATGR1*), and nitric oxide synthase 2 (inducible) (*NOS2*) genes are associated with NAFLD in Japanese individuals (Yoneda et al 2008; Yoneda et al. 2009; Yoneda et al. 2009).

Genome-wide association studies (GWASs) have revealed that SNPs in the patatin-like phospholipase domain containing 3 (*PNPLA3*) and other genes influence NAFLD and liver enzyme levels in the plasma (Romeo et al. 2008; Chalasani et al. 2010; Speliotes et al. 2011; Kawaguchi et al. 2012). We previously reported that the risk allele (G-allele) of *PNPLA3* rs738409 is strongly associated with

NAFLD as well as with increases in aspartate transaminase (AST), alanine transaminase (ALT), ferritin

levels, and fibrosis stage in the patients with NAFLD in the Japanese population (Hotta et al. 2010).

To elucidate the detailed genetic background of NAFLD in the Japanese population, we

performed genome-wide analysis for NAFLD.

Materials and Methods

Subjects

For GWAS, 392 Japanese patients with NAFLD (NAFLD-1; 345 with NASH and 47 with simple

steatosis) were enrolled.   Genome-wide scan data for 934 general Japanese control subjects (control-1)

described in the JSNP database (IMS-JST: Institute of Medical Science-Japan Science and Technology

Agency Japanese SNP database, http://snp.ims.u-tokyo.ac.jp/) were used for GWAS.   For the replication

study, 172 patients with NAFLD (NAFLD-2, 97 with NASH, 4 with simple steatosis, and 71 with

NAFLD) and 1012 control subjects (control-2) were analyzed.   Control-2 subjects included Japanese

volunteers who had undergone medical examination for common disease screening.   All the NAFLD-1

and 101 NAFLD-2 patients underwent liver biopsy.   Computed tomography (CT) or magnetic resonance

imaging (MRI) was performed on 71 NAFLD-2 patients.   Patients with the following diseases were

excluded from the study: viral hepatitis (hepatitis B and C, Epstein–Barr virus infection), autoimmune

hepatitis, primary biliary cirrhosis, sclerosing cholangitis, hemochromatosis, α_1-antitrypsin deficiency,

Wilson's disease, drug-induced hepatitis, and alcoholic hepatitis (present or past daily consumption of

more than 20 g alcohol per day).   None of the patients showed clinical evidence of hepatic

decompensation such as hepatic encephalopathy, ascites, variceal bleeding, or a serum bilirubin level

greater than 2-fold the normal upper limit.

Liver biopsy tissues were stained with hematoxylin and -eosin, reticulin, and Masson's

trichrome stain.   Histological criterion for NAFLD diagnosis is macrovesicular fatty change in

hepatocytes with displacement of the nucleus toward the cell edge (Sanyal 2002).   When more than 5%

of hepatocytes are affected by macrovesicular steatosis, patients are diagnosed as having either steatosis

or NASH.   The minimal criteria for the diagnosis of NASH includes the presence of >5%

macrovesicular steatosis, inflammation, and liver cell ballooning, typically with predominantly

centrilobular (acinar zone 3) distribution (Matteoni et al.1999; Teli et al. 1995).   Steatosis degree was

graded as follows based on the percentage of hepatocytes containing macrovesicular fat droplets: grade 0,

no steatosis; grade 1, <33% hepatocytes containing macrovesicular fat droplets; grade 2, 33–66% of

hepatocytes containing macrovesicular fat droplets; and grade 3, >66% of hepatocytes containing

macrovesicular fat droplets (Brunt 2001).   The activity of hepatitis (necroinflammatory grade) was also

determined on the basis of the composite NAFLD activity score (NAS) as described by Kleiner et al.

(2005).   NAS is the unweighted sum of the scores for steatosis, lobular inflammation, and hepatocellular

ballooning, and ranges from 0 to 8.   Fibrosis severity was scored according to the method of Brunt

(2001) and was expressed on a 4-point scale, as follows: 0, none; 1, perivenular and/or perisinusoidal

fibrosis in zone 3; 2, combined pericellular portal fibrosis; 3, septal/bridging fibrosis; 4, cirrhosis.

Entire study was conducted in accordance with the guidelines of the Declaration of Helsinki.

Written informed consent was obtained from each subject, and the protocol was approved by the ethics

committee of Kyoto University, Yokohama City University, Hiroshima University, and Kurume

University.

Clinical and laboratory evaluation

The weight and height of patients were measured using a calibrated scale after removing shoes and heavy

clothing, if present.   Venous blood samples were obtained from subjects after overnight fasting (12 h) to

measure plasma glucose, hemoglobin A1c (HbA1c), total cholesterol, high-density lipoprotein (HDL)

cholesterol, triglycerides, serum AST, ALT, iron, ferritin, hyaluronic acid, and type IV collagen 7S.   All

the laboratory biochemical parameters were measured using conventional methods.

DNA preparation, genome-wide genotyping and quality control

Genomic DNA was extracted using Genomix (Talent Srl, Trieste, Italy) for blood samples collected from

each subject.   Genome scans were conducted for NAFLD-1 patients using the Human660W-Quad

BeadChip (n = 104) or the HumanOmniExpress BeadChip (n = 288; Illumina, Inc., San Diego, CA, USA).

Genome scan data from control-1(n = 934) were genotyped using the Illumina HumanHap550 BeadChip

and 515,286 SNPs in the autosomal chromosome were available in the JSNP database.   A total of

295,887 common SNPs in the autosomal chromosomes were determined among 3 BeadChips.

Individual call rates were all > 0.99 in the patients and the control group.   A total of 31,177 SNPs with

minor allele frequency (<0.01), 901 SNPs with a lower success rate (<0.95), and 2,269 SNPs with

distorted Hardy-Weinberg equilibrium ($P < 0.001$) were excluded; thus, 261,540 SNPs were subjected to

case-control association analysis.   Using phase II and III HapMapJPT, HCB and CEU data

(http://hapmap.ncbi.nlm.nih.gov/), we confirmed that NAFLD-1 subjects in this study were derived from

the Japanese population by using multi-dimensional scaling (MDS) analysis (Supplementary Fig. 1).

The number of alleles that shared identity-by-descent (PI_HAT) was calculated, and it was found that the

PI_HAT value was less than 0.05 for the NAFLD-1 patients.

For the replication study, Invader probes (Third Wave Technologies, Madison, WI, USA) were

constructed for 56 SNPs with P-values less than 5.0×10^{-5}.   SNPs were genotyped for NAFLD-2 and

control-2 using Invader assays as previously described (Ohnishi et al. 2001).   The success rates of the

Invader assays were > 99.0%.   To validate GWAS, NAFLD-1 patients analyzed by GWAS were also

genotyped using the Invader assay and SNPs with concordance rate of both genotyping more than 99%

were used for further analysis.   Thirteen SNPs showed a lower concordance rate (<0.99) and were

excluded from further analysis.

Statistical analysis

A case-control association analysis was performed using the Cochran-Armitage trend test. Combined

P-values were obtained using Fisher's combined probability test. Hardy–Weinberg equilibrium was

assessed using the χ^2-test (Nielsen 1998). PI_HAT and MDS analysis were performed using PLINK

1.07 (http://pngu.mgh.harvard.edu/purcell/plink) (Purcell et al.2007). A Manhattan plot of GWAS and

linkage disequilibrium (LD) was drawn using HaploView (Barrett et al. 2005). We categorized the

genotypes as 0, 1, or 2 depending on the number of copies of risk alleles present. Multiple linear

regression analyses were performed to test the independent effect per allele of each SNP on biochemical

traits and histological parameters, accounting for effects of the other variables (i.e., age, gender, and body

mass index [BMI]). BMI, fasting plasma glucose, triglycerides, ferritin, hyaluronic acid, and type IV

collagen 7s values were logarithmically transformed before performing multiple linear regression analysis.

Statistical analyses were performed using R software (http://www.r-project.org/).

Results

Genome-wide case-control association studies

We performed GWAS using NAFLD-1 (n = 392) and control-1 (n = 934). The characteristics of the

study samples are presented in Table 1. After quality controls of genotyping results, 261,540 SNPs in

autosomal chromosomes were used for case-control association analysis. To assess population

stratification, we examined the quantile-quantile P-value plot (Fig. 1A). A slight inflation in P-values

was observed according to the genomic control method ($\lambda_{GC} = 1.09$).   Because we used the JSNP

Database as a control, we were unable to evaluate population stratification between NAFLD-1 and

control-1 subjects.   Instead, we confirmed that all the NAFLD-1 subjects were collected from the

Japanese population by using MDS analysis (Supplementary Fig. 1).

To identify SNPs susceptible to causing NAFLD, we compared NAFLD-1 and control-1

subjects using the trend test.   A Manhattan plot showed that 1 peak located on chromosome 22q13 was

significantly associated with NAFLD and that some SNPs were marginally associated with NAFLD (Fig.

1 B, Supplementary Fig. 3).   To evaluate significant and marginally NAFLD-associated SNPs, we

selected 56 SNPs with P-values less than 5.0×10^{-5}.   We performed a replication study of NAFLD-2 (n

= 172) and control-2 (n = 1012) subjects.   After the replication study, 12 SNPs remained with P-values

less than 1.0×10^{-5}, and 8 SNPs were significantly associated with NAFLD even when conservative

Bonferroni's correction was applied ($P < 1.0 \times 10^{-9}$, Table 2).   All the 8 SNPs were in the same LD

block (Fig. 1 B, Supplementary Fig. 2) and located at chromosome 22q13, which was previously reported

to be NAFLD-susceptible (Romeo et al. 2008; Speliotes et al. 2011; Kawaguchi et al. 2012; Hotta et al.

2010).   NAFLD patients have higher BMI compared with the Japanese general population (Table 1);

thus, we performed multiple logistic regression analysis using genotypes, age, gender, and BMI as

independent variables, involving NAFLD-1, NAFLD-2 and control-2 subjects.   We also genotyped

rs738409 since this SNP was most extensively examined.   After adjusting for age, gender, and BMI, 9

SNPs were strongly associated with NAFLD ($P < 1.0 \times 10^{-9}$, Supplementary Table 1). SNP rs738409

was most strongly associated with NAFLD before adjustment ($P = 2.1 \times 10^{-18}$); however, after adjustment,

8 SNPs (rs2896019, rs3810622, rs738491, rs3761472, rs2143571, rs6006473, rs5764455, and rs6006611)

had also smaller P-values (1.8×10^{-10} to 1.8×10^{-13}) compared to rs738409 ($P = 6.8 \times 10^{-14}$). Moreover,

9 SNPs had high odds ratios (ORs; 1.84 to 2.05).

When 8 SNPs was adjusted with rs738409 using NAFLD-1, NAFLD-2 and control-2 subjects,

no SNPs showed significant association (Supplementary Table 2). P-value of rs738409 was smaller

than other SNPs, suggesting that rs738409 is most important for the development of NAFLD. We also

examined association between 3-SNP (rs738409, rs2896019, and rs3810622) haplotype in the *PNPLA3*,

4-SNP (rs738491, rs3761472, rs2143571, and rs6006473) haplotype in the *SAMM50*, and 2 SNPs

(rs5764455, and rs6006611) haplotype in the *PARVB* genes, and NAFLD, using NAFLD-1, NAFLD-2

and control-2 subjects. Haplotype GGA in the *PNPLA3* ($P = 1.3 \times 10^{-13}$, OR = 2.19), ACAA in the

SAMM50 ($P = 1.3 \times 10^{-11}$, OR = 1.99), and TG in the *PARVB* ($P = 5.0 \times 10^{-12}$, OR = 2.06) genes were

strongly associated with NAFLD (Supplementary Table 3). Haplotype analysis suggested that *PNPLA3*

gene is most important for the pathogenesis for NAFLD.

Analysis of various quantitative and histological phenotypes

Next, we investigated metabolic traits and NAFLD-susceptible SNPs since NAFLD is considered to be a

part of metabolic syndrome (Marchesini et al. 2001; Stefan et al. 2008). Nine SNPs were associated

with decreased serum triglycerides in NAFLD patients, but not in the control group (Table 3).   Allelic

effects on decreased triglycerides levels in NAFLD patients were similar among the 9 SNPs.   These

SNPs were associated with increased AST and ALT both in NAFLD and control subjects.   Allelic

effects of SNPs in the *PNPLA3* gene on increased AST levels in NAFLD patients were higher than those

of other SNPs.   Other metabolic traits were not associated with the 9 SNPs.

The 9 SNPs were associated with lobular inflammation, ballooning, and NAS (Table 4).

Rs3810622 was not associated with lobular inflammation.   Six SNPs (rs738409, rs2896019, rs738491,

rs6006473, rs5764455, and rs6006611) were associated with fibrosis.   Allelic effects on NAS and

fibrosis of SNPs in the *PNPLA3* and *PARVB* gene were stronger than those in the *SAMM50* genes,

although the associations were not significant according to multiple testing.   Nine SNPs in the

chromosome 22q13 region were associated with increased serum ferritin (except rs5764455).   Five

SNPs in the *SAMMM50* and *PARVB* genes (rs738491, rs3761472, rs2143571, rs6006473, and 5764455)

were associated with hyaluronic acid, which were high in NASH.   No SNPs were associated with type

IV collagen 7s.   Nine SNPs showed different association levels with serum metabolic traits and

histological severity suggesting that 3 genes (*PNPLA3*, *SAMM50*, and *PARVB*) may be involved in both

the development and progression of NAFLD (Supplementary Fig. 4).

We have also performed the association tests of 9 SNPs in patients with NASH and simple

steatosis diagnosed by liver biopsy.   Although the number of simple steatosis was small, 9 SNPs were

associated with NASH (OR = 1.76 – 2.79, Supplementary Table 4). Rs5764455 in the *PARVB* gene was

most strongly associated with NASH ($P = 3.4 \times 10^{-6}$, OR = 2.79). Haplotype analysis revealed that most

strongly association with steatosis was haplotype GGA in the *PNPLA3* ($P = 5.0 \times 10^{-4}$) and that haplotype

TG in the *PARVB* gene was most strongly associated with NAS ($P = 9.6 \times 10^{-7}$) and fibrosis ($P = 4.4 \times 10^{-4}$,

Supplementary Table 5).

Discussion

To elucidate the genetic background of NAFLD, we identified the candidate genes (Yoneda et al. 2008;

Yoneda et al. 2009; Yoneda et al. 2009; Hotta et al. 2010). We performed GWAS and found that the

PNPLA3-SAMM50-PARVB genetic region was significantly associated with NAFLD in the Japanese

population. According to our previous study (Hotta et al. 2010), rs738409 in the *PNPLA3* gene was

most strongly associated with NAFLD. NAFLD patients are overweight to obese, and many have

metabolic syndrome (Marchesini et al. 2001; Stefan et al. 2008). Even after adjusting for age, gender,

and BMI, 3 SNPs (rs738409, rs2896019, and rs3810622) in the *PNPLA3*, 4 SNPs (rs738491, rs3761472,

rs2143571, and rs6006473) in the *SAMM50*, and 2 SNPs (rs5764455 and rs6006611) in the *PARVB* genes

showed significant *P*-values. Our previous study indicated that the *P*-value for the association between

rs738409 and NAFLD increased after adjusting for age, gender, and BMI (Hotta et al. 2010). A recently

reported GWAS showed the strongest association of rs738409 in the *PNPLA3* gene with NAFLD in the

Japanese population; however, the study did not adjust for age, gender, or BMI (Kawaguchi et al. 2012).

Although numerous reports and a meta-analysis have indicated that rs738409 is associated with NAFLD

(Sookoian et al. 2011) and that the *PNPLA3* gene is thought to be responsible for the NAFLD, we

demonstrated that *SAMM50*, and *PARVB*, and *PNPLA3* are probably involved in NAFLD development.

NAFLD-susceptible SNPs were also associated with histological severity; however, the effects

differed among the 9 SNPs.   Steatosis grade was equally affected by the 9 SNPs.   Association with

histological activity (NAS) and severity (fibrosis stage) of NAFLD were stronger with SNPs in the

PNPLA3 and *PARVB* genes.   Among biomarkers, AST and ALT, which are commonly used to evaluate

liver function, were highly associated with the 2 SNPs in the *PNPLA3* gene, in NAFLD and in the control

subjects.   Ferritin and hyaluronic acid, the level of which increase in NASH, were associated with SNPs

in the *SAMM50* gene.   SNP in the *PARVB* gene also showed strong association with NASH compared

with simple steatosis.   Haplotype analysis indicated that *PNPLA3* gene would be most important for the

development for NAFLD and that *PARVB* gene would be most important for the progression of NAFLD.

Our data suggested that SNPs in *PNPLA3*, *SAMM50* and *PARVB* contribute to the increased NAFLD

activity, resulting in the progression from simple steatosis to NASH.   It has been suggested that NASH

is induced in 2 consecutive steps (the so-called 2-hit hypothesis): (i) excess fat accumulation in the liver

and (ii) subsequent necroinflammation in the liver (Day and James 1998).   Our results indicate that

SNPs in *PNPLA3* may be involved in the first hit and that SNPs in *PNPLA3*, *SAMM50* and *PARVB* may

be involved in the second hit.   The associations were not significant for multiple tests; therefore, further

analysis is necessary.

We previously reported that rs738409 is associated with decreased serum triglycerides in

NAFLD patients (Hotta et al. 2010).   In this study, we observed that SNPs, particularly in the *SAMM50*

gene, were associated with deceased levels of serum triglycerides.   The association between SNPs in the

PNPLA3 gene and decreased triglycerides levels in NAFLD is controversial (Kollerits et al. 2009;

Speliotes et al. 2010; Speliotes et al. 2011).   Recent reports indicate that rs738409 are associated with

decreased serum triglycerides in type 2 diabetes (Palmer 2012; Krarup et al. 2012).   The controversy

may be due in part to the observation that SNPs in the *SAMM50* gene showed a stronger effect on

triglyceride levels than the SNP in the *PNPLA3* gene.   Further investigation is necessary to elucidate the

association between SNPs in *PNPLA3* and *SAMM50* genes and serum triglycerides levels.

PNPLA3 rs738409 has been extensively investigated, and a strong association with NAFLD

has been confirmed (Day and James 1998).   The *PNPLA3* gene is thought to be involved in abnormal

lipid metabolism in the liver of NAFLD patients.   *PNPLA3*-deficient mice and transgenic mice did not

show a fatty liver (Chen et al.2010; Basantani et al. 2011; Li et al. 2012).   Overexpression of

PNPLA3[I148M] in mouse liver developed to fatty liver, but not into NASH (Li et al. 2012).   Thus,

PNPLA3 plays an important in the development, but not in the progression of NAFLD.   Our study

suggests that the *SAMM50* and *PARVB* genes may also be involved in the progression

(necroinflammation and fibrosis) of NAFLD. Sam50, encoded by the *SAMM50* gene, is a member of

the sorting and assembly machinery for β-barrel proteins in the mitochondrial outer membrane. Sam50

was reported to be involved in the structural integrity of mitochondrial cristae, assembly of respiratory

complexes, and maintenance of mitochondrial DNA. Long-term depletion of Sam50 influences the

amounts of proteins in all the large respiratory complexes in the mitochondria (Ott et al. 2012).

Mitochondrial abnormalities (loss of mitochondrial cristae and paracrystalline inclusions) have been

described for liver biopsy specimens of patients with NASH (Sanyal et al. 2001; Caldwell et al. 1999).

These reports and our results suggest that the *SAMM50* gene may be involved in mitochondrial

dysfunction and subsequent decreased removal of reactive oxygen species (ROS), leading to progression

of NAFLD. The *PARVB* gene encodes parvin-β, which forms integrin-linked kinase-pinch-parvin

complex, transmits signals from integrin to Akt/protein kinase B (PKB) (Kimura et al. 2010). Integrins

are a large family of heterodimeric cell surface receptors that act as mechanoreceptors by relaying

information between cells and from the extracellular matrix (ECM) to the cell interior. Since integrin

receptors directly bind to ECM components to control remodeling, they are thought to play a crucial role

in the evolution and progression of liver fibrosis (Desgrosellier et al. 2010; Patsenker et al. 2011). Loss

of parvin-β contributes to increased integrin-linked kinase activity and cell-matrix adhesion.

Overexpression of parvin-β increases mRNA expression, serine 82 phosphorylation, and activity of

peroxisome proliferator-activated receptor γ (PPARγ), leading to a concomitant increase in lipogenic gene

expression (Johnstone et al. 2008).   Our data and previous reports suggest that the *PARVB* gene is

involved in lipid accumulation and/or fibrosis in the liver, resulting in NAFLD.

In summary, we demonstrated that polymorphisms in the *SAMM50* and *PARVB* genes, as well

as those in the *PNPLA3* gene, were associated with NAFLD development and progression.   SNPs in the

PNPLA3 gene may be involved in the first hit and the *SAMM50* and *PARVB* genes (and *PNPLA3* gene) in

the second hit, although further studies are necessary to confirm our results.

Conflict of interest

The authors declare that they have no conflict of interest.

Figure legend

Fig. 1   Quantile-quantile plot for genome-wide association (A) and regional plots of genome-wide

significant loci (B).

A. The -$\log_{10}$(P-value) of observed association statistics is shown in y-axis, compared with

-log10(P-value)of the association statistics expected under the null hypothesis of no association in x-axis.

B. SNPs are plotted by their position on the chromosome against their association with NAFLD using

GWAS data.   The SNPs surrounding the top SNP (rs2896019) are colored to reflect their LD with the

top SNP (using pairwise r^2 values from GWAS data of NAFLD-1 and control-1).   The positions of genes

as well as the direction of transcription are shown above the plots.

References

Angulo P (2002) Nonalcoholic fatty liver disease. N Engl J Med 18:1221-1231

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265

Basantani MK, Sitnick MT, Cai L, Brenner DS, Gardner NP, Li JZ, Schoiswohl G, Yang K, Kumari M, Gross RW, Zechner R, Kershaw EE (2011) Pnpla3/Adiponutrin deficiency in mice does not contribute to fatty liver disease or metabolic syndrome. J Lipid Res 52:318-329

Brunt EM (2001) Nonalcoholic steatohepatitis: definition and pathology. Semin Liver Dis 21:3-16

Caldwell SH, Swerdlow RH, Khan EM, Iezzoni JC, Hespenheide EE, Parks JK, Parker WD Jr. (1999) Mitochondrial abnormalities in non-alcoholic steatohepatitis. J Hepato l 31:430-434

Chalasani N, Guo X, Loomba R, Goodarzi MO, Haritunians T, Kwon S, Cui J, Taylor KD, Wilson L, Cummings OW, Chen YD, Rotter JI; Nonalcoholic Steatohepatitis Clinical Research Network (2010) Genome-wide association study identifies variants associated with histologic features of nonalcoholic Fatty liver disease. Gastroenterology 139:1567-1576

Chen W, Chang B, Li L, Chan L (2010) Patatin-like phospholipase domain-containing 3/adiponutrin deficiency in mice is not associated with fatty liver disease. Hepatology 52:1134-1142

Day CP, James OF (1998) Steatohepatitis: a tale of two "hits"? Gastroenterology, 114:842-845

Desgrosellier JS, Cheresh DA (2010) Integrins in cancer: biological implications and therapeutic opportunities. Nat Rev Cancer 10:9-22

Farrell GC (2003) Non-alcoholic steatohepatitis: what is it, and why is it important in the Asia-Pacific region? J Gastroenterol Hepatol 18:124-138

Hotta K, Yoneda M, Hyogo H, Ochi H, Mizusawa S, Ueno T, Chayama K, Nakajima A, Nakao K, Sekine A (2010) Association of the rs738409 polymorphism in PNPLA3 with liver damage and the development of nonalcoholic fatty liver disease. BMC Med Genet 11:172

Johnstone CN, Mongroo PS, Rich AS, Schupp M, Bowser MJ, Delemos AS, Tobias JW, Liu Y, Hannigan GE, Rustgi AK (2008) Parvin-beta inhibits breast cancer tumorigenicity and promotes CDK9-mediated peroxisome proliferator-activated receptor gamma 1 phosphorylation. Mol Cell Biol 28:687-704

Kawaguchi T, Sumida Y, Umemura A, Matsuo K, Takahashi M, Takamura T, Yasui K, Saibara T, Hashimoto E, Kawanaka M, Watanabe S, Kawata S, Imai Y, Kokubo M, Shima T, Park H, Tanaka H, Tajima K, Yamada R, Matsuda F; Takeshi Okanoue, Japan Study Group of Nonalcoholic Fatty Liver Disease (2012) Genetic polymorphisms of the human PNPLA3 gene are strongly associated with severity of non-alcoholic fatty liver disease in Japanese. PLoS One 7:e38322

Kimura M, Murakami T, Kizaka-Kondoh S, Itoh M, Yamamoto K, Hojo Y, Takano M, Kario K,

Shimada K, Kobayashi E (2010) Functional molecular imaging of ILK-mediated Akt/PKB

signaling cascades and the associated role of beta-parvin. J Cell Sci 123:747-755

Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, Ferrell LD, Liu

YC, Torbenson MS, Unalp-Arida A, Yeh M, McCullough AJ, Sanyal AJ; Nonalcoholic

Steatohepatitis Clinical Research Network (2005) Design and validation of a histological

scoring system for nonalcoholic fatty liver disease. Hepatology 41:1313-1321

Kollerits B, Coassin S, Beckmann ND, Teumer A, Kiechl S, Döring A, Kavousi M, Hunt SC,

Lamina C, Paulweber B, Kutalik Z, Nauck M, van Duijn CM, Heid IM, Willeit J, Brandstätter

A, Adams TD, Mooser V, Aulchenko YS, Völzke H, Kronenberg F (2009) Genetic evidence for

a role of adiponutrin in the metabolism of apolipoprotein B-containing lipoproteins. Hum Mol

Genet 18:4669-4676

Krarup NT, Grarup N, Banasik K, Friedrichsen M, Færch K, Sandholt CH, Jørgensen T,

Poulsen P, Witte DR, Vaag A, Sørensen T, Pedersen O, Hansen T (2012) The PNPLA3

rs738409 G-allele associates with reduced fasting serum triglyceride and serum cholesterol in

Danes with impaired glucose regulation. PLoS One.7:e40376

Li JZ, Huang Y, Karaman R, Ivanova PT, Brown HA, Roddy T, Castro-Perez J, Cohen JC, Hobbs HH (2012) Chronic overexpression of PNPLA3I148M in mouse liver causes hepatic steatosis. J Clin Invest, in press. doi:pii: 65179. 10.1172/JCI65179

Ludwig J, Viggiano TR, McGill DB, Oh BJ (1980) Nonalcoholic steatohepatitis: Mayo Clinic experiences with a hitherto unnamed disease. Mayo Clin Proc 55:434-438

Marchesini G, Brizi M, Bianchi G, Tomassetti S, Bugianesi E, Lenzi M, McCullough AJ, Natale S, Forlani G, Melchionda N (2001) Nonalcoholic fatty liver disease: a feature of the metabolic syndrome. Diabetes 50:1844-1850

Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ (1999) Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. Gastroenterology 116:1413-1419

Nielsen DM, Ehm MG, Weir BS (1998) Detecting marker-disease association by testing for Hardy–Weinberg disequilibrium at a marker locus. Am J Hum Genet 63:1531-1540

Ohnishi Y, Tanaka T, Ozaki K, Yamada R, Suzuki H, Nakamura Y (2001) A high-throughput SNP typing system for genome-wide association studies. J Hum Genet 46:471-477

Ott C, Ross K, Straub S, Thiede B, Götz M, Goosmann C, Krischke M, Mueller MJ, Krohne G, Rudel T, Kozjak-Pavlovic V (2012) Sam50 functions in mitochondrial intermembrane space bridging and biogenesis of respiratory complexes. Mol Cell Biol 32:1173-1188

Palmer CN, Maglio C, Pirazzi C, Burza MA, Adiels M, Burch L, Donnelly LA, Colhoun H, Doney AS, Dillon JF, Pearson ER, McCarthy M, Hattersley AT, Frayling T, Morris AD, Peltonen M, Svensson PA, Jacobson P, Borén J, Sjöström L, Carlsson LM, Romeo S (2012) Paradoxical lower serum triglyceride levels and higher type 2 diabetes mellitus susceptibility in obese individuals with the PNPLA3 148M variant. PLoS One 7:e39362

Patsenker E, Stickel F (2011) Role of integrins in fibrosing liver diseases. Am J Physiol Gastrointest Liver Physiol 301:G425–G434

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575

Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, Boerwinkle E, Cohen JC, Hobbs HH (2008) Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 40:1461-1465

Sanyal AJ, Campbell-Sargent C, Mirshahi F, Rizzo WB, Contos MJ, Sterling RK, Luketic VA, Shiffman ML, Clore JN (2001) Nonalcoholic steatohepatitis: association of insulin resistance and mitochondrial abnormalities. Gastroenterology 120:1183-1192

Sanyal AJ 2002 American Gastroenterological Association: AGA technical review on nonalcoholic fatty liver disease. Gastroenterology, 123:1705-1725

Sookoian S, Pirola CJ (2011) 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 53:1883-1894

Speliotes EK, Butler JL, Palmer CD, Voight BF, GIANT Consortium; MIGen Consortium;

NASH CRN, Hirschhorn JN (2010) PNPLA3 variants specifically confer increased risk for

histologic nonalcoholic fatty liver disease but not metabolic disease. Hepatology 52:904-912

Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R, Kim LJ, Palmer CD, Gudnason V,

Eiriksdottir G, Garcia ME, Launer LJ, Nalls MA, Clark JM, Mitchell BD, Shuldiner AR, Butler

JL, Tomas M, Hoffmann U, Hwang SJ, Massaro JM, O'Donnell CJ, Sahani DV, Salomaa V,

Schadt EE, Schwartz SM, Siscovick DS; NASH CRN; GIANT Consortium; MAGIC

Investigators, Voight BF, Carr JJ, Feitosa MF, Harris TB, Fox CS, Smith AV, Kao WH,

Hirschhorn JN, Borecki IB; GOLD Consortium (2011) Genome-wide association analysis

identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on

metabolic traits. PLoS Genet 7:e1001324

Stefan N, Kantartzis K, Häring HU (2008) Causes and metabolic consequences of fatty liver.

Endocr Rev 29:939-960

Teli MR, James OF, Burt AD, Bennett MK, Day CP (1995) The natural history of nonalcoholic

fatty liver: a follow-up study. Hepatology 22:1714-1719

Wilfred de Alwis NM, Day CP (2008) Genes and nonalcoholic fatty liver disease. Curr Diab

Rep 8:156-163

Yoneda M, Hotta K, Nozaki Y, Endo H, Uchiyama T, Mawatari H, Iida H, Kato S, Hosono K,

Fujita K, Yoneda K, Takahashi H, Kirikoshi H, Kobayashi N, Inamori M, Abe Y, Kubota K,

Saito S, Maeyama S, Wada K, Nakajima A (2008) Association between PPARGC1A

polymorphisms and the occurrence of nonalcoholic fatty liver disease (NAFLD). BMC

Gastroenterol 8:27-34

Yoneda M, Hotta K, Nozaki Y, Endo H, Uchiyama T, Mawatari H, Iida H, Kato S, Fujita K,

Takahashi H, Kirikoshi H, Kobayashi N, Inamori M, Abe Y, Kubota K, Saito S, Maeyama S,

Wada K, Nakajima A (2009) Association between angiotensin II type 1 receptor polymorphisms

and the occurrence of nonalcoholic fatty liver disease. Liver Int, 29:1078-1085

Yoneda M, Hotta K, Nozaki Y, Endo H, Tomeno W, Watanabe S, Hosono K, Mawatari H, Iida

H, Fujita K, Takahashi H, Kirikoshi H, Kobayashi N, Inamori M, Kubota K, Shimamura T,

Saito S, Maeyama S, Wada K, Nakajima A (2009) Influence of inducible nitric oxide synthase

polymorphisms in Japanese patients with non-alcoholic fatty liver disease. Hepatol Res

39:963-971

Table 1 Clinical characteristic of the subjects

	GWAS		Replication	
	NAFLD-1 (n=392)	Control-1 (n=934)	NAFLD-2 (n=172)	Control-2 (n=1012)
No. of NASH	345	-	97	-
Men/Women	199/193	-	95/77	500/512
Age (year)	49.9 ± 14.8	-	53.5 ± 13.8	53.1 ± 15.3
BMI (kg/m^2)	28.0 ± 5.0	-	27.4 ± 4.6	22.7 ± 3.2
FPG (mg/dL)	118.8 ± 37.3	-	114.8 ± 36.8	98.2 ± 19.0
Hb.A1c (%)	6.4 ± 1.3	-	6.3 ± 1.1	5.5 ± 0.7
T.Chol. (mg/dL)	213.7 ± 41.4	-	205.0 ± 39.6	208.5 ± 36.2
Triglycerides (mg/dL)	172.2 ± 120.6	-	153.3 ± 74.4	110.0 ± 88.5
HDL-C (mg/dL)	52.9 ± 15.7	-	53.8 ± 12.7	62.7 ± 15.5
SBP (mmHg)	127.5 ± 15.0	-	129.6 ± 14.0	124.5 ± 19.1
DBP (mmHg)	78.0 ± 11.7	-	81.1 ± 9.4	76.3 ± 11.6
AST (IU/L)	51.3 ± 31.5	-	47.9 ± 25.4	23.0 ± 10.2
ALT (IU/L)	84.3 ± 60.2	-	75.4 ± 53.6	20.3 ± 11.8
Ferritin (ng/ml)	237.1 ± 225.0	-	229.1 ± 227.3	-
Hyaluronic acid (ng/dL)	44.5 ± 70.2	-	74.8 ± 208.2	-
Type IV collagen 7s (ng/dL)	4.4 ± 1.3	-	6.2 ± 12.8	-
Steatosis grade (1-3)	1.6 ± 0.7	-	1.5 ± 0.8*	-
Lobular inflammation (0-3)	1.2 ± 0.8	-	1.5 ± 0.6*	-
Hepatocyte ballooning (0-2)	1.1 ± 0.7	-	1.2 ± 0.5*	-
NAS (0-8)	4.0 ± 1.7	-	4.2 ± 1.3*	-
Fibrosis stage (0-4)	1.6 ± 1.0	-	2.0 ± 1.0*	-

*, n = 101, AST, aspartate transaminase; ALT, alanine transaminase; DBP, diastolic blood pressure; FPG, fasting plasma glucose; HDL-C, high density lipoprotein cholesterol; SBP, systolic blood pressure; SNP, single-nucleotide polymorphism; T. Chol., total cholesterol.

Table 2　List of the SNPs showing combined $P < 1.0 \times 10^{-5}$

dbSNP ID	Chr	BP (Build36.3)	Nearby genes	Allele 1/2	GWAS Genotype (risk allele frequency) NAFLD-1	Control-1	P-value	OR (95%CI)	Replication Genotype (risk allele frequency) NAFLD-2	Control-2	P-value	OR (95%CI)	Combined P-value
rs6691847	1	30,038,239	PTPRU	T/C*	11/106/275 (0.84)	48/355/531 (0.76)	7.2×10^{-6}	1.63 (1.31 - 2.03)	8/46/116 (0.82)	47/366/599 (0.77)	0.063	1.32 (0.98 - 1.77)	7.3×10^{-6}
rs7552722	1	115,722,878	NGF	A*/G	70/180/142 (0.41)	100/390/444 (0.32)	7.9×10^{-6}	1.49 (1.26 - 1.78)	22/84/66 (0.37)	124/394/490 (0.32)	0.059	1.27 (1.00 - 1.61)	7.2×10^{-6}
rs2051090	13	35,352,193	DCLK1	A/G*	4/108/280 (0.85)	51/328/555 (0.77)	1.6×10^{-6}	1.72 (1.38 - 2.16)	2/52/118 (0.84)	40/299/669 (0.81)	0.27	1.19 (0.88 - 1.62)	7.1×10^{-6}
rs2896019	22	42,665,027	PNPLA3	T/G*	75/155/162 (0.61)	290/453/191 (0.45)	1.3×10^{-13}	1.94 (1.64 - 2.30)	23/83/66 (0.63)	300/509/202 (0.45)	2.5×10^{-9}	2.02 (1.60 - 2.56)	1.6×10^{-20}
rs3810622	22	42,669,467	PNPLA3	A*/G	208/141/43 (0.71)	291/470/172 (0.56)	3.9×10^{-12}	1.90 (1.59 - 2.27)	92/63/17 (0.72)	314/517/180 (0.57)	1.0×10^{-7}	1.95 (1.52 - 2.51)	1.7×10^{-17}
rs738491	22	42,685,444	SAMM50	A*/G	162/170/60 (0.63)	266/448/220 (0.52)	1.2×10^{-6}	1.54 (1.30 - 1.83)	66/87/19 (0.64)	247/506/258 (0.49)	1.0×10^{-6}	1.79 (1.41 - 2.27)	3.9×10^{-11}
rs2073082	22	42,691,340	SAMM50	A/G*	28/142/221 (0.75)	102/419/413 (0.67)	5.0×10^{-5}	1.48 (1.22 - 1.78)	11/66/95 (0.74)	140/457/412 (0.63)	1.0×10^{-4}	1.67 (1.29 - 2.17)	1.0×10^{-7}
rs3761472	22	42,699,455	SAMM50	T/C*	80/186/126 (0.56)	326/430/178 (0.42)	3.0×10^{-10}	1.74 (1.47 - 2.06)	33/86/53 (0.56)	328/515/168 (0.42)	1.5×10^{-6}	1.74 (1.38 - 2.19)	1.5×10^{-14}
rs2143571	22	42,723,019	SAMM50	A*/G	124/185/81 (0.56)	184/427/323 (0.43)	3.9×10^{-9}	1.68 (1.42 - 1.99)	53/86/33 (0.56)	171/518/323 (0.42)	2.9×10^{-6}	1.71 (1.36 - 2.15)	3.5×10^{-13}
rs6006473	22	42,724,408	SAMM50, PARVB	A*/G	163/170/59 (0.63)	274/446/214 (0.53)	3.6×10^{-6}	1.96 (1.65 - 2.32)	67/86/19 (0.64)	257/506/248 (0.50)	3.4×10^{-6}	1.74 (1.38 - 2.21)	3.2×10^{-10}
rs5764455	22	42,729,857	PARVB	T*/C	115/180/97 (0.52)	164/422/348 (0.40)	3.0×10^{-8}	1.63 (1.38 - 1.93)	47/90/35 (0.53)	155/495/362 (0.40)	1.5×10^{-6}	1.74 (1.38 - 2.19)	1.6×10^{-12}
rs6006611	22	42,732,031	PARVB	A/G*	51/167/174 (0.66)	187/471/276 (0.55)	2.4×10^{-7}	1.58 (1.33 - 1.88)	17/72/83 (0.69)	213/502/296 (0.54)	2.3×10^{-7}	1.90 (1.49 - 2.43)	1.8×10^{-12}

*, risk allele.   *P*-values were calculated by Cochran-Armitage trend test.   Odds ratios (ORs) were calculated for risk allele with 95% confidence interval (CI).   Combined *P*-values were obtained using Fisher's combined probability test.

Table 3　Association between significant SNPs and metabolic traits

SNP ID	Subjects	FPG (mg/dL)		T. Chol. (mg/dL)		Triglycerides (mg/dL)		HDL-C (mg/dL)	
		β (s.e.)	P-value	β (s.e.)	P-value	β (s.e.)	P-value	β (s.e.)	P-value
rs738409	Control-2	0.002 (0.003)	0.38	-0.269 (1.579)	0.86	0.007 (0.010)	0.47	0.518 (0.630)	0.41
	NAFLD-1,2	0.001 (0.006)	0.84	-2.732 (2.458)	0.27	-0.049 (0.013)	1.3×10^{-4}	0.463 (0.863)	0.59
rs2896019	Control-2	0.003 (0.003)	0.36	-0.862 (1.575)	0.58	0.011 (0.010)	0.26	0.390 (0.629)	0.54
	NAFLD-1,2	0.002 (0.006)	0.78	-2.750 (2.438)	0.26	-0.044 (0.013)	5.0×10^{-4}	0.448 (0.855)	0.60
rs3810622	Control-2	0.000 (0.003)	0.91	-0.260 (1.597)	0.87	0.011 (0.010)	0.25	0.396 (0.638)	0.53
	NAFLD-1,2	-0.003 (0.007)	0.69	-0.723 (2.621)	0.78	-0.047 (0.014)	5.1×10^{-4}	0.667 (0.915)	0.47
rs738491	Control-2	0.004 (0.003)	0.19	0 .084 (1.555)	0.96	0.010 (0.009)	0.28	0.779 (0.620)	0.21
	NAFLD-1,2	0.005 (0.006)	0.40	-4.251 (2.558)	0.097	-0.041 (0.013)	0.0024	-0.023 (0.895)	0.98
rs3761472	Control-2	0.002 (0.003)	0.44	-0.277 (1.607)	0.86	0.008 (0.010)	0.39	0.618 (0.642)	0.34
	NAFLD-1,2	0.003 (0.006)	0.66	-3.636 (2.491)	0.15	-0.051 (0.013)	9.1×10^{-5}	0.394 (0.874)	0.65
rs2143571	Control-2	0.004 (0.003)	0.19	-0.017 (1.608)	0.99	0.010 (0.010)	0.30	0.548 (0.642)	0.39
	NAFLD-1,2	0.002 (0.006)	0.71	-3.789 (2.494)	0.13	-0.052 (0.013)	6.3×10^{-5}	0.325 (0.876)	0.71
rs6006473	Control-2	0.005 (0.003)	0.055	0.191 (1.554)	0.90	0.012 (0.009)	0.21	0.515 (0.620)	0.41
	NAFLD-1,2	0.005 (0.006)	0.42	-4.341 (2.559)	0.090	-0.042 (0.013)	0.0016	-0.123 (0.896)	0.89
rs5764455	Control-2	0.005 (0.003)	0.077	0.017 (1.610)	0.99	0.013 (0.010)	0.18	1.004 (0.642)	0.12
	NAFLD-1,2	0.002 (0.006)	0.74	-1.593 (2.453)	0.52	-0.042 (0.013)	0.0011	-0.097 (0.859)	0.91
rs6006611	Control-2	0.002 (0.003)	0.41	0.060 (1.561)	0.97	0.012 (0.009)	0.21	0.934 (0.623)	0.13
	NAFLD-1,2	0.000 (0.007)	0.97	-1.181 (2.608)	0.65	-0.039 (0.014)	0.0045	-0.627 (0.906)	0.49

SNP ID	Subjects	SBP (mmHg)		DBP (mmHg)		AST (IU/L)		ALT (IU/L)	
		β (s.e.)	P-value	β (s.e.)	P-value	β (s.e.)	P-value	β (s.e.)	P-value
rs738409	Control-2	0.271 (0.726)	0.71	-0.571 (0.462)	0.22	0.026 (0.007)	2.4×10^{-4}	0.032 (0.009)	4.4×10^{-4}
	NAFLD-1,2	0.383 (1.135)	0.74	-0.101 (0.904)	0.91	0.063 (0.013)	1.2×10^{-6}	0.073 (0.015)	1.1×10^{-6}
rs2896019	Control-2	0.422 (0.724)	0.56	-0.337 (0.462)	0.47	0.025 (0.007)	3.5×10^{-4}	0.032 (0.009)	4.3×10^{-4}
	NAFLD-1,2	0.275 (1.121)	0.81	0.016 (0.893)	0.99	0.063 (0.013)	1.2×10^{-6}	0.071 (0.015)	1.9×10^{-6}
rs3810622	Control-2	-0.134 (0.734)	0.85	-0.534 (0.468)	0.25	0.016 (0.007)	0.022	0.020 (0.009)	0.033
	NAFLD-1,2	0.719 (1.240)	0.56	0.991 (0.987)	0.32	0.053 (0.014)	1.3×10^{-4}	0.050 (0.016)	0.0018
rs738491	Control-2	0.441 (0.714)	0.54	-0.333 (0.455)	0.47	0.021 (0.007)	0.0024	0.027 (0.009)	0.0027
	NAFLD-1,2	-0.006 (1.172)	1.00	-0.111 (0.933)	0.91	0.049 (0.014)	2.9×10^{-4}	0.052 (0.016)	9.5×10^{-4}
rs3761472	Control-2	0.278 (0.739)	0.71	-0.430 (0.471)	0.36	0.022 (0.007)	0.0018	0.030 (0.009)	0.0011
	NAFLD-1,2	-0.223 (1.144)	0.85	-0.079 (0.911)	0.93	0.047 (0.013)	4.2×10^{-4}	0.054 (0.015)	4.2×10^{-4}
rs2143571	Control-2	0.258 (0.739)	0.73	-0.499 (0.471)	0.29	0.022 (0.007)	0.0022	0.030 (0.009)	0.0013
	NAFLD-1,2	-0.095 (1.137)	0.93	-0.133 (0.905)	0.88	0.047 (0.013)	3.6×10^{-4}	0.054 (0.015)	4.4×10^{-4}
rs6006473	Control-2	0.292 (0.714)	0.68	-0.391 (0.455)	0.39	0.020 (0.007)	0.0041	0.028 (0.009)	0.0024
	NAFLD-1,2	0.156 (1.167)	0.89	-0.195 (0.930)	0.83	0.049 (0.014)	3.5×10^{-4}	0.053 (0.016)	6.9×10^{-4}

rs5764455	Control-2	0.213 (0.740)	0.77	-0.500 (0.471)	0.29	0.015 (0.007)	0.034	0.025 (0.009)	0.0070
	NAFLD-1,2	-0.414 (1.122)	0.71	0.059 (0.894)	0.95	0.046 (0.013)	4.0×10^{-4}	0.050 (0.015)	9.7×10^{-4}
rs6006611	Control-2	0.151 (0.718)	0.83	-0.492 (0.457)	0.28	0.018 (0.007)	0.012	0.026 (0.009)	0.0046
	NAFLD-1,2	0.102 (1.217)	0.93	0.701 (0.969)	0.47	0.043 (0.014)	0.0018	0.050 (0.016)	0.0019

Data were derived from linear regression analysis.   NAFLD-1, NAFLD-2 and control-2 were used for analysis.   Values of FPG, triglycerides, AST, and ALT were logarithmically transformed.   Each metabolic phenotype was adjusted for age, gender, and logarithmically transformed BMI.   AST, aspartate transaminase; ALT, alanine transaminase; DBP, diastolic blood pressure; FPG, fasting plasma glucose; HDL-C, high density lipoprotein cholesterol; SBP, systolic blood pressure; SNP, single-nucleotide polymorphism; T. Chol., total cholesterol.
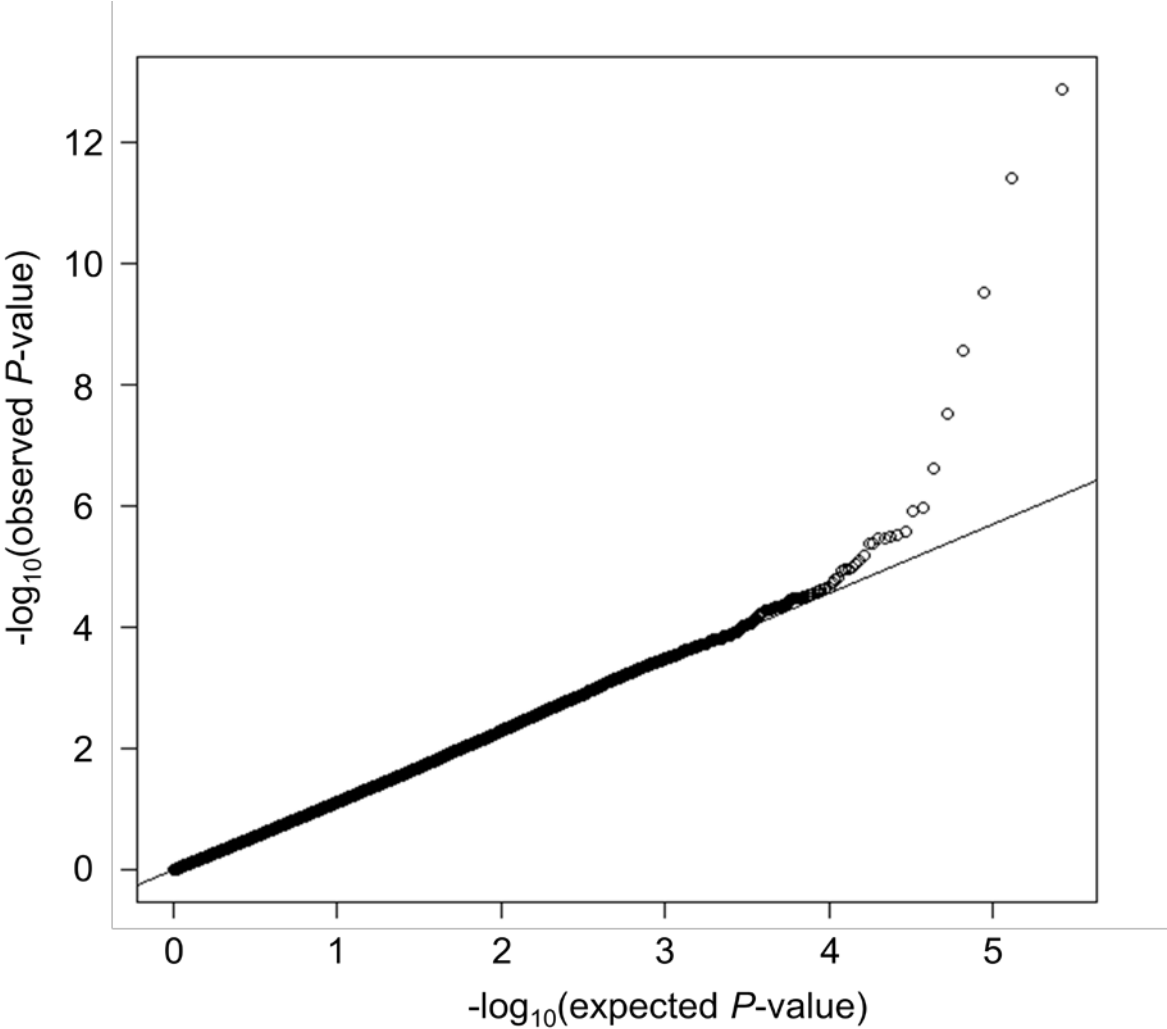
Table 4　Association between SNPs and histological traits and serum biomarker in NAFLD-1 and NAFLD-2 subjects

SNP ID	Steatosis grade*		Lobular inflammation*		Haptocyte ballooning*		NAS*	
	β (s.e.)	*P*-value	β (s.e.)	*P*-value	β (s.e.)	*P*-value	β (s.e.)	*P*-value
rs738409	0.136 (0.043)	0.0016	0.155 (0.044)	5.4×10^{-4}	0.137 (0.043)	0.0014	0.44 (0.097)	7.3×10^{-6}
rs2896019	0.141 (0.042)	9.7×10^{-4}	0.148 (0.044)	8.1×10^{-4}	0.137 (0.042)	0.0013	0.438 (0.096)	6.5×10^{-6}
rs3810622	0.147 (0.046)	0.0014	0.086 (0.048)	0.072	0.141 (0.046)	0.0023	0.395 (0.105)	1.9×10^{-4}
rs738491	0.116 (0.045)	0.0095	0.126 (0.046)	0.0069	0.131 (0.045)	0.0035	0.393 (0.102)	1.3×10^{-4}
rs3761472	0.131 (0.044)	0.0028	0.109 (0.045)	0.017	0.123 (0.044)	0.0052	0.387 (0.100)	1.2×10^{-4}
rs2143571	0.127 (0.044)	0.0039	0.106 (0.046)	0.021	0.120 (0.044)	0.0068	0.377 (0.100)	1.9×10^{-4}
rs6006473	0.103 (0.045)	0.022	0.118 (0.046)	0.011	0.134 (0.045)	0.0028	0.375 (0.102)	2.7×10^{-4}
rs5764455	0.113 (0.043)	0.0091	0.155 (0.044)	5.2×10^{-4}	0.136 (0.043)	0.0016	0.426 (0.098)	1.6×10^{-5}
rs6006611	0.122 (0.045)	0.0074	0.152 (0.047)	0.0012	0.189 (0.045)	2.7×10^{-5}	0.475 (0.102)	4.1×10^{-6}

| SNP ID	Fibrosis*		Ferritin (ng/ml)		Hyaluronic acid (ng/dL)		Type IV collagen 7s (ng/dL)	
	β (s.e.)	*P*-value	β (s.e.)	*P*-value	β (s.e.)	*P*-value	β (s.e.)	*P*-value
rs738409	0.18 (0.062)	0.0036	0.076 (0.025)	0.0028	0.042 (0.023)	0.061	0.003 (0.01)	0.77
rs2896019	0.183 (0.061)	0.0029	0.072 (0.025)	0.0044	0.042 (0.022)	0.062	0.004 (0.01)	0.66
rs3810622	0.110 (0.066)	0.097	0.083 (0.027)	0.0023	0.045 (0.024)	0.064	0.010 (0.011)	0.36
rs738491	0.149 (0.064)	0.021	0.066 (0.026)	0.013	0.079 (0.023)	7.8×10^{-4}	0.006 (0.011)	0.60
rs3761472	0.091 (0.063)	0.15	0.087 (0.026)	8.7×10^{-4}	0.069 (0.023)	0.0028	0.009 (0.010)	0.37
rs2143571	0.101 (0.063)	0.11	0.089 (0.026)	6.5×10^{-4}	0.066 (0.023)	0.0039	0.009 (0.011)	0.37
rs6006473	0.157 (0.064)	0.015	0.071 (0.026)	0.0076	0.072 (0.023)	0.0022	0.005 (0.011)	0.65
rs5764455	0.222 (0.061)	3.3×10^{-4}	0.046 (0.025)	0.068	0.059 (0.022)	0.0083	0.001 (0.010)	0.95
rs6006611	0.195 (0.065)	0.0027	0.057 (0.027)	0.033	0.041 (0.024)	0.089	0.004 (0.011)	0.73

Data were derived from linear regression analysis.　NAFLD-1, NAFLD-2 and control-2 were used for analysis.　Values of ferritin, hyaluronic acid, and type IV collagen 7s were logarithmically transformed.　Each phenotype was adjusted for age, gender, and logarithmically transformed BMI.　*, Histological diagnosed patients from NAFLD-1 (n=392) and NAFLD-2 (n = 101) subjects were used for analysis.

Fig. 1

A

B
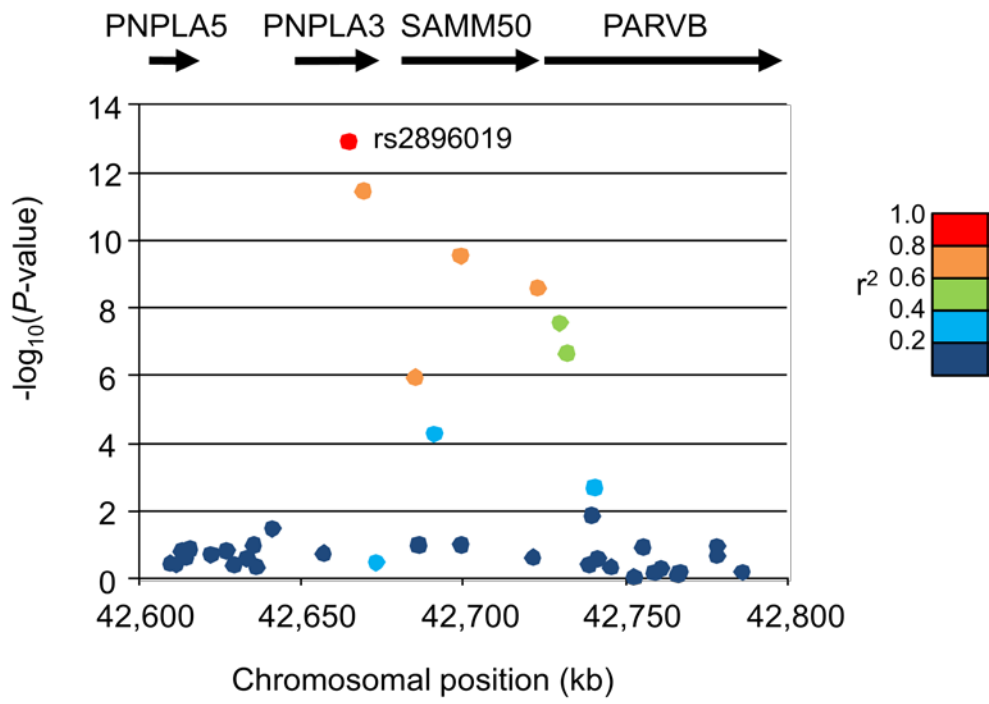
Supplementary Table 1   Association between SNPs and NAFLD before and after adjustment

dbSNP ID	Chr	Nearby genes	Allele 1/2	Genotype (risk allele frequency)		Unadjusted P-value	Adjusted P-value[†]	Adjusted OR (95%CI)[†]
				NAFLD	Control			
rs6691847	1	*PTPRU*	T/C*	19/152/391 (0.83)	47/366/599 (0.77)	$1.0×10^{-4}$	0.0041	1.39 (1.11 - 1.75)
rs7552722	1	*NGF*	A*/G	92/264/208 (0.40)	124/394/490 (0.32)	$1.8×10^{-5}$	$4.1×10^{-4}$	1.39 (1.16 - 1.66)
rs2051090	13	*DCLK1*	A/G*	6/160/398 (0.85)	40/299/669 (0.81)	0.012	0.13	1.20 (0.94 - 1.54)
rs738409	22	*PNPLA3*	C/G*	96/241/227 (0.62)	300/513/199 (0.45)	$2.1×10^{-18}$	$6.8×10^{-14}$	2.05 (1.70 - 2.47)
rs2896019	22	*PNPLA3*	T/G*	98/238/228 (0.62)	300/509/202 (0.45)	$8.3×10^{-18}$	$1.8×10^{-13}$	2.02 (1.67 - 2.43)
rs3810622	22	*PNPLA3*	A*/G	300/204/60 (0.71)	314/517/180 (0.57)	$1.3×10^{-15}$	$1.7×10^{-11}$	1.95 (1.60 - 2.36)
rs738491	22	*SAMM50*	A*/G	228/257/79 (0.63)	247/506/258 (0.49)	$2.1×10^{-13}$	$2.3×10^{-11}$	1.90 (1.57 - 2.29)
rs2073082	22	*SAMM50*	A/G*	39/208/316 (0.75)	140/457/412 (0.63)	$3.6×10^{-10}$	$1.5×10^{-8}$	1.78 (1.46 - 2.17)
rs3761472	22	*SAMM50*	T/C*	113/272/179 (0.56)	328/515/168 (0.42)	$1.1×10^{-13}$	$2.3×10^{-11}$	1.91 (1.58 - 2.31)
rs2143571	22	*SAMM50*	A*/G	177/271/114 (0.56)	171/518/323 (0.42)	$1.4×10^{-12}$	$1.8×10^{-10}$	1.85 (1.53 - 2.24)
rs6006473	22	*SAMM50* *PARVB*	A*/G	230/256/78 (0.63)	257/506/248 (0.50)	$3.1×10^{-12}$	$1.2×10^{-10}$	1.85 (1.53 - 2.23)
rs5764455	22	*PARVB*	T*/C	162/270/132 (0.53)	155/495/362 (0.40)	$4.4×10^{-12}$	$1.4×10^{-10}$	1.84 (1.53 - 2.21)
rs6006611	22	*PARVB*	A/G*	68/239/257 (0.67)	213/502/296 (0.54)	$9.7×10^{-12}$	$4.2×10^{-11}$	1.89 (1.56 - 2.28)

*, risk allele.   †, P-value and odds ratios (ORs) adjusted for age, gender, and logarithmically transformed BMI were calculated using multiple logistic regression analysis.   NAFLD-1, NAFLD-2, and control-2 subjects were used for analysis.

Supplementary Table 2   Association between SNPs and NAFLD adjusted with rs738409

dbSNP ID	Nearby genes	Explanatory variable			
		rs738409		Each dbSNP	
		P-value	OR (95%CI)	*P*-value	OR (95%CI)
rs2896019	*PNPLA3*	0.13	1.75 (0.85 - 3.61)	0.66	1.18 (0.57 - 2.42)
rs3810622	*PNPLA3*	1.9×10^{-4}	1.71 (1.29 - 2.26)	0.090	1.29 (0.96 - 1.72)
rs738491	*SAMM50*	2.2×10^{-4}	1.77 (1.31 - 2.40)	0.23	1.21 (0.89 - 1.63)
rs3761472	*SAMM50*	3.1×10^{-4}	1.89 (1.34 - 2.68)	0.60	1.10 (0.77 - 1.56)
rs2143571	*SAMM50*	4.0×10^{-5}	2.00 (1.44 - 2.79)	0.87	1.03 (0.74 - 1.44)
rs6006473	*SAMM50, PARVB*	5.5×10^{-5}	1.84 (1.37 - 2.47)	0.36	1.15 (0.85 - 1.55)
rs5764455	*PARVB*	1.8×10^{-5}	1.79 (1.37 - 2.34)	0.17	1.21 (0.93 - 1.57)
rs6006611	*PARVB*	2.7×10^{-5}	1.71 (1.33 - 2.19)	0.034	1.32 (1.02 - 1.70)

P-value and odds ratios (ORs) adjusted for age, gender, logarithmically transformed BMI, and rs738409 were calculated using multiple logistic regression analysis.   NAFLD-1, NAFLD-2, and control-2 were used for analysis.

Supplementary Table 3　Association between haplotypes and NAFLD

Gene	SNPs	Haplotype	Frequency		Unadjusted	Adjusted	Adjusted OR
			NAFLD	Control	P-value	P-value*	(95% CI)*
PNPLA3		GGA	0.603	0.440	8.1×10^{-18}	1.3×10^{-13}	2.19 (1.79 - 2.7)
	rs738409						
	rs2896019	CTG	0.275	0.423	4.2×10^{-16}	Reference	Reference
	rs3810622						
		CTA	0.100	0.118	0.15	0.095	1.32 (0.95 - 1.82)
SAMM50	rs738491	ACAA	0.551	0.413	1.3×10^{-13}	1.3×10^{-11}	1.99 (1.63 - 2.43)
	rs3761472						
	rs2143571	GTGG	0.363	0.496	1.6×10^{-12}	Reference	Reference
	rs6006473	ATGA	0.078	0.077	0.95	0.037	1.45 (1.02 - 2.06)
PARVB		TG	0.521	0.394	9.2×10^{-12}	5.0×10^{-12}	2.06 (1.68 - 2.52)
	rs5764455	CA	0.328	0.455	8.5×10^{-12}	Reference	Reference
	rs6006611						
		CG	0.146	0.147	0.94	0.0042	1.49 (1.14 - 1.97)

NAFLD-1, NAFLD-2, and control-2 were used for analysis.　The program Haplo.Stats
(http://mayoresearch.mayo.edu/mayo/research/schaid_lab/software.cfm) was used to explore associations of *PNPLA3*, *SAMM50* and *PARVB* haplotypes with NAFLD.　*, *P*-value and odds ratios (ORs) adjusted for age, gender, and logarithmically transformed BMI were calculated using multiple logistic regression analysis.

Supplementary Table 4 Association tests of SNPs in patients with NASH and simple steatosis

dbSNP ID	Chr	Nearby genes	Allele 1/2	Genotype (risk allele frequency)		P-value[†]	OR (95%CI)[†]
				NASH (n=442)	Steatosis (N=51)		
rs738409	22	*PNPLA3*	C/G*	72/183/187 (0.63)	17/24/10 (0.43)	1.8×10^{-4}	2.21 (1.46 - 3.35)
rs2896019	22	*PNPLA3*	T/G*	74/180/188 (0.63)	17/23/11 (0.44)	3.7×10^{-4}	2.10 (1.40 - 3.17)
rs3810622	22	*PNPLA3*	A*/G	237/158/47 (0.71)	18/24/9 (0.59)	0.0079	1.76 (1.16 - 2.67)
rs738491	22	*SAMM50*	A*/G	185/199/58 (0.64)	14/22/15 (0.49)	0.0014	1.99 (1.30 - 3.03)
rs2073082	22	*SAMM50*	A/G*	87/209/146 (0.57)	16/26/9 (0.43)	0.0075	1.78 (1.17 - 2.72)
rs3761472	22	*SAMM50*	T/C*	87/210/143 (0.56)	17/25/9 (0.42)	0.0057	1.83 (1.19 - 2.8)
rs2143571	22	*SAMM50*	A*/G	186/199/57 (0.65)	14/22/15 (0.49)	0.0013	2.01 (1.31 - 3.06)
rs6006473	22	*SAMM50* *PARVB*	A*/G	136/208/98 (0.54)	7/22/22 (0.35)	2.0×10^{-4}	2.28 (1.48 - 3.53)
rs5764455	22	*PARVB*	T*/C	211/181/50 (0.68)	11/25/15 (0.46)	3.4×10^{-6}	2.79 (1.81 - 4.3)
rs6006611	22	*PARVB*	A/G*	72/183/187 (0.63)	17/24/10 (0.43)	1.8×10^{-4}	2.21 (1.46 - 3.35)

*, risk allele. †, *P*-value and odds ratios (ORs) adjusted for age, gender, and logarithmically transformed BMI were calculated using multiple logistic regression analysis. Histological diagnosed patients from NAFLD-1 (345 NASH and 47 simple steatosis) and NAFLD-2 (97 NASH and 4 simple steatosis) subjects were used for analysis.

Supplementary Table 5   Association between haplotypes and histological traits

Gene	SNPs	Haplotype	Steatosis		NAS		Fibrosis	
			β (s.e.)	*P*-value*	β (s.e.)	*P*-value	β (s.e.)	*P*-value
PNPLA3	rs738409	GGA	0.167 (0.048)	5.0×10^{-4}	0.467 (0.108)	2.0×10^{-5}	0.164 (0.069)	0.017
	rs2896019	CTG	Reference	Reference	Reference	Reference	Reference	Reference
	rs3810622	CTA	0.102 (0.081)	0.21	0.097 (0.183)	0.60	-0.061 (0.116)	0.60
SAMM50	rs738491	ACAA	0.132 (0.046)	0.0045	0.422 (0.105)	7.1×10^{-5}	0.142 (0.066)	0.033
	rs3761472 rs2143571	GTGG	Reference	Reference	Reference	Reference	Reference	Reference
	rs6006473	ATGA	-0.008 (0.089)	0.93	0.213 (0.202)	0.29	0.262 (0.128)	0.041
PARVB		TG	0.137 (0.047)	0.0039	0.527 (0.106)	9.6×10^{-7}	0.239 (0.067)	4.4×10^{-4}
	rs5764455 rs6006611	CA	Reference	Reference	Reference	Reference	Reference	Reference
		CG	0.102 (0.070)	0.14	0.418 (0.156)	0.0078	0.083 (0.100)	0.41

Histological diagnosed patients from NAFLD-1 (345 NASH and 47 simple steatosis) and NAFLD-2 (97 NASH and 4 simple steatosis) subjects were used for analysis.   The program Haplo.Stats (http://mayoresearch.mayo.edu/mayo/research/schaid_lab/software.cfm) was used to explore associations of *PNPLA3*, *SAMM50* and *PARVB* haplotypes with NAFLD.   *, *P*-value and β adjusted for age, gender, and logarithmically transformed BMI were calculated using multiple regression analysis.
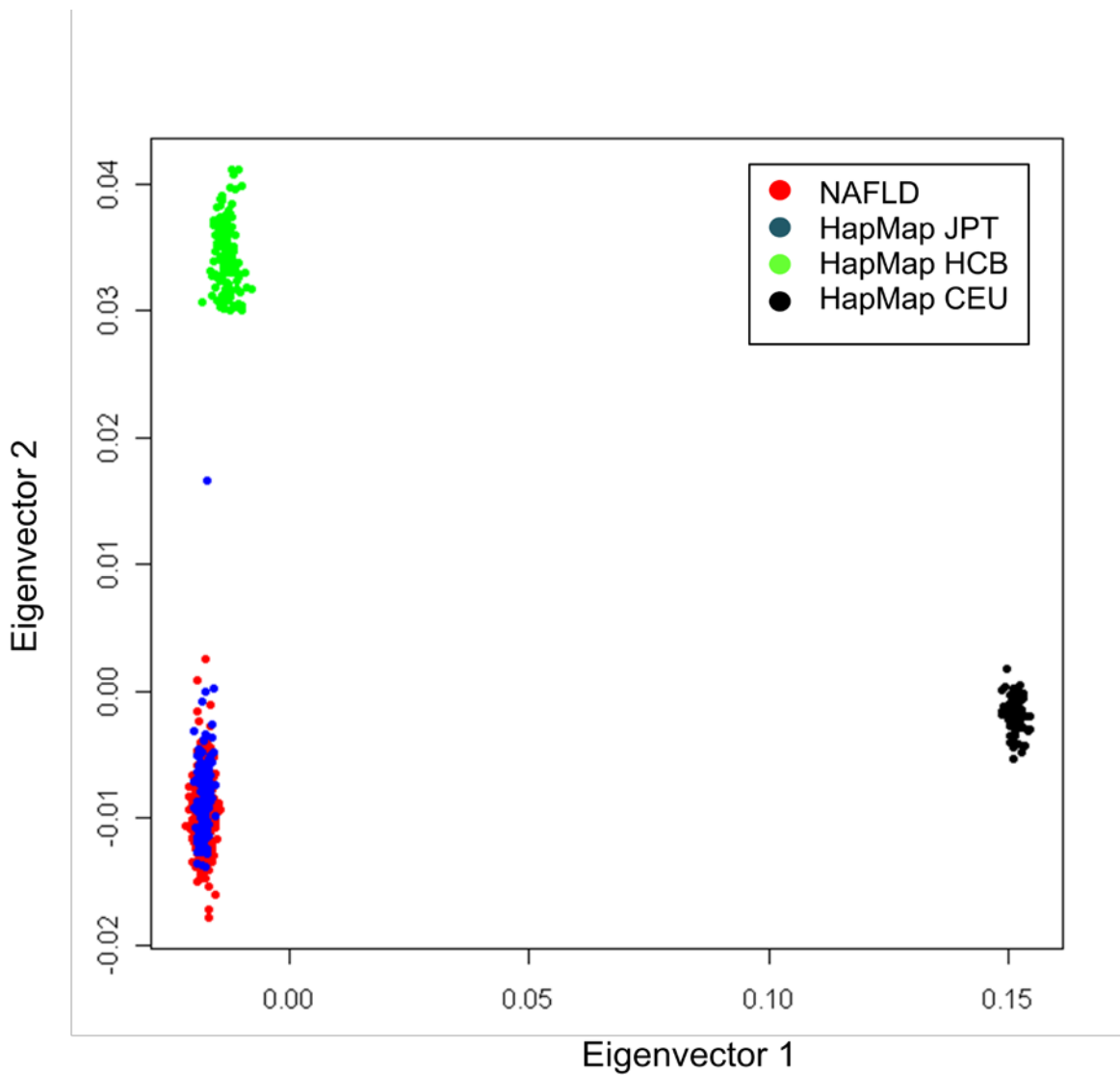
Supplementary Fig. 1   MDS plot for NAFLD, HapMap JPT, HapMap HCB and HapMap CEU
Relatedness between the NAFLD-1 (n = 392) patients as well as 71 European (CEU) and 250 East-Asian (113 JPT and 137 CHB) individuals from the phase II and III HapMap project was analyzed.   Genotype data for 256,869 SNPs were analyzed using PLINK 1.07 (http://pngu.mgh.harvard.edu/purcell/plink) (Purcell et al.2007). The individuals were plotted in a 2-dimensional graph, with the first (x-axis) and the second (y-axis) components of the Eigenvector factors.
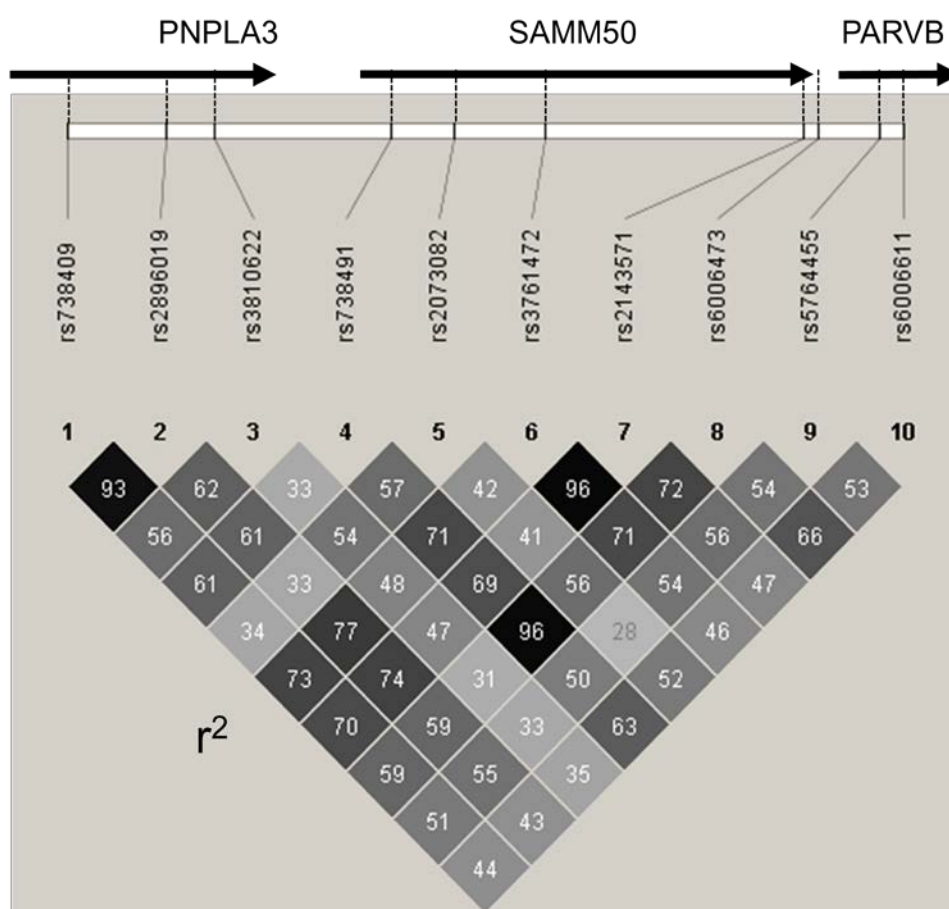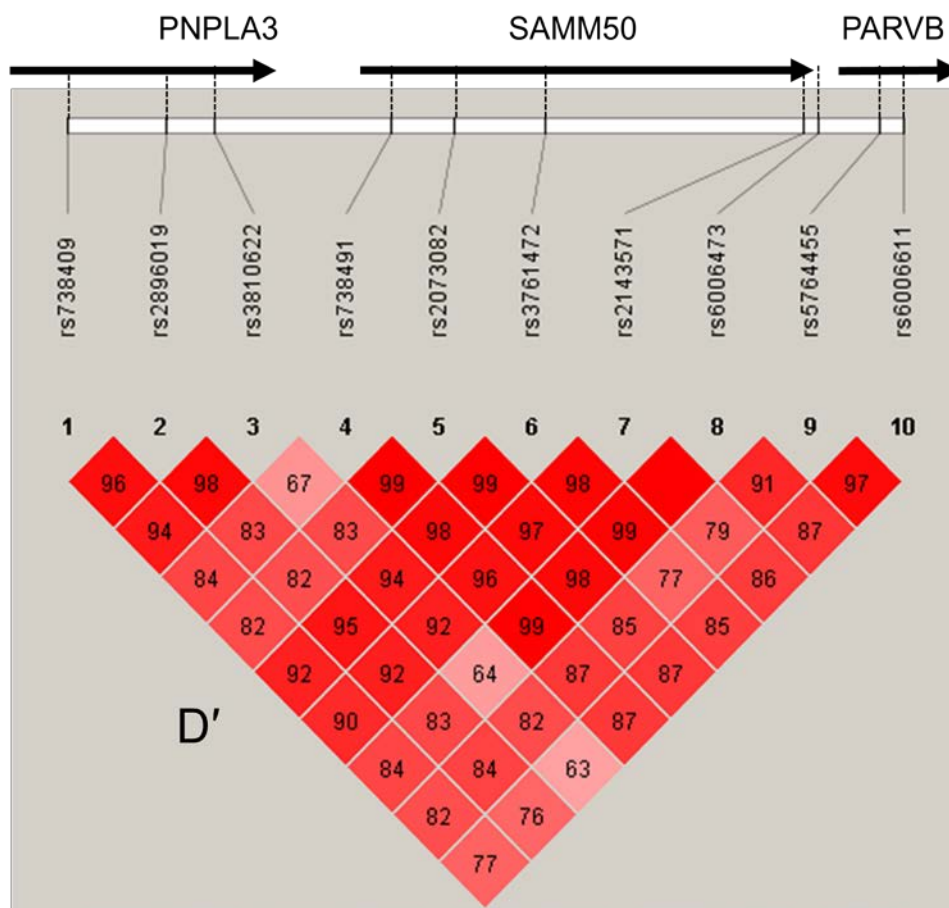
Supplementary Fig. 2 LD coefficients (D′ and r²) in the *PNPLA3*, *SAMM50* and *PARVB* genes
LD coefficients (D' and r²) were calculated for every pair of SNPs using HaploView (Barrett et al. 2005).
NAFLD-1, NAFLD-2, and control-2 were used for analysis.
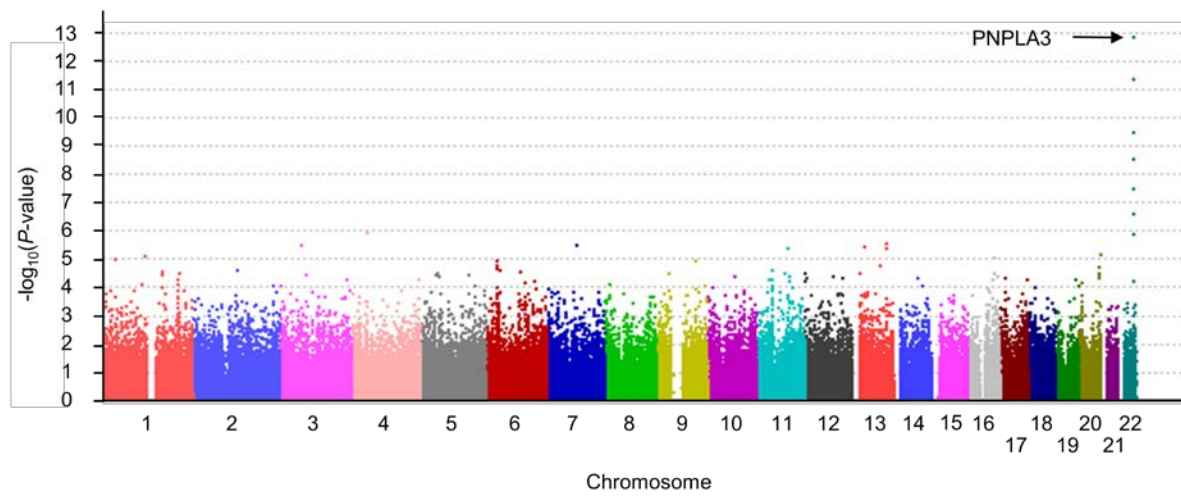
Supplementary Fig. 3   Manhattan plots of association of SNPs with NAFLD in the GWAS.
Association *P*-values are calculated by trend test and plotted along the chromosome in –log$_{10}$ scale.

A

B
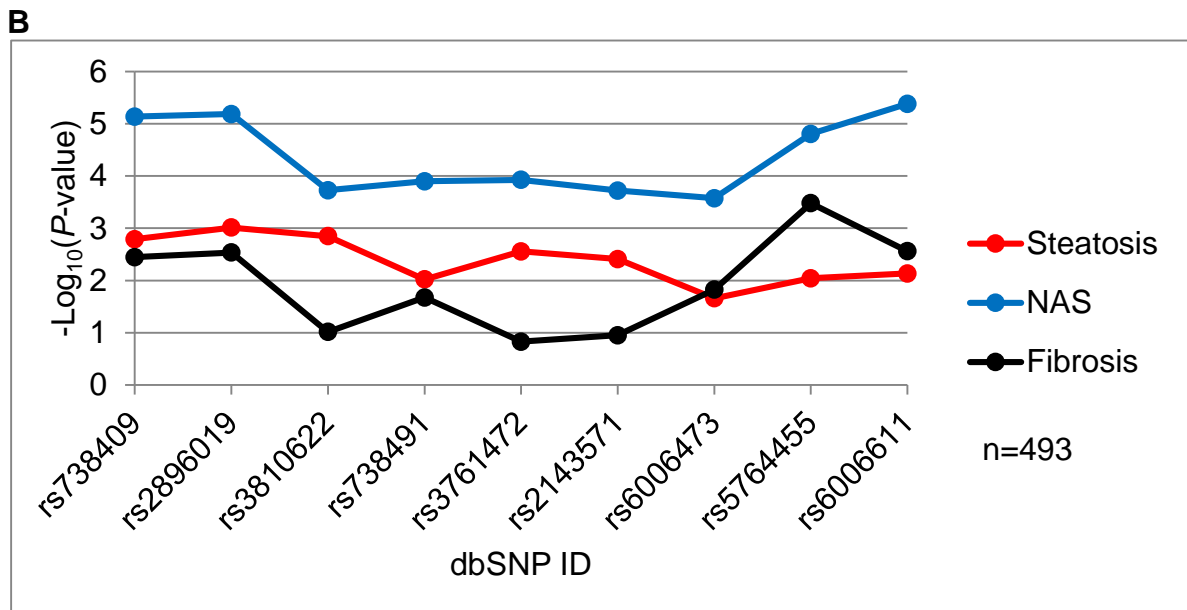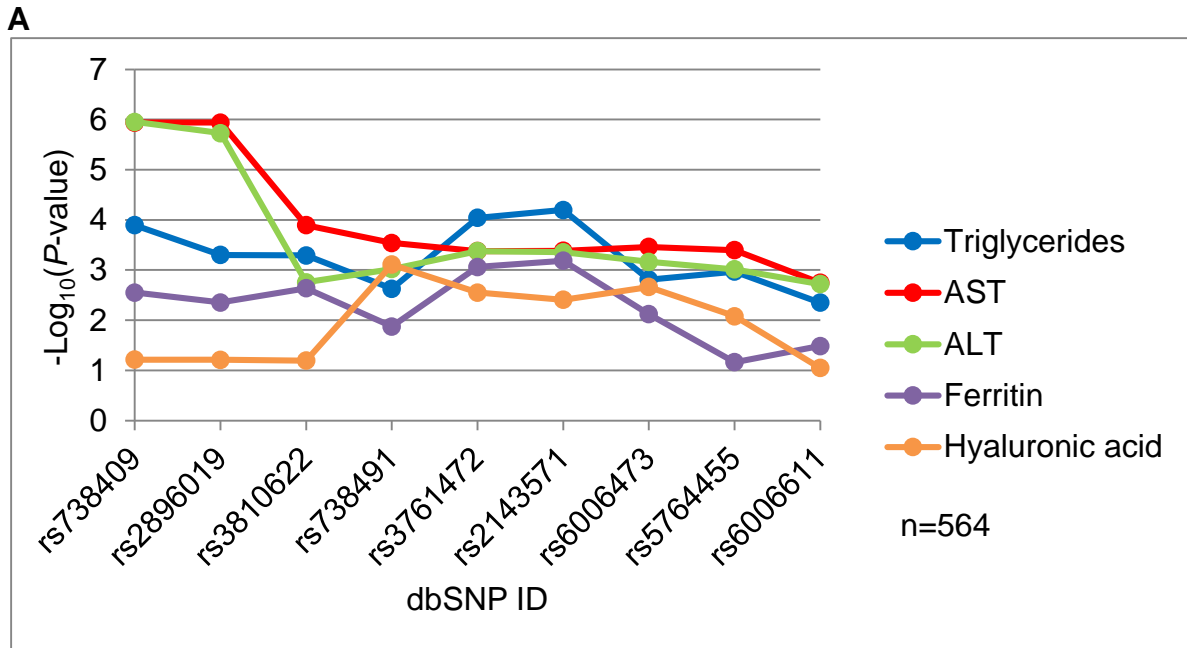
Supplementary Fig. 4 ***P*-values obtained using multiple regression analysis for the 7 SNPs**
Multiple linear regression analyses were performed to test the independent effect per allele of each SNP on biochemical traits (A) and histological parameters (B) by accounting for the effects of the other variables (i.e., age, gender, and BMI). The BMI, triglycerides, ferritin, and hyaluronic acid values were logarithmically transformed before performing multiple linear regression analysis. NAFLD-1 and NAFLD-2 were used for analysis.