Ph.D. Dissertation of Medicine

Genetic and metabolomic study of pediatric nonalcoholic fatty liver disease

Genetic variants of *PNPLA3, TM6SF2, SAMM50* and metabolomic profiles

소아 비알콜성 지방간질환의 유전자와 대사체 연구

August 2021

Graduate School of Medicine
Seoul National University
Pediatrics Major

Kyung Jae Lee

Genetic and metabolomic study of pediatric nonalcoholic fatty liver disease

- Genetic variants of *PNPLA3, TM6SF2, SAMM50* and metabolomic profiles-

지도교수 고 재 성

이 논문을 의학박사 학위논문으로 제출함

2021년 **4**월

서울대학교 대학원

의학과 소아청소년과

이 경 재

이경재의 박사 학위논문을 인준함

2021년 **7**월

위 원 장 _____

부위원장 _____

위 원 _____

위 원 _____

위 원 _____

Abstract

Background

Some genetic variants and different metabolomic profiles have been reported to be associated with nonalcoholic fatty liver disease (NAFLD). Few studies have reported various genetic variants and associated metabolites simultaneously, and no study has been conducted in pediatric populations.

Objective

The aim of this study was to investigate the effects of genetic variants on pediatric NAFLD and analyze metabolic differences between NAFLD patients and controls in a pediatric population. In addition, other risk factors for pediatric NAFLD were aimed to investigate.

Methods

NAFLD was defined if hepatic steatosis was shown on ultrasound. A total of 228 NAFLD patients (body mass index-Z [BMI-Z] = 2.51 ± 1.01) and 225 controls (BMI-Z = 0.22 ± 1.48) were included. All participants underwent examination by anthropometry and blood cell count and liver function analysis. Four variants of *PNPLA3* (rs738409), *TM6SF2* (rs58542926) and *SAMM50* (rs2073080, rs3761472) were genotyped by TaqMan allelic discrimination assay. Metabolic profiles were checked in children

with overweight. The pediatric NAFLD fibrosis score (PNFS), the AST/platelet ratio index (APRI) and fibrosis-4 (FIB-4) were used to evaluate the degree of hepatic fibrosis. Genetic risk factors for NAFLD in all participants were analyzed by adjusting for age, sex and BMI-Z. Subgroup analysis was conducted in children with overweight with more metabolic adjustments. The genetic risk score was calculated to evaluate the synergetic effects of 4 genetic variants.

Among them, 166 (105 NAFLD and 61 control) children were enrolled for metabolomic analysis. The plasma metabolome was quantified using a Biocrates AbsoluteIDQ p400 kit and Thermo Q Exactive Plus Orbitrap mass spectrometer.

Results

The four genetic variants (rs738409, rs58542926, rs2073080 and rs3761472), male sex and BMI-Z independently increased susceptibility to NAFLD. These variants remained significant risk factors with higher odds ratio in children with overweight in addition to fasting insulin and triglyceride. These variants increased the alanine aminotransferase level, PNFS, APRI, and FIB-4 independently. As the genetic risk score increased, aspartate transaminase, alanine aminotransferase, PNFS, APRI, and FIB-4

increased independently suggesting synergetic effects of these 4 variants.

NAFLD patients showed a higher plasma levels of branched chain amino acids (BCAAs, leucine, isoleucine, valine), tyrosine, phosphatidylcholines (PCs), sphingomyelins (SMs), diglyceride, triglycerides (TGs) than control. Some of these metabolites including BAAA, glutamate, PCs, SMs had positive association with homeostasis model assessment-estimated insulin resistance (HOMA-IR). Plasma levels of glutamine, glycine and serin were lower in NAFLD patients than control. Glutamine and glycine showed negative correlation with HOMA-IR. The carries of *TM6SF2* variants significantly showed lower plasma PCs, SMs and TG compared to wild type and the distribution of metabolites was reversed to the NAFLD.

Conclusion

The effects of genetic variants and metabolomic profiles in children with NAFLD was first demonstrated in this study. Genetic variants of *PNPLA3*, *TM6SF2* and *SAMM50* are associated with the development and severity of pediatric NAFLD and their effects are greater in children with overweight than normal weight. These variants have synergetic effects on severity of pediatric NAFLD. A

total of 49 metabolites showed significant differences between subjects with NAFLD and control, are associated with insulin resistance. While variants of *TM6SF2* results in lower plasma lipids, other variants did not show significant differences in metabolome.

Keywords: Nonalcoholic fatty liver disease; *Patatin-like phospholipase domain-containing 3; Transmembrane 6 superfamily member 2; Samm50*; Genetics; Metabolomics

Student Number: 2017-34184

Table of Contents

Chapter 1. Introduction

1.1. Study Background

Nonalcoholic fatty liver disease (NAFLD) is a common chronic liver disease that affects a large proportion of adults and children worldwide [1–3]. In Korea, the population of overweight pediatric patients significantly increased from 20.3% in 2010 to 25.5% in 2015, and the prevalence of pediatric NAFLD increased from 4.7% in 2010 to 5.9% in 2015 [4–6].

Aside from simple obesity, male sex, ethnicity, insulin resistance, visceral obesity, dietary factors and several genetic factors are associated with the development and severity of NAFLD and its natural course [7–14]. A previous pediatric study reported that the prevalence of NAFLD diagnosed with autopsy was higher in Hispanic and Asian children [15]. Another familial aggregation study showed that children without NAFLD showed a lower prevalence of NAFLD in their parents or siblings than children with NAFLD [16]. These studies suggested heritability of NAFLD. Recently, genome-wide association studies (GWAS) have shown that a large number of genetic variants are associated with susceptibility to and the severity of NAFLD [17–19]. Patatin-like phospholipase-containing domain 3 (*PNPLA3*) rs738409 and transmembrane 6 superfamily

member 2 (*TM6SF2*) rs58542926 variants are well known to be associated with pediatric NAFLD [20,21]. In cohorts of Caucasian, African American, and Hispanic obese children, *PNPLA3* rs768409 variants were found to increase hepatic fat content by magnetic resonance imaging without increasing insulin resistance [22], and the rs58542926 variant of *TM6SF2* was found to be associated with increased hepatic fat fraction, alanine aminotransferase (ALT) level, fibrosis stage and NAFLD activity score, although it showed protective effects against cardiovascular risk. A recent large-scale GWAS of Asian adults reported that rs3761472, rs2073080 and rs2143571 variants of *SAMM50* also affect the incidence and severity of NAFLD [18,19,23].

There have been no pediatric studies of the association between *SAMM50* and NAFLD. There was no genetic study of pediatric NAFLD in Korea. In addition, pediatric studies have been conducted only in children with obesity, and there have been no studies including children with normal weight.

In metabolomic studies, the levels of some branched-chain amino acids (BCAAs), such as valine, isoleucine, tryptophan, and lysine, were significantly increased in NAFLD children compared to obese controls and were negatively associated with insulin sensitivity [24]. In an adult study, several bile acids, free fatty acids, BCAAs,

glucose and pyruvate were significantly different, and these biomarkers could effectively separate healthy controls from NAFLD patients and healthy controls for nonalcoholic steatohepatitis (NASH) [25]. From these "omic" studies, people are expecting to understand the more precise mechanism of NAFLD and finally predict the individual course of the disease to be able to respond to therapeutic intervention.

To our knowledge, very few studies have conducted simultaneous analysis of genetic variants and related metabolomes.

1.2. Purpose of Research

The purpose of this study was to evaluate the association of *PNPLA3* rs738409, *TM6SF2* rs58542926 and *SAMM50* rs2073080 and rs3761472 with the development and severity of NAFLD in children. Another aim of this study was to investigate other risk factors in addition to genetic variants for the development of pediatric NAFLD. In addition, the metabolic profiles according to genetic variants and pediatric NAFLD were evaluated.

Chapter 2. Method

2.1. Study population

From January 2005 to April 2020, a total of 453 participants who visited the pediatric department at Seoul National University Children's Hospital and Hallym University Sacred Heart Hospital were enrolled in this study (Figure 1).

Patients were defined as NAFLD patients if they showed hepatic steatosis on abdominal ultrasound and controls if they did not. Controls were chosen among patients who visited clinic and received abdominal ultrasound for their diagnostic evaluation. Participants were further divided into two subgroups according to their body mass index Z (BMI-Z) score: normal weight (BMI-Z $\leq$ 1) and overweight (BMI >1) [26] (Figure 1). The BMI-Z score was calculated based on the 2017 Korean national growth chart for children and adolescents [27]. Finally, 228 NAFLD patients and 225 controls were recruited: 215 NAFLD patients with overweight (47.5%) and 13 NAFLD patients with normal weight (2.9%), 71 controls with overweight (15.7%) and 154 controls with normal weight (34%) for genetic analysis.

Metabolomic analysis in 166 children from 453 participants was

performed. These participants were recruited from January 2019 to May 2020. A total of 105 NAFLD patients and 61 controls were enrolled for metabolic analysis; 96 NAFLD patients were overweight (57.8%) 9 NAFLD patients were normal weight (5.4%), 22 controls were overweight (13.3%) and 39 controls were normal weight (23.5%).

Patients who took medication known to affect liver function test results or who consumed alcohol were excluded. Children with viral hepatitis, such as hepatitis A, hepatitis B, or hepatitis C, and children with Epstein-Barr virus, Wilson's disease, autoimmune hepatitis, and muscular disease were excluded. The Institutional Review Board (IRB No. 1811-149-98, 2018-10-015) of each hospital approved this study, and informed consent from parents and children was obtained.

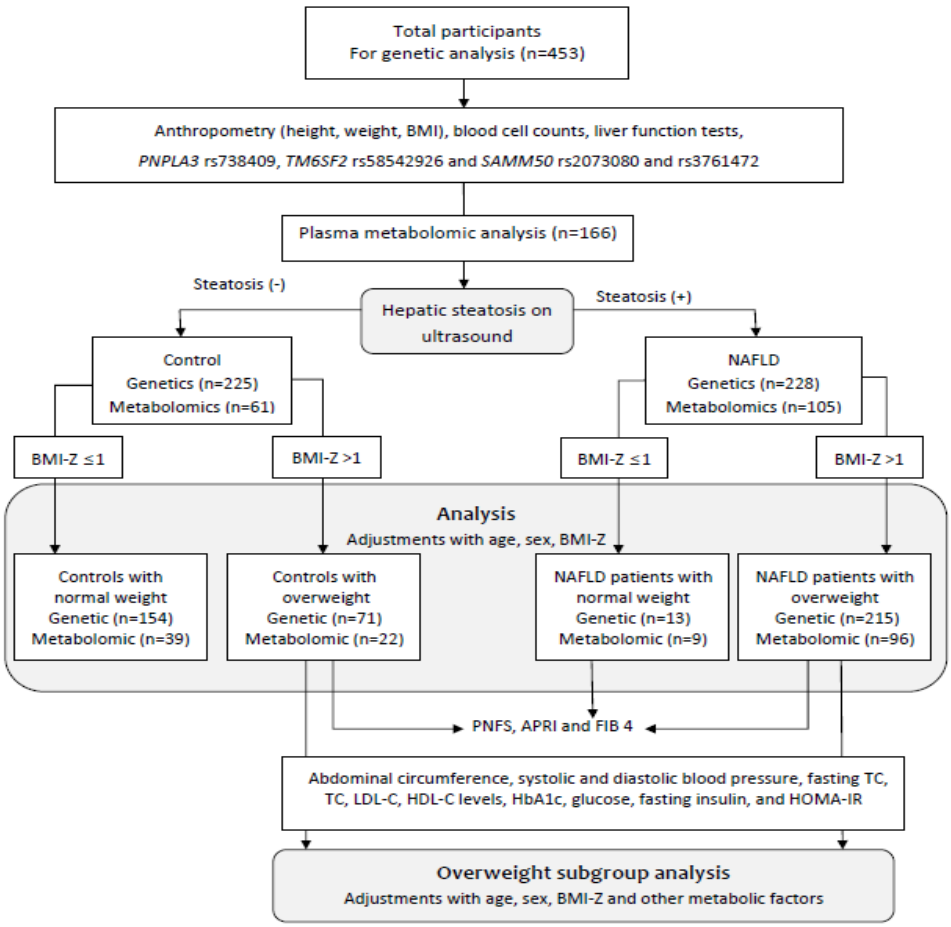

Figure 1. Study flow diagram of genetic analysis

Abbreviations: *PNPLA3, phospholipase-containing domain 3; TM6SF2, transmembrane 6 superfamily member 2*; BMI-Z score, body mass index Z score; NAFLD nonalcoholic fatty liver disease; PNFS, pediatric NAFLD fibrosis score; APRI, AST/platelet ratio index; FIB-4, fibrosis-4; TC, total cholesterol; TG, triglyceride; LDL-C, low-density lipoprotein cholesterol; HDL-C high-density

lipoprotein cholesterol, HOMA-IR, homeostatic model assessment of insulin resistance

2.2. Clinical and laboratory assessments

All participants underwent evaluation by anthropometry (height, weight and BMI-Z score), blood cell count and liver function analysis, including determination of the ALT, aspartate transaminase (AST), alkaline phosphatase (ALP), and gamma glutamyl transferase (GGT) levels, and abdominal ultrasonography.

Diagnosis of fatty liver and determination of the grade of steatosis were defined as increases in the echogenicity of the liver parenchyma compared to that of the right kidney cortex. The grading system was as follows: grade 0, normal hepatic structure with the same echogenicity as the kidney; grade 1, mild steatosis: slightly increased echogenicity compared to the kidney, clear visualization of the diaphragm and interface of hepatic veins, and sharp contours; grade 2, moderate steatosis: diffuse increased echogenicity with slightly impaired visualization of deeper parts of the liver, the diaphragm and hepatic veins and blunted contours; grade 3, severe steatosis: markedly increased echogenicity of the liver with poor or no visualization of the diaphragm or hepatic veins [28,29].

To validate the accuracy of hepatic ultrasound in this study, the

controlled attenuation parameters (CAP) by FibroScan® (Echosens, Paris, France) were measured in 54 participants (52 NAFLDs and 2 controls defined by ultrasound). CAP measures ultrasonic attenuation in the liver at 3.5 MHz at depth in the M prove and 2.5 MHz in the XL probe. CAP measures ultrasonic attenuation in the liver at 3.5 MHz at depth between 25 and 65 mm. The trained investigators performed CAP measurement. A previous study suggested that a CAP cutoff point of 225 dB/m for predicting steatosis was determined with 0.87 sensitivity, 0.83 specificity, positive predictive value 0.71, negative predictive value 0.93, and area under the curve 0.93 in pediatric NAFLD [30]. In this study, all 53 participants showed above 225 dB/m; mean CAP 324.28 ± 38.56 (range 253-400) dB/m. In this study, the diagnostic sensitivity of hepatic ultrasonography was 96.3%.

To screen metabolic risks in children with overweight, abdominal circumference (AC), systolic and diastolic blood pressure (SBP and DBP), HbA1c, fasting blood sugar (FBS), insulin, total cholesterol (TC), triglyceride (TG), low-density lipoprotein cholesterol (LDL-C), and high-density lipoprotein cholesterol (HDL-C) levels and homeostatic model assessment of insulin resistance (HOMA-IR) were measured in children with overweight.

2.3. Noninvasive fibrosis parameters

Several noninvasive fibrosis parameters have been used for pediatric NAFLD, such as the pediatric NAFLD fibrosis score (PNFS), the AST/platelet ratio index (APRI), and the fibrosis-4 (FIB-4) score [31,32]. A previous Korean study suggested that NAFLD showed that APRI and FIB-4 might be useful noninvasive fibrosis scores in children with NAFLD [32]. Another study showed that PNFS had better accuracy for identifying advanced fibrosis than the APRI and FIB-4 score [33]. Therefore, three noninvasive fibrosis parameters were used to evaluate the degree of hepatic fibrosis, as follows: the PNFS, determined using the ALT, ALP, and GGT levels and the platelet count [33]; the APRI, determined by (AST level/AST upper level of normal/platelet count) × 100 [34]; and the FIB-4 score, determined by (age × AST level/platelet count × $\sqrt{}$ ALT) [35].

These fibrosis scores were calculated in all NAFLD patients (n = 228) and controls with overweight (n = 71). Because of the sensitivity and specificity of the diagnostic accuracy of ultrasonography, especially in mild steatosis [36,37], these fibrosis scores were analyzed in controls with overweight to evaluate the potential risk of NAFLD and fibrosis.

2.4. Genotyping and genetic analysis

Genomic DNA was extracted from peripheral blood using a Gentra PureGene Blood Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Patients were genotyped for 4 variants of 3 genes, namely, *PNPLA3* (rs738409), *TM6SF2* (rs58542926) and *SAMM50* (rs2073080, rs3761472), by TaqMan allelic discrimination assay on an ABI 7900HT Real-Time PCR system (Applied Biosystem, Foster City, CA, USA). Predesigned assay primers and probes were purchased from Applied Biosystems.

A logistic regression test was performed to evaluate the effects of 4 genetic variants on the development of NAFLD in the genotype model and additive model. In the genotype model, genetic variants were considered as categorical variables. In the additive model, wild type was coded as 0, heterozygotes as 1, and homozygotes as 2. In the analysis of *TM6SF2* rs58542926, wild-type vs risk allele carriers (dominant model) were compared because of low minor allele frequency (MAF). To evaluate the synergetic effects of 4 genetic variants on liver function tests (AST, ALT and ALP) and fibrosis scores (PNFS, APRI and FIB-4), the genetic risk score was calculated as giving value as 1 for carrying 0 or 1 risk allele, 2 for carrying 2 to 3 risk alleles, 3 for carrying 4 to 5 risk alleles, and 4 for carrying 6 to 7 risk alleles, resulting in a range of 1 to 4.

2.5. Metabolomic analysis①

After overnight fasting, 4 mL of blood samples was collected and centrifuged at 4° C, and plasma samples were stored at −80° C until use. A commercially available high−resolution mass spectrometry−based AbsoluteIDQ® p400 HR Kit by Biocrates® (Innsbruck, Austria) was used. This quantitative kit covered 408 metabolites of several classes, including 21 amino acids, 21 biogenic amines, 1 monosaccharide, 55 acylcarnitines, 18 diglycerides (DGs), 42 TGs, 24 lysophosphatidylcholines (LPCs), 127 phosphatidylcholines (PCs), 31 sphingolipids including sphingomyelins (SMs) and ceramides (Cer) and 14 cholesterol esters (CEs).

Plasma samples for metabolomic profiling were prepared in accordance with the laboratory manual provided by Biocrates® and measured using a Q Exactive Plus orbitrap mass spectrometer with a heated electrospray ionization source coupled with an Ultimate 3000 ultra−performance liquid chromatography (Thermo Fisher Scientific, MA, USA). In brief, thawed plasma samples on ice were centrifuged, and the supernatant was used. For quantifying amino

① Metabolomic experiments and raw data generation were conducted by Woori Chae and Prof. Joo−Youn Cho.

acids and biogenic amines, liquid chromatography-based mass spectrometry was used. Quantitation of raw data and metabolites via calibration curves was performed using Xcalibur software by Thermo Fisher Scientific (MA, USA). Concentration data were exported in.csv format and imported to MetIDQ software by Biocrates® (Innsbruck, Austria) for further processing. Flow injection-based mass spectrometry was applied to assess hexoses, lipids, acylcarnitines and cholesterol esters. Raw data processing and metabolite quantitation were performed using MetIDQ software. To minimize batch effects, the entire concentration of the plasma metabolome was normalized by pooled quality control samples. Then, adjusted concentration data were exported for further analyses in.xlsx format.

2.6. Statistical analysis

Differences in categorical variables (e.g., patient group, sex) according to genotype were compared by the chi-squared test, and those in continuous variables (e.g., age, BMI-Z score, AST level, ALT level, noninvasive fibrosis indexes) were compared by analysis of variance with post hoc analysis or T-test. The Kruskal-Wallis or Mann-Whitney U test was used to compare fasting glucose, HbA1c and HOMA-IR, which were not normally

distributed. A general linear model was used to evaluate differences in continuous variables according to genotype. A logistic regression analysis was performed to calculate the odds ratio (OR) and 95% confidence interval (CI) of the presence of NAFLD according to genotype. Since three variants, *PNPLA3* rs738409, *SAMM50* rs2073080, and *SAMM50* rs3761472, showed multicollinearity (variance inflation factors >10), logistic regression analysis was performed using *TM6SF2* rs58542926 and one of these three variants as covariates. A univariate analysis was performed to find potential compounding variables for NAFLD. Based on the univariate analysis, sex, BMI-Z score, AC, blood pressure (SBP, DBP), fasting lipids (TC, TG, LDL-C, HDL-C), fasting insulin and HOMA-IR were significant variables. To reduce multicollinearity, BMI-Z score age, sex, SBP, TGs and insulin were selected as compounding variables. The associations of the genetic risk score with liver function tests and fibrosis scores were analyzed using a general linear model adjusted for confounding factors. Statistical analysis was performed using SPSS Statistics software, version 23.0 (IBM Corp., Armonk, NY, USA), and data are expressed as the mean ± standard deviation and median ± 95% CI according to their distributions. Two-sided P values less than 0.05 were considered statistically significant in genetic analysis.

Web-based MetaboAnalyst 5.0 (http://www.metaboanalyst.ca/) was used to perform descriptive statistics and to identify metabolomic biomarkers. To evaluate the pathways, enrichment analysis was conducted, performing metabolite set enrichment analysis of human and mammalian species based on several libraries containing ~9000 groups of metabolite sets. Missing values were replaced by 1/5 of the minimum positive values of their corresponding variables. Some variables were appropriately log-transformed when required. The t-test was performed to evaluate differences between subjects with NAFLD and controls. Because BCAAs, glycine, carnitine and lipids such as TG and DG are known metabolites associated with NAFLD [38-42], valine, isoleucine, leucine, glycine, carnitine and TG(50:2) were used as positive controls to confirm the quality of metabolomic samples. Pearson's correlation coefficient was used to evaluate the association between HOMA-IR and metabolites.

Metabolomic differences of each genetic variant were also analyzed. P-values were corrected for multiple testing by controlling the false discovery rate (FDR) for metabolomic analysis.

Chapter 3. Results

Baseline characteristics

The mean age of participants for genetic analysis was 12.6 ± 3.2 (range 5.8 ~ 20.3) years. There were significantly more boys in the NAFLD group (n = 178, 78.1%) than in the control group (n = 138, 61.3%) ($P < 0.001$). There were no age differences between NAFLD patients and controls; however, the BMI-Z score was significantly higher in the NAFLD patients than in the controls ($P < 0.001$). The mean levels of AST, ALT, ALP, GGT and uric acid were significantly higher in the NAFLD patients than in the controls (Table 1). There were no differences between the control with overweight and normal weight controls. Regarding metabolic profiles, NAFLD patients showed significantly higher AC, SBP, DBP, fasting glucose, insulin, TC, TG, LDL-C and HOMA-IR than controls. The mean HDL-C level was significantly lower in the NAFLD patients than in the controls ($P = 0.003$). However, there was no significant difference in HbA1c levels between the two groups (Table 1).

These differences between NAFLD patients and controls were also shown in participants for metabolomic analysis (Table 1). More

boys (n = 83, 79%) and significantly higher BMI-Z, AC, AST, ALT, ALP, GGT, uric acid, SBP, fasting insulin, TC, TG, LDL-C and HOMA-IR were observed in the NAFLD group.

3.1. Genetic association of pediatric NAFLD

3.1.1 Genotype frequencies

The allele frequencies of the genotype distribution of the 4 variants were in Hardy-Weinberg equilibrium ($P > 0.05$). The risk allele frequencies of *PNPLA3* rs738409 (G), *TM6SF2* rs58542926 (T), *SAMM50* rs2073080 (T) and *SAMM50* rs3761472 (G) were 0.499, 0.099, 0.488 and 0.478, respectively. These results are consistent with the findings of the Allele Frequency Database 1000 Genomes Project (https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/). Risk allele frequencies of *PNPLA3* rs738409 (G), *SAMM50* rs2073080 (T) and *SAMM50* rs3761472 (G) were higher in Korean children and Asians (Korean adults, Han Chinese and Japanese) than in other ethnicities (Table 2).

Table 1. Baseline

	Genetic analysis (N = 453)			Metabolomic analysis (N = 166)		
	Control (n = 225)	NAFLD (n = 228)	P	Control (n = 61)	NAFLD (n = 105)	P
Male	138 (61.3%)	178 (78.1%)	<0.001	39 (63.9%)	83 (79%)	0.033
Female	87 (38.7%)	50 (21.9%)		22 (36.1%)	22 (21%)	
Age	12.54 ± 3.43	12.59 ± 3.06	0.587	13.52 ± 3.09	12.97 ± 3.04	0.267
BMI-Z	0.22 ± 1.48	2.51 ± 1.01	<0.001	0.31 ± 1.38	2.45 ± 1.07	<0.001
AC, cm	79.25 ± 10.9	90.88 ± 11.68	<0.001	80.10 ± 10.18	91.19 ± 12.4	0.001
AST, IU/L	22.67 ± 8.84	66.9 ± 51.7	<0.001	22.56 ± 12.96	58.68 ± 41.83	<0.001
ALT, IU/L	16.03 ± 5.27	123.82 ± 108.35	<0.001	20.38 ± 26.12	106.76 ± 97.74	<0.001
ALP, U/L	206.56 ± 04.27	263.18 ± 127.0	<0.001	205.2 ± 05.81	260.8 ± 139.03	0.008
GGT, IU/L	14.53 ± 7.91	48.78 ± 39.89	<0.001	16.13 ± 19.25	45.76 ± 37.19	<0.001
Uric acid, mg/dL	5.34 ± 1.51	6.67 ± 1.55	<0.001	5.67 ± 1.47	6.58 ± 1.70	0.001
SBP (mmHg)	111.38 ± 13.61	123.59 ± 12.78	<0.001	117.76 ± 3.42	124.17 ± 1.79	0.029
DBP (mmHg)	66.51 ± 0.03	72.58 ± 11.3	<0.001	72.52 ± 73.79	73.79 ± 9.03	0.581
FBS (mg/dL)	98 (93.5-100.7)	99 (99.6-109)	0.709	103 (98.2-107.3)	101 (96-116.6)	0.170

Insulin (μU/mL)	12.2 ± 5.90	19.7 ± 9.95	<0.001	9.98 ± 5.62	19.35 ± 10.69	0.001
TC (mg/dL)	155.62 ± 29.38	178.14 ± 34.47	<0.001	157.61 ± 30.34	174.25 ± 35.78	0.004
TG (mg/dL)	101.41 ± 5.21	138.38 ± 78.97	0.001	95.11 ± 41.14	137.39 ± 73.01	0.005
HDL-C (mg/dL)	51.58 ± 9.34	47.19 ± 9.85	0.003	52.96 ± 10.05	48.66 ± 11.09	0.075
LDL-C (mg/dL)	105.55 ± 27.87	116.03 ± 33.57	0.033	103.23 ± 26.51	110.56 ± 26.15	0.207
HbA1c (%)	5.35 (5.23-5.43)	5.5 (5.36-6.43)	0.012	5.13 (5.14-5.2)	5.3 (5.29-5.88)	0.048
HOMA-IR	2.78 (2.42-3.35)	4.58 (4.46-5.57)	<0.001	2.61 (1.34-4.64)	4.18 (3.87-5.73)	<0.001

Abbreviations: NAFLD; nonalcoholic fatty liver disease; BMI-Z score, body mass index Z score; AC, abdominal circumference; AST, aspartate transaminase; ALT, alanine aminotransferase; ALP, alkaline phosphatase; GGT, gamma glutamyl transferase; SBP, systolic blood pressure; DBP, diastolic blood pressure; FBS, fasting blood sugar; TC, total cholesterol; TG, triglyceride; HDL-C high-density lipoprotein cholesterol, LDL-C, low-density lipoprotein cholesterol; HOMA-IR, homeostatic model assessment of insulin resistance

Table 2. Comparison of risk allele frequencies by countries

		Risk allele frequency in this study		Risk allele frequency in Korean adults[19]		Risk allele frequency in other countries					
Gene name	Genetic variants	Control	NAFLD	Korean adult control	Korean adults NAFLD	Han Chinese	Japanese	Americans of African	Indian	British in England	Finnish in Finland
PNPLA3	rs738409 (G)	0.429	0.571	0.398	0.466	0.391	0.423	0.172	0.221	0.253	0.172
TM6SF2	rs58542926 (T)	0.069	0.129	0.066	0.094	0.114	0.072	0.082	0.123	0.071	0.061
SAMM50	rs2073080 (T)	0.422	0.553	0.395	0.454	0.403	0.433	0.218	0.216	0.236	0.232
SAMM50	rs3761472 (G)	0.402	0.553	0.39	0.45	0.395	0.428	0.148	0.216	0.187	0.232

Abbreviations: NAFLD; nonalcoholic fatty liver disease; *PNPLA3, phospholipase-containing domain 3; TM6SF2, transmembrane 6 superfamily member*

Figure 2 shows the genotype distributions of the 4 variants in subjects with NAFLD and the controls. Homozygous carriers of the risk alleles *PNPLA*3 (GG) and *SAMM50* (TT and GG) had a significantly higher frequency of NAFLD than the controls. Carriers of the *TM6SF2* CT/TT genotypes showed a significantly higher frequency of NAFLD (65.5%) than carriers of the wild type (CC) (46.7%, P = 0.02).

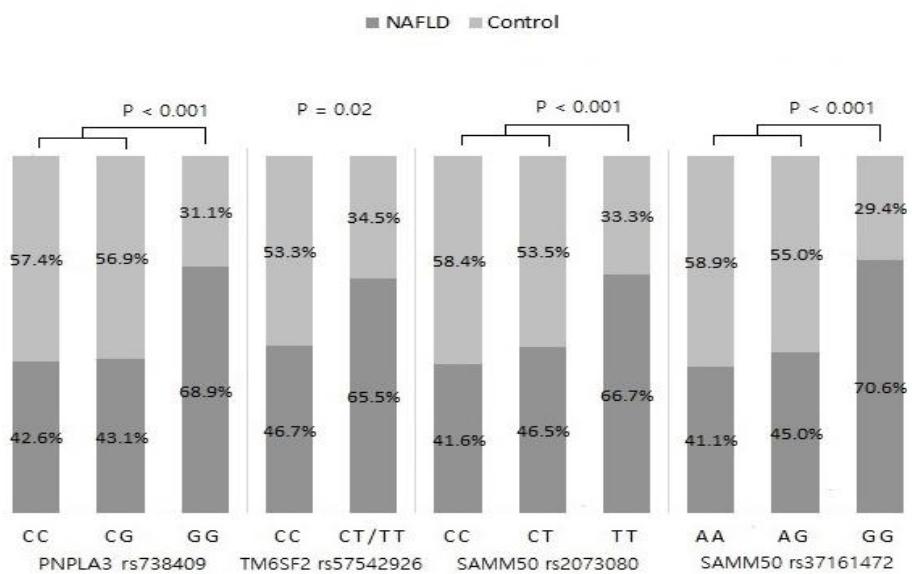

Figure 2. Genotype distribution of *PNPLA3* rs738409, *TM6SF2* rs58542926, and *SAMM50* rs2073080 and rs3761472 in subjects with NAFLD and controls The frequencies of the *PNPLA3* GG, *TM6SF2* CC/TT, and *SAMM50* TT and GG genotypes were significantly higher in the NAFLD group than in the control group.
Abbreviations: *PNPLA3, phospholipase-containing domain 3; TM6SF2, transmembrane 6 superfamily member 2*

3.1.2 Clinical characteristics according to genotype

In all participants, individuals with the *PNPLA3* rs738409 (GG), *TM6SF2* (CT/TT), and *SAMM50* rs2073080 (TT) and rs3761472 (GG) genotypes had significantly higher AST and ALT levels than those with other genotypes (Table 3). The grade of steatosis on sonography was significantly different among all 4 variants.

In the analysis of metabolic features in children with overweight, only *TM6SF2* showed a significant difference in the lipid profile (Table 3). The CT/TT genotype of *TM6SF2* rs58542926 was associated with significantly lower plasma TC and TG levels than the wild type (P = 0.033 and P = 0.038). In simple linear regression analysis, the risk allele (T) of *TM6SF2* rs57542926 also significantly reduced the TC (ß = −10.093, P = 0.036) and TG (ß = −20.962, P = 0.035) levels. There were no significant differences in metabolic features, such as blood pressure, fasting glucose, insulin, HbA1c and HOMA-IR, for all 4 variants.

Table 3. Clinical characteristics and metabolic features by genotype

	PNPLA3 rs738409				*TM6SF2* rs58542926			*SAMM50* rs2073080				*SAMM50* rs3761472			
	CC	CG	GG	P-value	CC	CT/TT	P-value	CC	CT	TT	P-value	AA	AG	GG	P-value
All participants	(n = 122)	(n = 204)	(n = 122)		(n = 366)	(n = 87)		(n = 125)	(n = 213)	(n = 114)		(n = 129)	(n = 211)	(n = 109)	
BMI-Z score	1.31 ± 1.6	1.3 ± 1.8	1.6 ± 1.6	0.155	1.3 ± 1.7	1.49 ± 1.5	0.473	1.3 ± 1.7	1.3 ± 1.8	1.6 ± 1.6	0.175	1.32 ± 1.68	1.19 ± 1.8	1.8 ± 1.4	0.011[b]
AST (IU/L)	40.0 ± 41.6	39.5 ± 38.7	59.1± 49.6	<0.001[a]	42.1 ± 39.9	57.0 ± 53.8	0.004	39.8 ± 42.6	40.0 ± 36.2	59.8 ± 52.3	<0.001[a]	38.2 ± 40.3	40.4 ± 37.5	61.8 ± 52.7	<0.001[a]
ALT (IU/L)	58.2 ± 92.8	58.1 ± 81.7	102.4 ± 109.2	<0.001[a]	63.9 ± 87.6	97.3 ± 115.9	0.003	57.0 ± 94.1	60.8 ± 78.8	102.8 ± 113.5	<0.001[a]	55.4 ± 92.5	60.0 ± 78.7	107.7 ± 113.9	<0.001[a]
ALP (U/L)	218.1 ± 119.1	224.5 ±116.7	267.9 ±119.7	0.001[a]	233.6 ± 118.0	241.4 ± 126.4	0.584	209.1 ± 116.5	231.0 ± 117.3	271.3 ± 119.6	<0.001[a]	212.2 ± 117.9	229.1 ± 118.2	274.2 ± 117.9	<0.001[a]
GGT (IU/L)	28.1 ± 40.1	16.5 ± 30.7	39.3 ± 34.2	0.007[b]	28.7 ± 31.2	38.5 ± 46.5	0.03	28.3 ± 39.8	27.8 ± 31.4	37.8 ± 33.6	0.05	28.02 ± 39.21	27.4 ± 31.5	39.4 ± 33.8	0.016[b]
Children with overweight	(n = 76)	(n = 114)	(n = 90)		(n = 225)	(n = 59)		(n = 76)	(n = 122)	(n = 85)		(n = 80)	(n = 115)	(n = 86)	
BMI-Z score	2.3 ± 0.9	2.6 ± 1	2.4 ± 0.9	0.079	2.5 ± 1.0	2.3 ± 1.0	0.252	2.4 ± 1.0	2.6 ± 1.0	2.4 ± 0.9	0.331	2.4 ± 1.0	2.6 ± 1.0	2.4 ± 0.9	0.313
AC	92 ± 12.5	89.3 ± 11.6	88.8 ± 12.1	0.367	89.8 ± 12.1	89.5 ± 11.7	0.895	92.7 ± 12.6	88.9 ± 11.9	88.7 ± 11.5	0.178	92.6 ± 2.4	89.3 ± 12.4	88.4 ± 11.6	0.175
SBP (mmHg)	122.3 ± 13.8	122.7 ± 12.9	121.2 ± 13.8	0.765	122.2 ± 13.2	122.5 ± 14.5	0.863	122.4 ± 13.5	123.6 ± 13.2	120.2 ± 13.6	0.262	122.3 ± 13.5	123.4 ± 13.1	120.5 ± 13.7	0.363
DBP	73.8 ±	71.8 ±	70.3 ±	0.194	72.0 ±	71.4 ±	0.733	74.1 ±	71.8 ±	70.1 ±	0.132	73.7 ±	72.0 ±	70.0 ±	0.154

(mmHg)	10.5	10.6	12.9		11.3	11.7		10.5	10.3	13.1		10.4	10.5	13.0	
FBS (mg/dL)	106.1 ± 35.1	100.8 ± 27.5	100.1 ± 15.2	0.288	101.3 ± 23.5	103.9 ± 36.1	0.515	105.8 ± 35.3	100.5 ± 26.7	100.3 ± 15.3	0.330	105.6 ± 34.5	100.6 ± 27.4	100.3 ± 15.2	0.359
Insulin (μU/mL)	18.7 ± 10.3	18.8 ± 10.9	18.8 ± 9.3	0.999	19.4 ± 10.0	16.6 ± 8.7	0.080	18.3 ± 10.2	19.2 ± 10.1	18.4 ± 9.0	0.810	18.8 ± 10.2	18.8 ± 10.3	18.4 ± 9.0	0.959
TC (mg/dL)	176.2 ± 32.6	175.1 ± 35.6	175.2 ± 32.9	0.975	177.7 ± 33.8	167.1 ± 32.9	0.033	174.2 ± 34.2	176.4 ± 34.0	175.5 ± 33.5	0.906	173.8 ± 33.9	177.3 ± 34.5	175.2 ± 33.5	0.773
TG (mg/dL)	123.9 ± 59.3	123.8 ± 57.6	142.5 ± 82.7	0.116	134.2 ± 69.3	113.0 ± 57.9	0.038	115.3 ± 53.8	130.8 ± 62.3	140.5 ± 82.0	0.079	117.8 ± 54.2	130.6 ± 63.2	140.0 ± 81.6	0.136
HDL-C (mg/dL)	49.4 ± 7.8	48.5 ± 10.1	46.7 ± 11.3	0.229	48.5 ± 10.4	46.3 ± 8.3	0.136	49.2 ± 7.7	47.7 ± 10.1	47.6 ± 11.6	0.549	48.9 ± 7.8	48.0 ± 10.3	47.5 ± 11.5	0.706
LDL-C (mg/dL)	116.1 ± 28.1	113.5 ± 28.8	114.4 ± 40.2	0.887	116.0 ± 33.8	108.8 ± 28.0	0.150	113.4 ± 28.9	115.0 ± 27.6	114.9 ± 41.3	0.944	114.2 ± 28.9	114.8± 28.0	114.7 ± 41.1	0.991
HbA1c (%)	5.6 ± 0.7	6.1 ± 5.1	5.6 ± 0.9	0.589	5.9 ± 3.7	5.5 ± 0.5	0.548	5.6 ± 0.6	6.0 ± 4.9	5.6 ± 0.9	0.607	5.6 ± 0.6	6.1 ± 5.0	5.6 ± 0.9	0.550
HOMA-IR	4.9 ± 3.3	4.9 ± 4.3	4.7 ± 2.7	0.932	5.0 ± 3.8	4.2 ± 2.2	0.139	4.8 ± 3.3	5.0 ± 4.3	4.6 ± 2.6	0.777	4.9 ± 3.3	5.0 ± 4.4	4.6 ± 2.6	0.831

[a] Homozygous vs. heterozygous and wild type [b] Homozygous vs. heterozygous

Abbreviations: *PNPLA3, phospholipase-containing domain 3; TM6SF2, transmembrane 6 superfamily member 2*; BMI-Z score, body mass index Z score; AC, abdominal circumference; AST, aspartate transaminase; ALT, alanine aminotransferase; ALP, alkaline phosphatase; GGT, gamma glutamyl transferase; SBP, systolic blood pressure; DBP, diastolic blood pressure; FBS, fasting blood sugar; TC, total cholesterol; TG, triglycerides; LDL-C, low-density lipoprotein cholesterol; HDL-C high-density lipoprotein cholesterol, HOMA-IR, homeostatic model assessment of insulin resistance

3.1.3 Genetic association of pediatric NAFLD

Noninvasive fibrosis scores, such as the PNFS and APRI, were significantly higher in risk allele carriers for each gene (Figure 3). The mean PNFS was significantly higher in carriers of *PNPLA3* GG (77.41 ± 18.05) than in carriers of CC (65.40 ± 18.87, P < 0.001) or CG (68.0 ± 20.5, P = 0.001). For *SAMM50* rs2073080 and rs3761472, the mean PNFS was also significantly higher in homozygous carriers than in others: rs2073080 TT (77.35 ± 18.17) vs CC (63.39 ± 20.5, P < 0.001) and CT (69.39 ± 20.57, P = 0.009); rs3761472 GG (77.41 ± 18.08) vs AA (63.93 ± 18.59, P < 0.001) and AG (69.55 ± 20.52, P = 0.010). For *TM6SF2*, the CT/TT genotypes (75.77 ± 18.67) showed a significantly higher PNFS than the CC genotype (68.85 ± 19.99, P = 0.013). The mean APRI was also higher in risk allele homozygotes than in those with other genotypes for *PNPLA3* rs738409 GG (0.58 ± 0.43) vs CC (0.40 ± 0.45, P = 0.014) and CG (0.44 ± 0.42, P = 0.039) and *SAMM50* rs2073080 TT (0.59 ± 0.45) vs CC (0.41 ± 0.46, P = 0.027) and CT (0.43 ± 0.39, P = 0.026). The homozygous risk allele carrier of *SAMM50* rs3761472 showed a significantly higher APRI than the wild type GG (0.58 ± 0.45) vs AA (0.38 ± 0.41, P = 0.007). For *TM6SF2*, those with the CT/TT genotype had a

significantly higher APRI (0.60 ± 0.54) than those with the CC genotype (0.44 ± 0.39, P = 0.009).

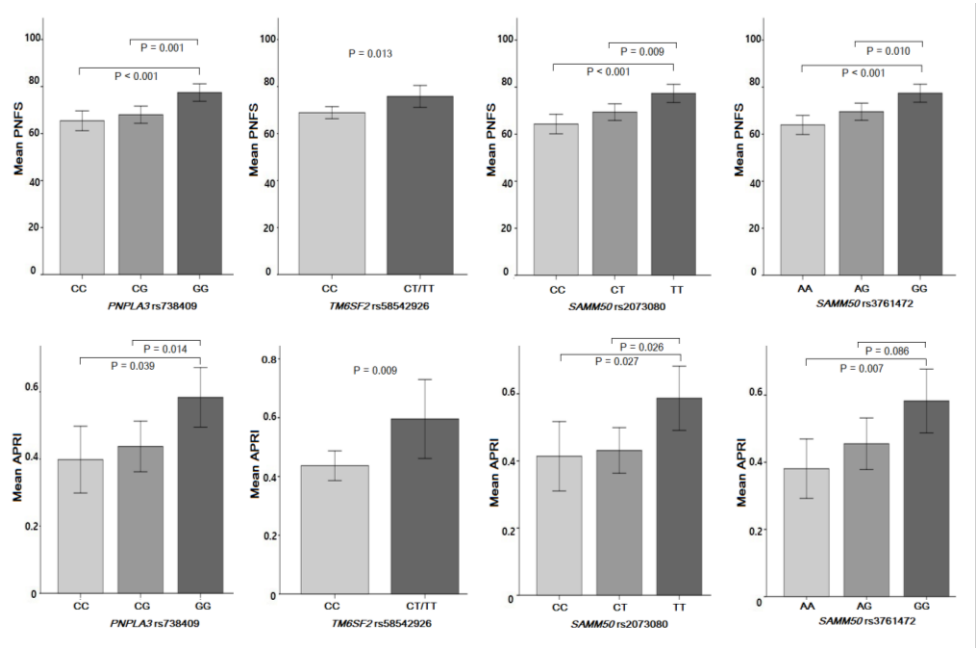

Figure 3. Pediatric NAFLD fibrosis score (PNFS) and AST/platelet ratio index (APRI) according to genotype in NAFLD patients and controls with overweight.

The PNFS and APRI were significantly higher in risk allele carriers for the 4 variants. For *PNPLA3* and *SAMM50*, homozygous risk alleles were associated with a significantly higher PNFS and APRI than the wild type. For *TM6SF2*, homozygous and heterozygous risk alleles significantly increased the PNFS and APRI compared with the wild type. Abbreviations: *PNPLA3,* phospholipase–containing domain 3; *TM6SF2*, transmembrane 6 superfamily member 2; PNFS, pediatric NAFLD fibrosis score; APRI, AST/platelet ratio index

3.1.4 Independent risk factors for the development and severity of NAFLD

In all participants, after adjusting for sex, age, and BMI-Z score, the 4 risk alleles remained significant risk factors for NAFLD. Subjects with homozygous variants of *PNPLA3* rs738409 (genotype model 1, OR = 5.78, 95% CI = 2.61-12.8, P < 0.001), *SAMM50* rs2073080 (genotype model 2, OR = 5.18, 95% = CI 2.33-11.49, P < 0.001) and *SAMM50* rs3761472 (genotype model 3, OR = 5.71, 95% CI = 2.57-12.7, P < 0.001) had a five- to 6-fold higher risk of pediatric NAFLD than wild-type subjects (Table 4). Subjects with the CT/TT genotype of *TM6SF2* rs58542926 showed a 4 times higher risk of pediatric NAFLD (OR = 3.87-4.21, 95% CI = 1.85-8.75) than wild-type subjects (P < 0.001) (Table 4). Heterozygotes of *PNPLA3* rs738409, *SAMM50* rs2073080 and *SAMM50* rs3761472 did not increase the risk of NAFLD.

In the analysis of the additive model, as the risk alleles of *PNPLA3* rs738409 (additive model 1), *SAMM50* rs2073080 (additive model 2), *SAMM50* rs3761472 (additive model 3) and *TM6SF2* rs58542926 (additive model 1-3) increased, the risk of NAFLD independently increased (Table 5). Male sex and BMI-Z

score were also independent risk factors in all participants.

In the subgroup analysis among children with overweight, these 4 variants remained significant risk factors for NAFLD with a higher OR after adjusting for sex, age, BMI-Z score, SBP, fasting insulin and TGs. Subjects with the CT/TT genotype of *TM6SF2* rs58542926 showed a 36- to 42-fold higher risk of pediatric NAFLD (genotype model 4-6, OR = 35.78-46.37, 95% CI = 3.18-519.93) than wild-type subjects in children who were overweight ($P < 0.05$) (Table 6). Subjects with homozygous variants of *PNPLA3* rs738409 (genotype model 4, OR = 24.5, 95% CI = 3.65-164.52, P = 0.001), *SAMM50* rs2073080 (genotype model 5, OR = 13.32, 95% CI = 2.6-68.25, P = 0.002) and *SAMM50* rs3761472 (genotype model 6, OR = 15.51, 95% CI = 2.99-80.48, P = 0.001) had a 13- to 25-fold higher risk of pediatric NAFLD than wild-type subjects (Table 6). In the analysis of the additive model, as the risk alleles of *PNPLA3* rs738409 (additive model 4), *SAMM50* rs2073080 (additive model 5), *SAMM50* rs3761472 (additive model 6) and *TM6SF2* rs58542926 (additive model 4-6), increased, the risk of NAFLD independently increased in children with overweight (Table 7). Genetic effects were greater in

children with overweight than in all participants. In addition, male sex, fasting insulin and triglycerides were also independent risk factors for NAFLD in children.

All 4 variants were significantly positively associated with the AST and ALT levels after adjusting for age, sex, and BMI-Z score (Table 8). As the number of risk alleles increased, the ALT level independently increased by 18.6 IU/L for *PNPLA3*, 18.6 to 20.0 IU/L for *SAMM50* and 33.8 IU/L for *TM6SF2*. The steatosis grade independently increased according to the number of risk alleles for all 4 variants.

In children with overweight, the increase in the ALT level related to the risk allele was greater than that in all participants, at 18.6 IU/L for *PNPLA3*, 22.5 to 22.8 IU/L for *SAMM50* and 47.3 IU/L for *TM6SF2*. When one risk allele increased, the PNFS independently increased by 4.5 for *PNPLA3*, 5.3 to 5.4 for *SAMM50,* and 6.6 for *TMS6SF2*. The APRI and FIB-4 score independently increased according to risk alleles for all 4 variants (APRI: 0.08-0.192, FIB-4 score: 0.021-0.035) (Table 8).

Table 4. Independent risk factors for pediatric NAFLD by genotype model

Genotype model 1				Genotype model 2				Genotype model 3			
Variable	OR	95% CI	P-value	Variable	OR	95% CI	P-value	Variable	OR	95% CI	P-value
PNPLA3 rs738409 (GG)[a]	5.78	2.61-12.80	<0.001	*SAMM50* rs2073080 (TT)[c]	5.18	2.33-11.49	<0.001	*SAMM50* rs3761472 (GG)[d]	5.71	2.57-12.7	<0.001
TM6SF2 rs58542926 (CT/TT)[b]	3.87	1.85-8.1	<0.001	*TM6SF2* rs58542926 (CT/TT)[b]	4.21	2.02-8.75	<0.001	*TM6SF2* rs58542926 (CT/TT)[b]	4.04	1.92-8.50	<0.001
Male sex	3.22	1.70-6.08	<0.001	Male sex	2.91	1.56-5.44	0.001	Male sex	2.99	1.59-5.60	0.001
BMI-Z	5.03	3.69-6.85	<0.001	BMI-Z	4.9	3.63-6.61	<0.001	BMI-Z	4.89	3.61-6.63	<0.001

Adjusted for age, sex, and BMI-Z score

a The *PNPLA3* rs738409 genetic variant has CC, CG, and GG genotypes; the CC genotype, the wild type, served as the reference group.

b The *TM6SF2* rs58542926 genetic variant has CC, CT and TT genotypes; the CC genotype, the wild type, served as the reference group.

c The *SAMM50* rs2073080 genetic variant has CC, CT and TT genotypes; the CC genotype, the wild type, served as the reference group.

d The *SAMM50* rs3761472 genetic variant has AA, AG and GG genotypes; the AA genotype, the wild type, served as the reference group.

Abbreviations: *PNPLA3, phospholipase-containing domain 3; TM6SF2, transmembrane 6 superfamily member 2*; BMI-Z score, body mass index Z score; OR, odds ratio, CI, confidence interval

Table 5. Independent risk factors for pediatric NAFLD by additive model

Additive model 1				Additive model 2				Additive model 3			
Variable	OR	95% CI	P-value	Variable	OR	95% CI	P-value	Variable	OR	95% CI	P-value
PNPLA3 rs738409 (G)	2.315	1.575-3.403	<0.001	*SAMM50* rs2073080 (T)	2.256	1.519-3.351	<0.001	*SAMM50* rs3761472 (G)	2.374	1.596-3.531	<0.001
TM6SF2 rs58542926 (T)	3.956	1.939-8.071	<0.001	*TM6SF2* rs58542926 (T)	4.195	2.065-8.521	<0.001	*TM6SF2* rs58542926 (T)	4.088	1.987-8.408	<0.001
Male	2.969	1.591-5.541	0.001	Male	2.842	1.525-5.296	0.001	Male	2.924	1.563-5.470	0.001
BMI-Z	4.939	3.648-6.688	<0.001	BMI-Z	4.910	3.639-6.626	<0.001	BMI-Z	4.948	3.648-6.711	<0.001

Adjusted for age, sex, and BMI-Z score

Abbreviations: *PNPLA3, phospholipase-containing domain 3; TM6SF2, transmembrane 6 superfamily member 2*; BMI-Z score, body mass index Z score; OR, odds ratio, CI, confidence interval

Table 6. Independent risk factors for pediatric NAFLD among children with overweight by genotype model

Genotype model 4				Genotype model 5				Genotype model 6			
Variable	OR	95% CI	P-value	Variable	OR	95% CI	P-value	Variable	OR	95% CI	P-value
PNPLA3 rs738409 (GG) [a]	24.5	3.65-164.5	0.001	*SAMM50* rs2073080 (TT) [c]	13.32	2.60-68.25	0.002	*SAMM50* rs3761472 (GG) [d]	15.51	2.99-80.48	0.001
TM6SF2 rs58542926 (CT/TT) [b]	35.78	3.18-402.3	0.004	*TM6SF2* rs58542926 (CT/TT) [b]	42.97	3.92-471.49	0.002	*TM6SF2* rs58542926 (CT/TT) [b]	46.37	4.14-519.9	0.002
Male sex	6.9	1.77-26.92	0.005	Male sex	5.56	1.59-19.52	0.007	Male sex	5.48	1.55-19.43	0.008
BMI-Z	3.0	1.32-6.83	0.009	BMI-Z	2.64	1.24-5.63	0.012	BMI-Z	2.66	1.24-5.72	0.012
Insulin	1.16	1.05-1.27	0.002	Insulin	1.14	1.05-1.25	0.003	Insulin	1.15	1.05-1.25	0.002
Triglyceride	1.02	1.00-1.03	0.019	Triglyceride	1.02	1.00-1.03	0.013	Triglyceride	1.02	1.00-1.03	0.013

Adjusted for age, sex, BMI-Z score, SBP, insulin and triglyceride

a The *PNPLA3* rs738409 genetic variant has CC, CG, and GG genotypes; the CC genotype, the wild type, served as the reference group.

b The *TM6SF2* rs58542926 genetic variant has CC, CT and TT genotypes; the CC genotype, the wild type, served as the reference group.

c The *SAMM50* rs2073080 genetic variant has CC, CT and TT genotypes; the CC genotype, the wild type, served as the reference group.

d The *SAMM50* rs3761472 genetic variant has AA, AG and GG genotypes; the AA genotype, the wild type, served as the reference group.

Abbreviations: *PNPLA3, phospholipase-containing domain 3; TM6SF2, transmembrane 6 superfamily member 2*; BMI-Z score, body mass index Z score; OR, odds ratio, CI, confidence interval

Table 7. Independent risk factors for pediatric NAFLD among children with overweight by additive model

Additive model 4				Additive model 5				Additive model 6			
Variable	OR	95% CI	P-value	Variable	OR	95% CI	P-value	Variable	OR	95% CI	P-value
PNPLA3 rs738409 (G)	3.488	1.674-7.267	0.001	*SAMM50* rs2073080 (T)	3.077	1.517-6.241	0.002	*SAMM50* rs3761472 (G)	3.418	1.648-7.088	0.001
TM6SF2 rs58542926 (T)	46.493	4.22-512.9	0.002	*TM6SF2* rs58542926 (T)	45.79	4.29-489.37	0.002	*TM6SF2* rs58542926 (T)	50.345	4.57-555.0	0.001
Male	5.066	1.451-17.69	0.011	Male	4.786	1.432-15.998	0.011	Male	4.885	1.429-16.70	0.011
BMI-Z	2.440	1.153-5.164	0.020	BMI-Z	2.419	1.169-5.002	0.017	BMI-Z	2.500	1.188-5.258	0.016
Insulin	1.154	1.058-1.259	0.001	Insulin	1.143	1.050-1.244	0.002	Insulin	1.152	1.057-1.256	0.001
Triglyceride	1.017	1.003-1.032	0.018	Triglyceride	1.016	1.003-1.030	0.017	Triglyceride	1.017	1.003-1.030	0.016

Adjusted for age, sex, BMI-Z score, SBP, insulin and triglyceride

Abbreviations: *PNPLA3, phospholipase-containing domain 3; TM6SF2, transmembrane 6 superfamily member 2*; BMI-Z score, body mass index Z score; OR, odds ratio, CI, confidence interval

Table 8. Associations of the 4 variants with clinical traits

	PNPLA3 rs738409			*TM6SF2 rs58542926*			*SAMM50* rs2073080			*SAMM50* rs3761472		
	Coefficient estimate	SE	P-value	Coefficient estimate	SE	P-value	Coefficient estimate	SE	P-value	Coefficient estimate	SE	P-value
AST (IU/L) [a]	8.0	2.63	0.003	15.3	4.6	0.001	8.0	2.67	0.003	9.18	2.67	0.001
AST (IU/L) [b]	7.2	3.95	0.07	22.91	6.94	0.001	8.5	4.05	0.037	9.01	3.98	0.025
ALT (IU/L) [a]	18.57	5.57	0.001	33.84	9.75	0.001	18.56	5.65	0.001	19.97	5.66	<0.001
ALT (IU/L) [b]	18.6	8.64	0.033	47.33	15.29	0.002	22.5	8.87	0.012	22.8	8.71	0.01
PNFS [c]	4.52	1.552	0.004	6.573	2.778	0.019	5.348	1.601	0.001	5.382	1.571	0.001
APRI [c]	0.080	0.034	0.020	0.192	0.059	0.001	0.091	0.035	0.010	0.094	0.034	0.007
FIB-4 [c]	0.022	0.010	0.025	0.035	0.017	0.036	0.021	0.010	0.037	0.021	0.010	0.028

[a] Analysis of all participants, adjusted for age, sex, and BMI-Z score

[b] Analysis in children in overweight, adjusted for age, sex, BMI-Z score, systolic blood pressure, insulin and triglyceride level

[c] Analysis of NAFLD patients and overweight controls, adjusted for age, sex, BMI-Z score, systolic blood pressure, insulin and triglyceride level

Abbreviations: *PNPLA3, phospholipase-containing domain 3; TM6SF2, transmembrane 6 superfamily member 2*; AST, aspartate transaminase; ALT, alanine aminotransferase, PNFS, pediatric NAFLD fibrosis score; APRI, AST/platelet ratio index; FIB-4, fibrosis-4; SE, standard error

3.1.5 Synergetic effects on the severity of NAFLD

Synergetic effects of these 4 variants were found on the severity of pediatric NAFLD. As the genetic risk score increased, the mean AST ($P < 0.001$), ALT ($P < 0.001$), ALP ($P = 0.013$) and GGT ($P = 0.026$) independently increased after adjustments for age, sex and BMI-Z (Figure 4). Furthermore, as the genetic risk score increased, the mean PNFS ($P < 0.001$), APRI ($P < 0.001$), and FIB-4 ($P = 0.004$) also significantly increased in NAFLD patients and control with overweight (Figure 4).

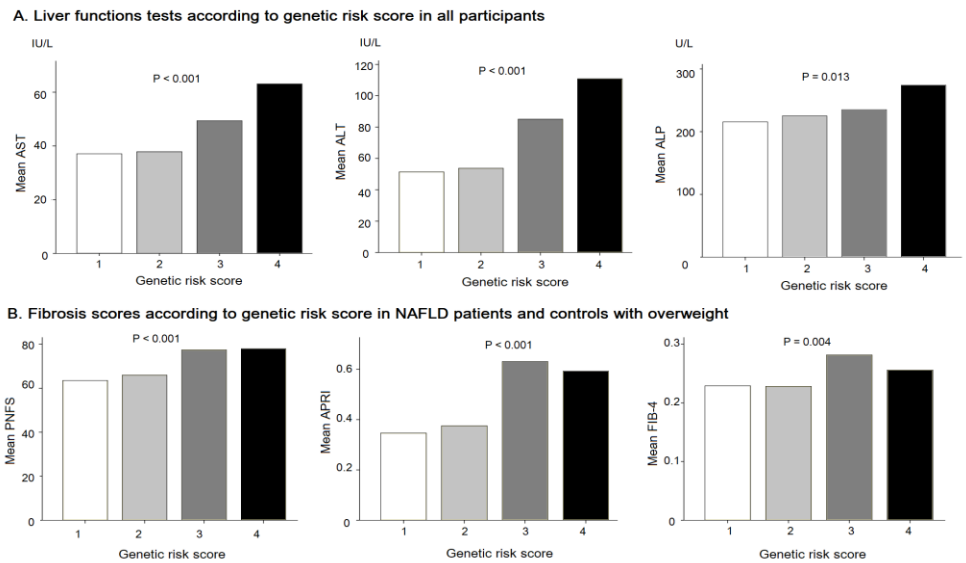

Figure 4. Synergetic effects of *PNPLA3* rs738409, *TM6SF2* rs58542926, and *SAMM50* rs2073080 and rs3761472 on liver function tests and fibrosis scores.

Abbreviations: AST, aspartate transaminase; ALT, alanine aminotransferase; ALP, alkaline phosphatase; PNFS, pediatric NAFLD fibrosis score; APRI, AST/platelet ratio index; FIB-4, fibrosis-4

3.2 Metabolomic difference between NAFLD and control children

Principal component analysis (PCA) score plots discriminating the control (red) from the NAFLD patients (green, Figure 5). Samples were appropriate for metabolomic analysis because quality control samples (blue) were located at the center of the PCA score plot. In the positive control analysis, the plasma levels of BCAAs, carnitine and TG(50:2) were significantly higher in NAFLD than in the control, while the glycine level was lower in NAFLD. These results were the same as those of previous studies suggesting that the quality of metabolomic samples was appropriate (Figure 5) [38-42].

A total of 101 metabolites were significantly different between NAFLD and the control: 11 amino acids, 5 biogenic amines, 4 acylcarnitines, 33 PCs, 6 LPCs, 14 sphingolipids and 28 neutral lipids (Supplementary Table 1). Forty-nine metabolites (1 AC, 8 amino acids, 15 PCs, 5 DGs, 14 TGs, 6 SMs) showed a significant correlation with HOMA-IR (Table 9). Higher metabolites in NAFLD were positively correlated with HOMA-IR, and lower metabolites in NAFLD were negatively correlated with HOMA-IR. However, metabolites were not significantly different between overweight and normal-weight NAFLD subjects.

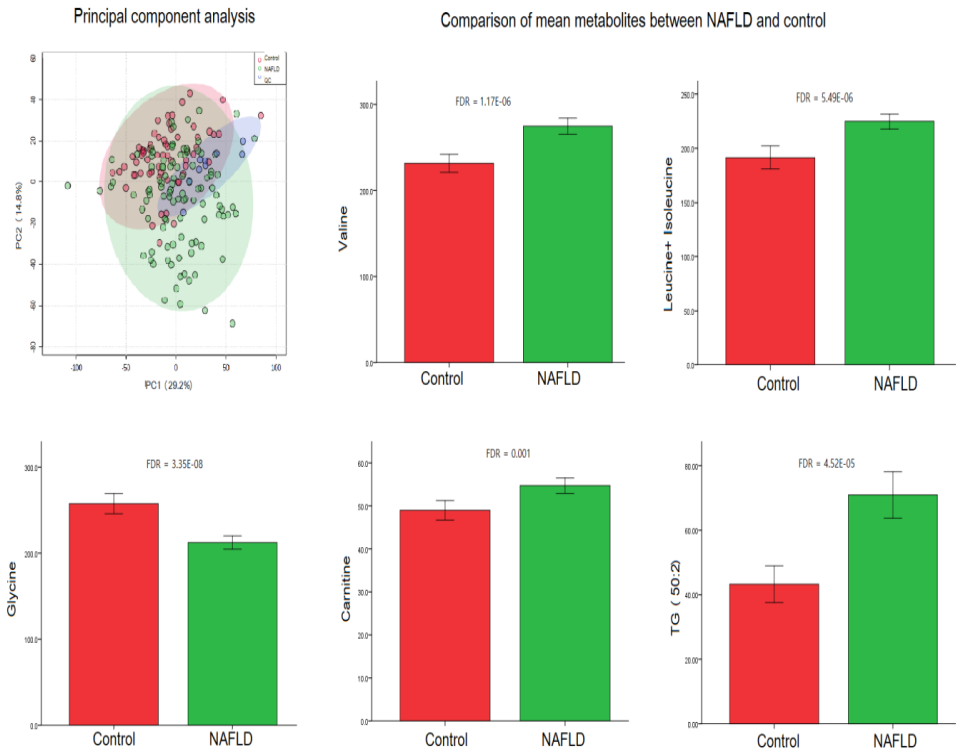

Figure 5. Principal component analysis of whole metabolomes and positive control analysis

Abbreviations: NAFLD, nonalcoholic fatty liver disease. TG. Triglyceride

Table 9. Metabolites associated with pediatric NAFLD and insulin resistance

	Control vs NAFLD			Correlation with HOMA-IR	
	t value[a]	P-value	FDR	r^2	P-value
Methylmalonylcarnitine	-2.615	0.009755	0.0315	0.284	0.007
Glycine	6.6613	3.92E-10	3.35E-08	-0.27	0.011
Glutamine	4.25	3.58E-05	0.00031	-0.491	1.16E-06
Glutamate	-4.291	3.03E-05	0.00028	0.534	8.45E-08
Valine	-5.8618	2.45E-08	0.000001	0.285	0.007
Leucine + Isoleucine	-5.4052	2.25E-07	0.000005	0.364	0.0005
Alanine	-3.2592	0.0014	0.0066	0.274	0.01
Lysine	-4.3482	2.40E-05	0.0002	0.22	0.039
Ornitine	-2.7277	0.0071	0.0252	0.278	0.009
DG(34:1)	-5.7514	4.23E-08	1.61E-06	0.535	7.92E-08
DG(34:3)	-2.6815	0.0081	0.0279	0.303	0.004
DG(36:2)	-4.353	2.36E-05	0.0002	0.458	7.41E-06
DG(36:3)	-2.4811	0.0141	0.0419	0.340	0.001
DG-O(34:1)	-4.4617	1.50E-05	0.0002	0.450	1.11E-05
TG(44:4)	-3.1339	0.002	0.0091	0.332	0.002
TG(48:1)	-4.4778	1.41E-05	0.0002	0.378	0.0003
TG(48:2)	-3.3397	0.001	0.0054	0.299	0.005
TG(49:1)	-4.2951	2.98E-05	0.0003	0.409	7.68E-05
TG(49:2)	-2.6869	0.008	0.0278	0.316	0.003
TG(50:1)	-5.8907	2.12E-08	1.17E-06	0.486	1.6E-06
TG(50:2)	-5.2666	4.31E-07	9.83E-06	0.419	4.77E-05
TG(50:3)	-3.883	0.0001	0.001	0.335	0.001
TG(50:4)	-2.7846	0.006	0.0223	0.249	0.019
TG(51:2)	-4.1913	4.52E-05	0.0004	0.265	0.012
TG(52:2)	-5.1532	7.28E-07	1.38E-05	0.405	8.96E-05

TG(52:3)	-3.5419	0.0005	0.003	0.310	0.003
TG(54:3)	-4.8426	2.94E-06	5.04E-05	0.287	0.007
TG(56:6)	-5.9019	2.00E-08	1.17E-06	0.314	0.003
PC(32:0)	-3.1296	0.0021	0.0091	0.351	0.001
PC(32:1)	-4.5377	1.09E-05	0.0002	0.235	0.028
PC(34:1)	-2.4994	0.0134	0.041	0.275	0.009
PC(36:3)	-2.4738	0.0144	0.0421	0.213	0.046
PC(39:4)	5.1885	6.19E-07	1.32E-05	-0.413	6.45E-05
PC(40:3)	-2.7559	0.0065	0.0237	0.232	0.029
PC(42:4)	-2.9016	0.0042	0.0166	0.305	0.004
PC(43:6)	-2.7939	0.0058	0.0219	0.397	0.0001
PC(44:1)	-5.8395	2.73E-08	1.17E-06	0.497	8.61E-07
PC(46:2)	-5.7227	4.87E-08	1.65E-06	0.396	0.0001
PC(O-40:4)	3.1778	0.001774	0.0082	-0.217	0.042
PC(O-42:4)	3.0025	0.003098	0.0129	-0.350	0.001
PC(O-42:5)	2.6296	0.009362	0.0308	-0.252	0.018
PC(O-42:6)	3.188	0.001716	0.008	-0.263	0.013
PC(O-44:6)	5.5998	8.86E-08	2.33E-06	-0.374	0.0003
Cer(42:1)	-3.2513	0.001395	0.00672	0.214	0.046
SM(37:1)	-4.0165	8.98E-05	0.0006	0.296	0.005
SM(38:3)	-6.7011	3.16E-10	3.35E-08	0.342	0.001
SM(39:2)	-2.623	0.0095	0.0311	0.270	0.011
SM(43:1)	-4.5274	1.14E-05	0.0002	0.238	0.025
SM(44:1)	-2.4048	0.0173	0.0489	0.266	0.012

a NAFLD served as the reference group.

Abbreviations: HOMA-IR, homeostatic model assessment of insulin resistance; FDR, false discovery rate; DG, diglyceride; TG, triglyceride; PC, phosphatidylcholine; Cer, ceramide; SM, sphingomyelin

3.2.1 Amino acids and biogenic amines

The mean concentrations of plasma BCAAs (leucine, isoleucine, valine), glutamate, tyrosine, alanine, lysine, phenylalanine, and ornithine were higher in subjects with NAFLD than in the controls (Figure 6). Some of these metabolites had a positive correlation with HOMA-IR: glutamate (r^2 = 0.534, P = 8.45E-08), leucine + isoleucine (r^2 = 0.364, P = 0.0005), valine (r^2 = 0.285, P = 0.007), ornithine (r^2 = 0.278, P =0.009), alanine (r^2 = 0.274, P = 0.011) and lysine (r^2 = 0.220, P = 0.007) (Figure 7). Furthermore, glycine and glutamine were lower in NAFLD than in the control (Figure 6) and showed a negative correlation with HOMA-IR (r^2 = −0.027, P = 0.011 and r^2 = −0.491, P < 0.001) (Figure 7). In addition, some biogenic amines, such as kynurenine, asymmetric dimethylarginine, spermidine, taurine and sarcosine, were significantly higher in NAFLD than in the control (FDR < 0.05) (Supplementary Table 1).

3.2.2 Lipids

Most of the lipid metabolites, such as carnitine, glycerophospholipid, sphingolipids and neutral fats, showed significantly higher plasma concentrations in NAFLD than in the control (Supplementary Table 1). The plasma levels of carnitine, octenoylcarnitine, methylmalonyl carnitine and hexanoylcarnitine

were significantly elevated in NAFLD patients compared to controls. Meanwhile, methylmalonyl carnitine had a positive correlation with HOMA-IR (r^2 = 0.284, P = 0.007) (Table 9).

Significant differences in the levels of sphingolipids such as Cer(42:1) and several SMs were significantly higher in NAFLD than in the control (Figure 6, Supplementary table 1). Among the 14 SMs, Cer(42:1), SM(37:1), SM(38:3), SM(39:2), SM(43:1) and SM(44:1) were positively associated with HOMA-IR (P < 0.05) (Table 9, Figure 7).

Although many glycerophospholipids were significantly different between the NAFLD and control groups, the pattern of differences was not consistent. Most of the PCs, such as PC(44:1), PC(46:2), PC(39:4), PC(38:3), PC(32:1), etc. showed higher concentrations in NAFLD subjects than in control subjects (Figure 6). Among these PCs that were higher in subjects with NAFLD, 9 PCs showed a positive correlation with HOMA-IR. However, 12 PCs, PC(39:4), PC(35:2), PC(O-44:6), PC(O-34:3), etc. and 6 LPCs; LPC(18:2), LPC(18:1), LPC(20:4), LPC(17:0), etc. were significantly lower in subjects with NAFLD than in controls (Supplementary table 1). Among these, 6 PCs showed a negative correlation with HOMA-IR (Table 9, Figure 7).

Mean levels of CE(16:2), 7 DGs and 20 TGs were higher in pediatric NAFLD patients than in controls (Supplementary table 1). Among 28 significant neutral fats, 5 DGs and 14 TGs showed a positive correlation with HOMA-IR (Table 9, Figure 7).

Figure 6. Box plots of plasma amino acids, diglycerides, triglycerides, phosphatidylcholines and spingomyelins of NAFLD patients and controls (FDR < 0.05)

Abbreviations: NAFLD, nonalcoholic fatty liver disease; DG, diglyceride, TG, triglyceride; PC, phosphatidylcholine; SM, sphingomyelin; FDR, false discovery rate

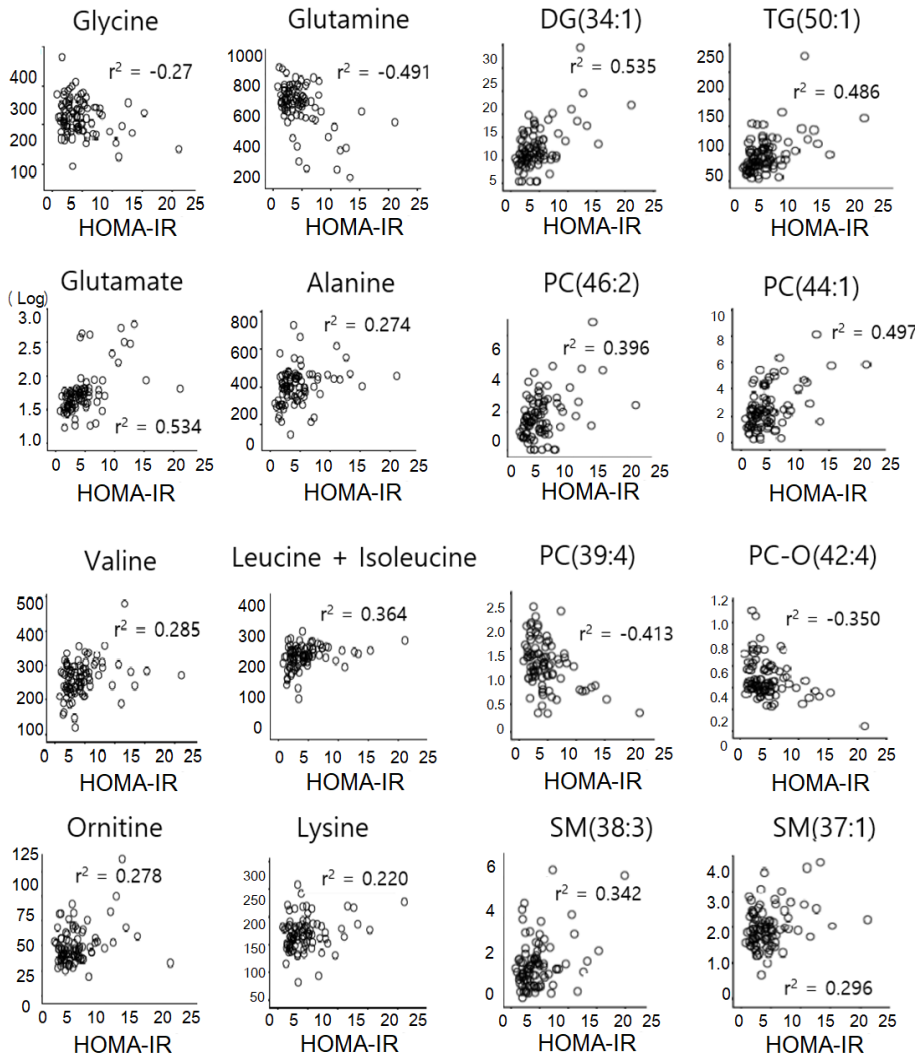

Figure 7. Metabolomic correlations with insulin resistance ($P < 0.05$)

Abbreviations: DG, diglyceride, PC, TG, triglyceride; PC, phosphatidylcholine; SM, sphingomyelin; HOMA-IR, homeostatic model assessment of insulin resistance

3.3 Metabolomic profiles according to genetic variants and NAFLD

Among 101 significant metabolites between subjects with NAFLD and controls, the differences according to genetic variants were further analyzed. These metabolites were described with raw P-values ($P < 0.05$) in addition to FDR (Supplementary Table 2).

Although most of the lipids were higher in NAFLD patients than controls, the risk allele carriers of the *TM6SF2* rs58542926 variant showed significantly lower plasma PC(32:0), PC(36:3) PC(36:4), PC(38:4), PC(40:7), SM(32:1), SM(38:1), SM(39:1) and TG(50:4) after FDR adjustments (Figure 8A). The distributions of these lipid metabolites between risk allele carrier vs noncarrier of *TM6SF2* rs58542926 and NAFLD vs control were inversed. A total of 35 lipids CE(16:1), Cer(42:1), 19 PCs, 3 LPCs, 7 SMs and 4 TGs) were lower in *TM6SF2* variant carriers by raw P-value ($P < 0.05$) (Supplementary Table 2).

Homozygous risk allele carriers of *SAMM50* rs2073080 (TT) and rs3761472 (GG) showed higher concentrations of glutamate, PC(39:4), PC(40:4), PC(42:7) and TG(56:6), while the level of glutamine was lower than that of the wild type according to the raw P-value (Supplementary Table 2). These distributions of

metabolites between *SAMM50* variant genotypes were similar to the distribution between NAFLD vs control (Figure 8B).

Homozygous risk allele carriers of *PNPLA3* rs738409 (GG) showed higher kynurein, PC(40:4), PC(42:7) and TG(56:6) than CC/CT type ($P < 0.05$). However, metabolomic differences among *PNPLA3* and *SAMM50* genotypes were not significant after FDR correction (Supplementary Table 2).

Figure 8. Comparisons of metabolomic distribution according to genetic variants and presence of NAFLD

(A) Metabolomic distributions according to TM6SF2 rs58542929 genotypes and presence of NAFLD

(B) Metabolomic distributions according to SAMM 50 rs2073080, rs3761472 genotypes and presence of NAFLD

Abbreviations: *TM6SF2,* transmembrane 6 superfamily member 2; NAFLD, nonalcoholic fatty liver disease; PC, phosphatidylcholine; TG, triglyceride; SM, sphingomyelin; FDR, false discovery rate

Chapter 4. Discussion

This was the first study to analyze various genetic variants with metabolomic differences simultaneously in a pediatric population. This is the first study of *SAMM50* in Asian children with NAFLD. The impact of genetic variants on the risk of NAFLD was greater in the children with overweight than in all participants. The risk alleles for *PNPLA3, TM6SF2* and *SAMM50* are independent risk factors for NAFLD and independently increase the steatosis grade and fibrosis scores, regardless of sex, age, BMI-Z score and other metabolic factors. In the metabolomic analysis, NAFLD patients showed higher levels of BCAAs, tyrosine, PCs, SMs, DGs, and TGs and lower levels of glycine in both children with overweight and normal weight.

PNPLA3 is the most well-known gene associated with NAFLD in both adult and pediatric populations [17,22,43-49]. This gene belongs to the patatin-like phospholipase domain-containing family of proteins and encodes a protein called adiponutrin. The rs738409 C/G variant of this gene results in the amino acid substitution I148 M and is regarded as a major contributor to the development of fatty liver. These results support those of previous studies showing that *PNPLA3* rs738409 increased the susceptibility to NAFLD in

children and increased the ALT level and grade of hepatic steatosis on sonography, in addition to the incidence of suspected fibrosis. The MAF of *PNPLA3* rs738409 is 0.42–0.5 in the Asian population and 0.483 [19,50] among Hispanic children [22], similar to the MAF in this study (0.49). These MAFs are higher than those in Caucasian (0.324) and African American (0.183) children [22], which could explain the higher prevalence of fatty liver in Asian and Hispanic children (10.2~11.8%) than in black and white children (1.5~8.6%) reported in a previous study [51].

TM6SF2 is known to be involved in lipid metabolism, affecting NAFLD pathogenesis [52–55]. The loss-of-function mutation results in reduced TG secretion in the liver and therefore intrahepatic TG accumulation [56]. Although the risk allele frequency was the lowest, the OR of *TM6SF2* rs58542926 was higher than that of the other variants in this study. In a previous study conducted in obese children, *TM6SSF2* rs58542926 also showed a higher OR (2.58~4.13) than other genetic variants (OR = 1.34~1.66) [20, 57]. In this study, risk allele carriers (CT/TT) showed significantly lower TC and TG levels than wild-type allele carriers. Loss of *TM6SF2* function results in a higher risk of hepatic steatosis but a favorable lipid profile, including low plasma TG and LDL levels [52,58].

For the first time, the association between *SAMM50* and fatty liver in a pediatric population was demonstrated. Some variants s3761472, rs2073080 and rs2143571 of *SAMM50* have recently been found to be associated with NAFLD in GWASs of Asian adults [18,19,23] and Hispanic individuals [59]. *SAMM50* encodes Sam50, which is part of the sorting and assembly machinery required for the assembly of mitochondrial outer membrane β-barrel proteins and plays a crucial role in maintaining mitochondrial function [60].

Mitochondrial dysfunction is important because of insulin resistance, which is a major contributor to the development and progression of NAFLD [61]. In a study of Japanese adults, *SAMM50* showed a strong association with NAFLD and was positively associated with AST and ALT levels and histological traits. Although the metabolic effects of *SAMM50* remain controversial [18,19,23], the TG level was higher in risk allele carriers than in noncarriers in this study. Further research for accurate elucidation of the pathogenesis of NAFLD and the effect of *SAMM50* on lipid metabolism is needed.

Because of the invasiveness of liver biopsy, noninvasive fibrosis scores have been used to predict hepatic fibrosis in children with NAFLD. Three scores, including the PNFS, APRI and FIB-4 score

were used [32]. The APRI and FIB-4 score are useful fibrosis scores to distinguish mild (stage 0-1) and significant fibrosis (stage 2-3) (area under the receiving operating characteristic curve: FIB-4 score, 0.81; APRI, 0.70) [32]. The PNFS is better for identifying advanced fibrosis (stage 3-4) than the APRI and FIB-4 score [33]. In this study, the mean PNFS and APRI were significantly higher in *PNPLA3, TM6SF2 and SAMM50* risk allele carriers. Furthermore, each risk allele independently increased the PNFS, APRI and FIB-4 score even after correcting for sex, BMI-Z score, age, and metabolic factors. The risk alleles of the 4 variants might affect the progression of fibrosis in pediatric NAFLD patients.

A recessive inheritance pattern of *PNPLA3* rs738409, *SAMM50* rs2073080 and *SAMM50* rs3761472 and a dominant inheritance pattern of *TM6SF2* rs58542926 were found for the development of NAFLD and the severity of NAFLD, such as the mean AST, ALT, PNFS and APRI. In addition, dose-dependent increase in AST, ALT, PNFS, APRI and FIB-4 were found according to the number of risk alleles of all 4 variants. A study from Asian obese children also showed a recessive inheritance model of rs738409 with a risk allele frequency of 0.392. They showed a greater OR of the GG genotype (OR 5.84, $P < 0.001$) than the GC genotype (OR 2.96, $P = 0.0008$)

compared to the wild type (CC) of rs738409. They reported that only the GG genotype increased ALT levels compared to the wild type independently, not in the CG type [46]. However, most of the studies of *PNPLA3* demonstrated that the rs738409 G allele was strongly associated with the development and severity of NAFLD in a dose-dependent manner (additive model) with the highest risk of GG type [20,44,62,63]. Previous studies of *SAMM50* analyzed the genetic risk as an additive model [18,19,59]. Hence, this study is the first demonstration of a recessive pattern of *SAMM50* rs2073080 and rs3761472. Because of the lower risk allele frequency of *TM6SF2* rs58542926, it reveals the dominant effects of the risk allele (T). These dominant effects on NAFLD development and severity were also replicated in a multiethnic group study and several other pediatric studies [20,52,54].

In addition to genetic effects, male sex was a significant risk factor for NAFLD in this study, which is consistent with the findings of several studies in pediatric and adult populations [64,65]. One study of NAFLD in obese children and adolescents indicated that the prevalence was highest in postpubertal boys (51.2%) and lowest in postpubertal girls (12.2%), suggesting a sex-specific role of sex steroids in the development of NAFLD [64]. The fasting insulin level

was an independent risk factor for the development of NAFLD in the overweight population. It is well known that insulin resistance is one of the most important factors in the development of NAFLD and its progression to NASH [8,66]

This study supports these synergetic effects of *PNPLA3* rs738409, *TM6SF2* rs58542926, and *SAMM50* rs2073080 and rs3761472 variants on NAFLD severity. As the genetic risk score calculated by the sum of risk alleles increased, AST, ALT, ALT, GGT and fibrosis scores, such as PNFS, APRI and FIB4, also increased independently. This report is the first to evaluate the synergetic effects of variants of *SAMM50* along with *PNPLA3* and *TM6SF2,* especially in children. A previous study also showed these synergetic effects of *TM6SF2* rs58542926 and *PNPLA3* rs738409 in nonalcoholic steatohepatitis and significant fibrosis [63]. Another pediatric study conducted an analysis of 11 genetic variants, and *TM6SF2* rs58542926, *PNPLA3* rs738409 and glucokinase regulatory protein rs1260326 remained independently associated with NAFLD. They also developed weighted genetic risk scores based on 11 genetic variants with clinical risk factors with high accuracy to predict NAFLD [57].

A total of 101 metabolites showed significantly different plasma levels between subjects with NAFLD and controls. It is well known

that BCAAs are significantly higher in NAFLD patients and that circulating BCCAs are correlated with insulin resistance [24,40].

BCAAs are also associated with progression to simple steatosis to NASH or hepatic fat contents. Adipose tissue dysfunction might have a key role in the development of NAFLD because NAFLD patients showed down-regulation of BCAA catabolism in adipose tissue and mitochondrial energy metabolism, resulting in increased serum BCAA concentrations. The rate of BCAA metabolism was negatively associated with plasma BCAA concentrations, liver fat content, and indexes of insulin resistance [67],[38]. In this study, significantly higher levels of BCAAs (isoleucine, leucine, valine), hydrophobic amino acids (alanine), glutamate, ornitine and lysine were also reported in NAFLD patients than in controls. A positive association between BCAAs, glutamate, alanine and HOMA-IR was observed which supports previous results that insulin resistance has a positive association with these metabolites.

Lower levels of glycine, glutamine and serine in subjects with NAFLD than in controls, and a negative association of glycine and glutamine with HOMA-IR were found in this study. In a previous study, glycine, serine and glutamine were known to have a negative association with insulin resistance, and a higher glutamine-to-

glutamate ratio was associated with a lower risk of diabetes [68]. Furthermore, the glutamate/[serine + glycine] index could be used as an index of hepatic insulin resistance and inflammation [42]. A previous study suggested that subjects with hepatic steatosis have reduced de novo synthesis of glutathione due to limited availability of glycine and serine substrates. With this, serine (a precursor to glycine) supplementation was proposed for NAFLD treatment [69].

This study reported several lipids, such as acylcarnitines, PCs, SMs, CEs, DEs and TGs, that were higher in the NAFLD group than in the control group. However, some circulating PCs and LPCs were lower in NAFLD patients than in controls. Similar to these results, previous studies on hepatic or plasma glycerophospholipids and sphigolipids are still inconsistent [41]. PCs are the major components of glycerophospholipids in mammals, which are important components of structural cellular membranes, and only phospholipid molecules can regulate the assembly and secretion of lipoproteins. Up to 30% of hepatic PC comes from the conversion of phosphatidylethanolamine (PE) to PC by the enzyme phosphatidylethanolamine N-methyltransferase (PEMT) [70]. However, the pathophysiology is not well understood since low hepatic PC levels and an altered hepatic PC/PE ratio are known to

have major implications in the development of NAFLD, and loss-of-function mutation of the *PEMT* gene might confer susceptibility to NAFLD [70,71]. The possible pathophysiology of NAFLD associated with PCs is synthesis impairment of PC in hepatocytes. Low hepatic PCs could decrease very low-density lipoprotein (VLDL) secretion from the liver and result in hepatic TG accumulation [72]. In terms of circulating lipid metabolites, the total contents of serum PC and SM significantly increased in NAFLD patients compared with controls [73,] which was similar to these results.

An association between phospholipids and HOMA-IR was reported in this study. Interestingly, PCs that were higher in NAFLD patients showed a positive association with HOMA-IR, while PCs that were lower in NAFLD patients had a negative association with HOMA-IR. Although more precise functional studies are still required, phospholipids are related to several insulin actions. Phospholipids regulate insulin secretion from the pancreas, mediate insulin action on adipocytes and skeletal muscle, modulate gene expression related to glucose uptake and control mitochondrial function [74]. Since the relationship between insulin resistance and phospholipids is very complex, phospholipids could be a promising target for controlling the physiological or pathophysiological processes

affecting insulin sensitivity [74].

LPC is a biologically active lipid generated from PC and is considered an important mediator of hepatic lipotoxicity[41]. In this study, several LPCs were significantly lower in NAFLD patients. Recent studies also reported that plasma LPC species were decreased in patients with NASH or that LPC was diminished in patients with NAFLD [75,76]. Although these findings seem to be related to TG, HDL metabolism and insulin sensitivity, the exact pathophysiology needs to be investigated further [70,75,77].

In this study, several SMs, Cer(42:1), showed significantly higher plasma levels in NAFLD subjects than in controls, and 5 SMs and Cer(42:1) had a positive association with HOMA-IR. The plasma levels of SM have been previously reported to correlate with BMI, and ceramide is a well-known contributing factor to insulin resistance [41,78]. Although many studies have shown different results, they have provided some evidence that sphingolipid metabolism is related to NAFLD development and severity [76,79]. Several previous studies support these results that the serum levels of some SM and ceramide species were higher in patients with NAFLD than in healthy controls [73,80,81]. Regarding insulin resistance, fasting plasma insulin and HOMA-IR were positively associated with SM contents

of erythrocyte and adipocyte membranes in obese women [82,83].

The levels of several DGs and TGs were significantly higher in NAFLD patients than in controls and positively associated with HOMA-IR in this study. A previous study reported that NAFLD is characterized by both TG and free cholesterol accumulation without any corresponding increase in CE [39]. They reported a stepwise increase in the mean TG (product)/DG (precursor) ratio from normal livers to simple steatosis to NASH [39]. The same study group found that plasma DG and TG were significantly increased in subjects with NAFLD compared with controls, which is the same as our results [84]. The functional measure of enzyme activity revealed a marked step up from DG to TG in NAFLD, which supports that the enzyme diacylglycerol acyl transferase plays an important role in the development of hepatic steatosis [39]. The accumulation of specific lipid metabolites (DGs and/or Cers) in liver and skeletal muscle is a common final pathway leading to impaired insulin signaling and insulin resistance [86].

Insulin resistance is a complex metabolic disorder related to ectopic lipid accumulation [85]. Among 101 significant metabolites, 49 metabolites (1 AC, 8 amino acids, 15 PCs, 5 DGs, 14 TGs, 6 SMs) had significant correlations with HOMA-IR. Correlation directions

between HOMA-IR and metabolites were the same as differences between NAFLD and the control. Higher metabolites in NAFLD were positively correlated with HOMA-IR, and lower metabolites in NAFLD were negatively correlated with HOMA-IR. This result suggests that the metabolic difference between NAFLD patients and controls could be partially explained by insulin resistance.

This was the first study to analyze numerous metabolites with several genetic variants simultaneously between NAFLD and controls in children. A total of 101 metabolites that showed significant differences between NAFLD and the control were further analyzed by genetic variants to evaluate their differences by each genetic variant. *TM6SF2* rs58542926-T was strongly associated with NAFLD development and severity, but the distribution of metabolites with/without NAFLD and with/without the T allele showed the opposite direction, especially in lipid profiles such as 5 PCs, TG(50:4), and 3 SMs, which was similar to a previous study [87]. *TM6SF2* is known to be associated with VDLD secretion from the liver; therefore, NAFLD patients with the rs58542926-T allele showed reduced concentrations of circulating lipids, unlike other NAFLD patients with obesity [52,53]. Because PCs are the only necessary phospholipids to assemble VDLD particles, PC deficiency

increases intrahepatic degradation of VLDL particles and thereby reduces their secretion [88,89]. The mechanism leading to reduced secretion of VLDL particles in *TM6SF2* rs58542926-T carriers seems to be caused by deficiency in polyunsaturated phosphatidylcholines and increased hepatic TG [88]. These metabolomic results of *TM6SF2* rs58542926 are compatible with previous lipidomic studies that variant carriers showed metabolic silencing, although they pose a great risk of NAFLD [87,88,90].

These results also suggested that several metabolites may also be affected by *PNPLA3* rs738409, *SAMM50* rs2073080, and rs3761472 variants, which showed the same patterns as NAFLD presence by raw p-value (Supplementary Table 2); however, this is a very preliminary result that needs further study. A previous study showed that *PNPLA3* has no effect on circulating metabolites, and another suggested that it is associated with glycerophospholipids and retinol [87,91]. Only one study reported that *SAMM50* was associated with only one metabolite [91]. However, only a few genomic and metabolomic studies have been conducted simultaneously, and additional evidence needs to be accumulated to confirm the metabolomic effects of these genetic variants.

To our knowledge, this was the first study that analyzed several

genetic variants and various metabolites. Very few studies have been conducted on various genetic variants and associated metabolomic differences, and none have been conducted in pediatric populations. The effects of genetic variants and metabolomic differences for NAFLD were analyzed in both overweight and normal weight children, and the genetic effect was greater in overweight children than in normal weight children. A simultaneous analysis of several variants and synergetic effects related to NAFLD were conducted in a pediatric population. Both genotype and additive models were analyzed among several genetic variants. All genetic variants had additive effects on the development and severity of pediatric NAFLD. This was the first report of the association of *SAMM50* and pediatric NAFLD in addition to the recessive effect of *SAMM50*. Carriers of the *TM6SF2* variant showed significantly lower circulating lipids while they had a higher risk of NAFLD than those with wild-type. A possible mechanism related to metabolic differences between NAFLD patients and controls might be insulin resistance.

This study also has some limitations. First, liver biopsy was not performed to evaluate the degree of steatosis and fibrosis in patients with NAFLD. Liver biopsy is usually performed to evaluate

NAFLD patients with an increased risk of NASH and/or advanced fibrosis [49]. Second, this study did not investigate lifestyle factors, such as diet and exercise.

Chapter 5. Conclusion

Risk alleles of *PNPLA3 rs738409*, *TM6SF2* rs58542926, *SAMM50* rs2073080 and rs37361472 were independent risk factors for NAFLD in this study. These variants also independently increased the severity of NAFLD, regardless of sex, age, BMI-Z score and other metabolic factors. The effects of genetic variants on the development of NAFLD were greater in the overweight subgroup than in all participants. As the genetic risk score increases, AST, ALT and fibrosis scores increase independently, suggesting the synergetic effects of the four genetic variants. Male sex, fasting insulin and triglyceride levels were also independent risk factors for NAFLD in children.

Metabolomic differences between NAFLD patients and controls were reported in the pediatric population. A total of 101 metabolites, such as amino acids, PCs, SMs and neutral fats, showed significant differences between subjects with NAFLD and controls. Among them, 49 metabolites were associated with HOMA-IR, supporting that metabolomic differences in NAFLD are associated with insulin resistance.

Some of these metabolites were significantly correlated with HOMA-IR with the same pattern. Carriers of the *TM6SF2* variant

showed significantly lower circulating lipids than wild-type carriers, and they were in the opposite direction to NAFLD patients.

PNPLA3 rs738409, *SAMM50* rs2073080, *SAMM50* rs3761472 variants, some amino acids and lipids play an important role in the development of pediatric NAFLD. This study will provide the basis of precision medicine for pediatric NAFLD.

Chapter 6. Abbreviations

NAFLD, nonalcoholic fatty liver disease

GWAS, genome-wide association study

PNPLA3, phospholipase-containing domain 3

TM6SF2, transmembrane 6 superfamily member 2

ALT, alanine aminotransferase

BCAA, branched-chain amino acid

NASH, nonalcoholic steatohepatitis

BMI-Z score, body mass index Z score

AST, aspartate transaminase

ALP, alkaline phosphatase

GGT, gamma glutamyl transferase

CAP, controlled attenuation parameter

AC, abdominal circumference

SBP, systolic blood pressure

DBP, diastolic blood pressure

FBS, fasting blood sugar

TC, total cholesterol

TG, triglyceride

LDL-C, low-density lipoprotein cholesterol

HDL-C, high-density lipoprotein cholesterol

HOMA-IR, homeostatic model assessment of insulin resistance

PNFS, pediatric NAFLD fibrosis score

APRI, AST/platelet ratio index

FIB-4, fibrosis-4

MAF, minor allele frequency

DG, diglyceride

LPC, lysophosphatidylcholine

PC, phosphatidylcholine

SM, sphingomyelin

Cer, ceramide

CE, cholesterol ester.

OR, odds ratio

CI, confidence interval

FDR, false discovery rate

PE, phosphatidylethanolamine

PEMT, phosphatidylethanolamine N-methyltransferase

VLDL, very low-density lipoprotein

Chapter 7. Bibliography

1. Mitra S, De A, Chowdhury A. Epidemiology of non-alcoholic and alcoholic fatty liver diseases. *Transl Gastroenterol Hepatol*. 2020;5. doi:10.21037/TGH.2019.09.08
2. Araújo AR, Rosso N, Bedogni G, Tiribelli C, Bellentani S. Global epidemiology of non-alcoholic fatty liver disease/non-alcoholic steatohepatitis: What we need in the future. *Liver Int*. 2018;38:47-51. doi:10.1111/liv.13643
3. Bellentani S, Scaglioni F, Marino M, Bedogni G. Epidemiology of non-alcoholic fatty liver disease. *Dig Dis*. 2010;28(1):155-161. doi:10.1159/000282080
4. Kang Y, Park S, Kim S, Koh H. Estimated prevalence of adolescents with nonalcoholic fatty liver disease in Korea. *J Korean Med Sci*. 2018;33(14). doi:10.3346/jkms.2018.33.e109
5. Kim MJ, Lee KJ. Analysis of the dietary factors associated with suspected pediatric nonalcoholic fatty liver disease and potential liver fibrosis: Korean National Health and Nutrition Examination Survey 2014-2017. *BMC Pediatr*. 2020;20(1). doi:10.1186/s12887-020-02022-y
6. Kim JW, Lee KJ, Yang HR, et al. Prevalence and risk factors of elevated alanine aminotransferase among Korean adolescents: 2001-2014. *BMC Public Health*. 2018;18(1). doi:10.1186/s12889-018-5548-9
7. Marzuillo P, Del Giudice EM, Santoro N. Pediatric non-alcoholic fatty liver disease: New insights and future directions. *World J Hepatol*. 2014;6(4):217-225.

doi:10.4254/wjh.v6.i4.217

8. Buzzetti E, Pinzani M, Tsochatzis EA. The multiple-hit pathogenesis of non-alcoholic fatty liver disease (NAFLD). *Metabolism*. 2016;65(8):1038-1048. doi:10.1016/j.metabol.2015.12.012
9. Miyaaki H, Nakao K. Significance of genetic polymorphisms in patients with nonalcoholic fatty liver disease. *Clin J Gastroenterol*. 2017;10(3):201-207. doi:10.1007/s12328-017-0732-5
10. Sookoian S, Pirola CJ. Genetic predisposition in nonalcoholic fatty liver disease. *Clin Mol Hepatol*. 2017;23(1):1-12. doi:10.3350/cmh.2016.0109
11. Eslam M, Valenti L, Romeo S. Genetics and epigenetics of NAFLD and NASH: Clinical impact. *J Hepatol*. 2018;68(2):268-279. doi:10.1016/j.jhep.2017.09.003
12. Goyal NP, Schwimmer JB. The Genetics of Pediatric Nonalcoholic Fatty Liver Disease. *Clin Liver Dis*. 2018;22(1). doi:10.1016/j.cld.2017.08.002
13. Sookoian S, Pirola CJ. Precision medicine in nonalcoholic fatty liver disease: New therapeutic insights from genetics and systems biology. *Clin Mol Hepatol*. 2020;26(4):461-475. doi:10.3350/cmh.2020.0136
14. Ko JS. New perspectives in pediatric nonalcoholic fatty liver disease: Epidemiology, genetics, diagnosis, and natural history. *Pediatr Gastroenterol Hepatol Nutr*. 2019;22(6):501-510. doi:10.5223/pghn.2019.22.6.501
15. Schwimmer JB, Deutsch R, Kahen T, Lavine JE, Stanley C, Behling C. Prevalence of fatty liver in children and

adolescents. *Pediatrics*. 2006;118(4):1388-1393. doi:10.1542/peds.2006-1212

16. Schwimmer JB, Celedon MA, Lavine JE, et al. Heritability of Nonalcoholic Fatty Liver Disease. *Gastroenterology*. 2009;136(5):1585-1592. doi:10.1053/j.gastro.2009.01.050
17. Romeo S, Kozlitina J, Xing C, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet*. 2008;40(12):1461-1465. doi:10.1038/ng.257
18. Kitamoto T, Kitamoto A, Yoneda M, 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. *Hum Genet*. 2013;132(7):783-792. doi:10.1007/s00439-013-1294-3
19. Chung GE, Lee Y, Yim JY, et al. Genetic Polymorphisms of PNPLA3 and SAMM50 Are Associated with Nonalcoholic Fatty Liver Disease in a Korean Population. *Gut Liver*. 2017;12(3):316-323. doi:10.5009/gnl17306
20. Lin YC, Chang PF, Chang MH, Ni YH. Genetic determinants of hepatic steatosis and serum cytokeratin-18 fragment levels in Taiwanese children. *Liver Int*. 2018;38(7):1300-1307. doi:10.1111/liv.13689
21. Marzuillo P, Miraglia del Giudice E, Santoro N. Pediatric fatty liver disease: Role of ethnicity and genetics. *World J Gastroenterol*. 2014;20(23):7347-7355. doi:10.3748/wjg.v20.i23.7347
22. Santoro N, Kursawe R, D' Adamo E, et al. A common variant

in the patatin-like phospholipase 3 gene (PNPLA3) is associated with fatty liver disease in obese children and adolescents. *Hepatology*. 2010;52(4):1281-1290. doi:10.1002/hep.23832

23. Chen L, Lin Z, Jiang M, et al. Genetic Variants in the SAMM50 Gene Create Susceptibility to Nonalcoholic Fatty Liver Disease in a Chinese Han Population. *Hepat Mon*. 2015;15(10). doi:10.5812/hepatmon.31076
24. Goffredo M, Santoro N, Tric D, et al. A Branched-Chain Amino Acid-Related Metabolic Signature Characterizes Obese Adolescents with Non-Alcoholic Fatty Liver Disease. Published online 2017:1-12. doi:10.3390/nu9070642
25. Kalhan SC, Guo L, Edmison J, et al. Plasma metabolomic profile in nonalcoholic fatty liver disease. *Metabolism*. 2011;60(3):404-413. doi:10.1016/j.metabol.2010.03.006
26. De Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Organ*. 2007;85(9):660-667. doi:10.2471/BLT.07.043497
27. Kim JH, Yun S, Hwang SS, et al. The 2017 Korean national growth charts for children and adolescents: Development, improve ment, and prospects. *Korean J Pediatr*. 2018;61(5):135-149. doi:10.3345/kjp.2018.61.5.135
28. Charatcharoenwitthaya P, Lindor KD. Role of Radiologic Modalities in the Management of Non-alcoholic Steatohepatitis. *Clin Liver Dis*. 2007;11(1):37-54. doi:10.1016/j.cld.2007.02.014
29. Ferraioli G, Monteiro LBS. Ultrasound-based techniques for

the diagnosis of liver steatosis. *World J Gastroenterol*. 2019;25(40):6053-6062. doi:10.3748/wjg.v25.i40.6053

30. Desai NK, Harney S, Raza R, et al. Comparison of controlled attenuation parameter and liver biopsy to assess hepatic steatosis in pediatric patients. *J Pediatr*. 2016;173:160-164.e1. doi:10.1016/j.jpeds.2016.03.021
31. Mansoor S, Collyer E, Alkhouri N. A Comprehensive Review of Noninvasive Liver Fibrosis Tests in Pediatric Nonalcoholic Fatty Liver Disease. *Curr Gastroenterol Rep*. 2015;17(6). doi:10.1007/s11894-015-0447-z
32. Yang HR, Kim HR, Kim MJ, Ko JS, Seo JK. Noninvasive Parameters and hepatic fibrosis scores in children with nonalcoholic fatty liver disease. *World J Gastroenterol*. 2012;18(13):1525-1530. doi:10.3748/wjg.v18.i13.1525
33. Alkhouri N, Mansoor S, Giammaria P, Liccardo D, Lopez R, Nobili V. The development of the Pediatric NAFLD Fibrosis Score (PNFS) to predict the presence of advanced fibrosis in children with nonalcoholic fatty liver disease. *PLoS One*. 2014;9(8):1-7. doi:10.1371/journal.pone.0104558
34. Loaeza-del-castillo A, Paz-pineda F, Oviedo-cárdenas E, Sánchez-ávila F. AST to platelet ratio index (APRI) for the noninvasive evaluation of liver fibrosis. *Ann Hepatol*. 2019;7(4):350-357. doi:10.1016/S1665-2681(19)31836-8
35. Vallet-Pichard A, Mallet V, Nalpas B, et al. FIB-4: An inexpensive and accurate marker of fibrosis in HCV infection. Comparison with liver biopsy and FibroTest. *Hepatology*. 2007;46(1):32-36. doi:10.1002/hep.21669
36. Shannon A, Alkhouri N, Carter-Kent C, et al.

Ultrasonographic quantitative estimation of hepatic steatosis in children With NAFLD. *J Pediatr Gastroenterol Nutr*. 2011;53(2):190-195. doi:10.1097/MPG.0b013e31821b4b61

37. Awai HI, Newton KP, Sirlin CB, Behling C, Schwimmer JB. Evidence and recommendations for imaging liver fat in children, based on systematic review. *Clin Gastroenterol Hepatol*. 2014;12(5):765-773. doi:10.1016/j.cgh.2013.09.050
38. Newgard CB, An J, Bain JR, et al. A Branched-Chain Amino Acid-Related Metabolic Signature that Differentiates Obese and Lean Humans and Contributes to Insulin Resistance. *Cell Metab*. 2009;9(4):311-326. doi:10.1016/j.cmet.2009.02.002
39. Puri P, Baillie RA, Wiest MM, et al. A lipidomic analysis of nonalcoholic fatty liver disease. *Hepatology*. 2007;46(4):1081-1090. doi:10.1002/hep.21763
40. Lake AD, Novak P, Shipkova P, et al. Branched chain amino acid metabolism profiles in progressive human nonalcoholic fatty liver disease. *Amino Acids*. 2015;47(3):603-615. doi:10.1007/s00726-014-1894-9
41. Kartsoli S, Kostara CE, Tsimihodimos V, Bairaktari ET, Christodoulou DK. Lipidomics in non-alcoholic fatty liver disease. *World J Hepatol*. 2020;12(8):436-450. doi:10.4254/wjh.v12.i8.436
42. Cassader M, Rosso C, Buzzigoli E, et al. Altered amino acid concentrations in NAFLD: Impact of obesity and insulin resistance. *Hepatology*. 2017;67(1):145-158. doi:10.1002/hep.29465
43. Romeo S, Sentinelli F, Cambuli VM, et al. The 148M allele of

the PNPLA3 gene is associated with indices of liver damage early in life. *J Hepatol*. 2010;53(2):335-338. doi:10.1016/j.jhep.2010.02.034

44. Valenti L, Al-Serri A, Daly AK, et al. Homozygosity for the patatin-like phospholipase-3/adiponutrin i148m polymorphism influences liver fibrosis in patients with nonalcoholic fatty liver disease. *Hepatology*. 2010;51(4):1209-1217. doi:10.1002/hep.23622
45. Miraglia del Giudice E, Grandone A, Cirillo G, et al. The association of PNPLA3 variants with liver enzymes in childhood obesity is driven by the interaction with abdominal fat. *PLoS One*. 2011;6(11):1-6. doi:10.1371/journal.pone.0027933
46. Lin YC, Chang PF, Hu FC, Yang WS, Chang MH, Ni YH. A common variant in the PNPLA3 gene is a risk factor for non-alcoholic fatty liver disease in obese Taiwanese children. *J Pediatr*. 2011;158(5):740-744. doi:10.1016/j.jpeds.2010.11.016
47. Hyysalo J, Stojkovic I, Kotronen A, et al. Genetic variation in PNPLA3 but not APOC3 influences liver fat in non-alcoholic fatty liver disease. *J Gastroenterol Hepatol*. 2012;27(5):951-956. doi:10.1111/j.1440-1746.2011.07045.x
48. Kawaguchi T, Sumida Y, Umemura A, et al. Genetic polymorphisms of the human PNPLA3 gene are strongly associated with severity of non-alcoholic fatty liver disease in Japanese. *PLoS One*. 2012;7(6). doi:10.1371/journal.pone.0038322
49. Shen J, Wong GLH, Chan HLY, et al. PNPLA3 gene

polymorphism accounts for fatty liver in community subjects without metabolic syndrome. *Aliment Pharmacol Ther*. 2014;39(5):532-539. doi:10.1111/apt.12609

50. Lee SS, Byoun YS, Jeong SH, et al. Role of the PNPLA3 I148M Polymorphism in Nonalcoholic Fatty Liver Disease and Fibrosis in Korea. *Dig Dis Sci*. 2014;59(12):2967-2974. doi:10.1007/s10620-014-3279-z
51. Schwimmer JB, Deutsch R, Kahen T, Lavine JE, Stanley C, Behling C. Prevalence of fatty liver in children and adolescents. *Pediatrics*. 2006;118(4):1388-1393. doi:10.1542/peds.2006-1212
52. Grandone A, Cozzolino D, Marzuillo P, et al. TM6SF2 Glu167Lys polymorphism is associated with low levels of LDL-cholesterol and increased liver injury in obese children. *Pediatr Obes*. 2016;11(2):115-119. doi:10.1111/ijpo.12032
53. Mahdessian H, Taxiarchis A, Popov S, et al. TM6SF2 is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic lipid droplet content. *Proc Natl Acad Sci U S A*. 2014;111(24):8913-8918. doi:10.1073/pnas.1323785111
54. Goffredo M, Caprio S, Feldstein AE, et al. Role of TM6SF2 rs58542926 in the pathogenesis of nonalcoholic pediatric fatty liver disease: A multiethnic study. *Hepatology*. 2016;63(1):117-125. doi:10.1002/hep.28283
55. Akuta N, Kawamura Y, Arase Y, et al. Relationships between Genetic Variations of PNPLA3, TM6SF2 and Histological Features of Nonalcoholic Fatty Liver Disease in Japan. *Gut Liver*. 2015;10(3):1-9. doi:10.5009/gnl15163

56. Kozlitina J, Smagris E, Stender S, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet*. 2014;46(4):352-356. doi:10.1038/ng.2901
57. Zusi C, Mantovani A, Olivieri F, et al. Contribution of a genetic risk score to clinical prediction of hepatic steatosis in obese children and adolescents. *Dig Liver Dis*. 2019;51(11):1586-1592. doi:10.1016/j.dld.2019.05.029
58. Speliotes EK, Yerges-Armstrong LM, Wu J, et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet*. 2011;7(3). doi:10.1371/journal.pgen.1001324
59. Edelman D, Kalia H, Delio M, et al. Genetic analysis of nonalcoholic fatty liver disease within a caribbean-hispanic population. *Mol Genet Genomic Med*. 2015;3(6):558-569. doi:10.1002/mgg3.168
60. Ott C, Ross K, Straub S, et al. Sam50 Functions in Mitochondrial Intermembrane Space Bridging and Biogenesis of Respiratory Complexes. *Mol Cell Biol*. 2012;32(6):1173-1188. doi:10.1128/mcb.06388-11
61. Rector RS, Thyfault JP, Uptergrove GM, et al. Mitochondrial dysfunction precedes insulin resistance and hepatic steatosis and contributes to the natural history of non-alcoholic fatty liver disease in an obese rodent model. *J Hepatol*. 2010;52(5):727-736. doi:10.1016/j.jhep.2009.11.030
62. Valenti L, Alisi A, Galmozzi E, et al. I148M patatin-like phospholipase domain-containing 3 gene variant and severity

of pediatric nonalcoholic fatty liver disease. *Hepatology*. 2010;52(4):1274-1280. doi:10.1002/hep.23823

63. Koo BK, Joo SK, Kim D, et al. Additive effects of PNPLA3 and TM6SF2 on the histological severity of non-alcoholic fatty liver disease. *J Gastroenterol Hepatol*. 2018;33(6):1277-1285. doi:10.1111/jgh.14056

64. Denzer C, Thiere D, Muche R, et al. Gender-specific prevalences of fatty liver in obese children and adolescents: roles of body fat distribution, sex steroids, and insulin resistance. *J Clin Endocrinol Metab*. 2009;94(10):3872-3881. doi:10.1210/jc.2009-1125

65. Ballestri S, Nascimbeni F, Baldelli E, Marrazzo A, Romagnoli D, Lonardo A. NAFLD as a Sexual Dimorphic Disease : Role of Gender and Reproductive Status in the Development and Progression of Nonalcoholic Fatty Liver Disease and Inherent Cardiovascular Risk. *Adv Ther*. 2017;34(6):1291-1326. doi:10.1007/s12325-017-0556-1

66. Utzschneider KM, Kahn SE. Review: The role of insulin resistance in nonalcoholic fatty liver disease. *J Clin Endocrinol Metab*. 2006;91(12):4753-4761. doi:10.1210/jc.2006-0587

67. Cheng S, Wiklund P, Autio R, et al. Adipose tissue dysfunction and altered systemic amino acid metabolism are associated with non-alcoholic fatty liver disease. *PLoS One*. 2015;10(10). doi:10.1371/journal.pone.0138889

68. Cheng S, Rhee EP, Larson MG, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. *Circulation*. 2012;125(18):2222-2231.

doi:10.1161/CIRCULATIONAHA.111.067827

69. Mardinoglu A, Bjornson E, Zhang C, et al. Personal model-assisted identification of NAD + and glutathione metabolism as intervention target in NAFLD . *Mol Syst Biol*. 2017;13(3):916. doi:10.15252/msb.20167422

70. O' Doherty PJA. Regulation of ethanolamine and choline phosphatide biosynthesis in isolated rat intestinal villus cells. *Can J Biochem*. 1980;58(7):527–533. doi:10.1139/o80-071

71. Kartsoli S, Kostara CE, Tsimihodimos V, Bairaktari ET, Christodoulou DK. Lipidomics in non-alcoholic fatty liver disease Conflict-of-interest statement. *World J Hepatol*. 2020;12(8):436–450. doi:10.4254/wjh.v12.i8.436

72. Waite KA, Cabilio NR, Vance DE. Choline deficiency-induced liver damage is reversible in Pemt-/- mice. *J Nutr*. 2002;132(1):68–71. doi:10.1093/jn/132.1.68

73. Tiwari-Heckler S, Gan-Schreier H, Stremmel W, Chamulitrat W, Pathil A. Circulating Phospholipid Patterns in NAFLD Patients Associated with a Combination of Metabolic Risk Factors. *Nutrients*. 2018;10(5):649. doi:10.3390/nu10050649

74. Chang W, Hatch GM, Wang Y, Yu F, Wang M. The relationship between phospholipids and insulin resistance: From clinical to experimental studies. *J Cell Mol Med*. 2019;23(2):702–710. doi:10.1111/jcmm.13984

75. Orešič M, Hyötyläinen T, Kotronen A, et al. Prediction of non-alcoholic fatty-liver disease and liver fat content by serum molecular lipids. doi:10.1007/s00125-013-2981-2

76. Zhou Y, Orešič M, Leivonen M, et al. Noninvasive Detection of

Nonalcoholic Steatohepatitis Using Clinical Markers and Circulating Levels of Lipids and Metabolites. *Clin Gastroenterol Hepatol*. 2016;14(10):1463–1472.e6. doi:10.1016/j.cgh.2016.05.046

77. Lehmann R, Franken H, Dammeier S, et al. Circulating Lysophosphatidylcholines Are Markers of a Metabolically Benign Nonalcoholic Fatty Liver. Published online 2013. doi:10.2337/dc12–1760

78. Meikle PJ, Wong G, Barlow CK, et al. Plasma Lipid Profiling Shows Similar Associations with Prediabetes and Type 2 Diabetes. Maya–Monteiro CM, ed. *PLoS One*. 2013;8(9):e74341. doi:10.1371/journal.pone.0074341

79. Chiappini F, Coilly A, Kadar H, et al. Metabolism dysregulation induces a specific lipid signature of nonalcoholic steatohepatitis in patients. *Sci Rep*. 2017;7. doi:10.1038/srep46658

80. Barr J, Vázquez–Chantada M, Alonso C, et al. Liquid chromatography – Mass spectrometry–based parallel metabolic profiling of human and mouse model serum reveals putative biomarkers associated with the progression of nonalcoholic fatty liver disease. *J Proteome Res*. 2010;9(9):4501–4512. doi:10.1021/pr1002593

81. Anjani K, Lhomme M, Sokolovska N, et al. Circulating phospholipid profiling identifies portal contribution to NASH signature in obesity. *J Hepatol*. 2015;62(4):905–912. doi:10.1016/j.jhep.2014.11.002

82. Zeghari N, Vidal H, Younsi M, Ziegler O, Drouin P, Donner M. Adipocyte membrane phospholipids and PPAR–γ expression

in obese women: Relationship to hyperinsulinemia. *Am J Physiol - Endocrinol Metab*. 2000;279(4 42-4). doi:10.1152/ajpendo.2000.279.4.e736

83. Younsi M, Quilliot D, Al-Makdissy N, et al. Erythrocyte membrane phospholipid composition is related to hyperinsulinemia in obese nondiabetic women: Effects of weight loss. *Metabolism*. 2002;51(10):1261-1268. doi:10.1053/meta.2002.35184

84. Puri P, Wiest MM, Cheung O, et al. The plasma lipidomic signature of nonalcoholic steatohepatitis. *Hepatology*. 2009;50(6):1827-1838. doi:10.1002/hep.23229

85. Samuel VT, Shulman GI. Mechanisms for insulin resistance: Common threads and missing links. *Cell*. 2012;148(5):852-871. doi:10.1016/j.cell.2012.02.017

86. Finck BN, Hall AM. Does diacylglycerol accumulation in fatty liver disease cause hepatic insulin resistance? *Biomed Res Int*. 2015;2015. doi:10.1155/2015/104132

87. Sliz E, Sebert S, Wü Rtz P, et al. NAFLD risk alleles in PNPLA3, TM6SF2, GCKR and LYPLAL1 show divergent metabolic effects. *Hum Mol Genet*. 2018;27(12):2214-2223. doi:10.1093/hmg/ddy124

88. Luukkonen PK, Zhou Y, Nidhina Haridas PA, et al. Impaired hepatic lipid synthesis from polyunsaturated fatty acids in TM6SF2 E167K variant carriers with NAFLD. *J Hepatol*. 2017;67(1):128-136. doi:10.1016/j.jhep.2017.02.014

89. Cole LK, Vance JE, Vance DE. Phosphatidylcholine biosynthesis and lipoprotein metabolism. *Biochim Biophys Acta - Mol Cell Biol Lipids*. 2012;1821(5):754-761.

doi:10.1016/j.bbalip.2011.09.009

90. Zhou Y, Llauradó G, Orešič M, Hyötyläinen T, Orho-Melander M, Yki-Järvinen H. Circulating triacylglycerol signatures and insulin sensitivity in NAFLD associated with the E167K variant in TM6SF2. *J Hepatol*. 2015;62(3):657-663. doi:10.1016/j.jhep.2014.10.010

91. Silva LF, Vangipurapu J, Kuulasmaa T, Laakso M. Phenotypic characterization of seven genetic variants associated with non-alcoholic fatty liver disease. *Metabolism*. 2020;104:154079. doi:10.1016/j.metabol.2019.12.025

Chapter 8. Supplementary tables

Supplementary table 1. A total of 101 metabolites were significantly different between NAFLD patients and controls in the pediatric population

Metabolites	t value[a]	P-value	FDR	Metabolites	t value[a]	P-value	FDR
Amino acids (11)				Acylarnitines (4)			
Glycine	6.66	3.92E-10	3.35E-08	AC(0:0)	-3.88	0.000149	0.001
Tyrosine	-6.66	3.83E-10	3.35E-08	AC(8:1)	-2.96	0.00348	0.0142
Valine	-5.86	2.45E-08	1.17E-06	AC(4:0-DC)	-2.6	0.009755	0.0315
Leucine+ Isoleucine	-5.40	2.25E-07	5.49E-06	AC(6:0)	-2.40	0.017741	0.0497
Lysine	-4.34	2.40E-05	0.0002	Sphingolipids (14)			
Glutamine	4.25	3.58E-05	0.00031	Cer(42:1)	-3.25	0.0014	0.00671
Glutamate	-4.29	3.03E-05	0.00028	SM(38:3)	-6.70	3.16E-10	3.35E-08
Phenylalanine	-3.64	0.0004	0.0022	SM(36:1)	-5.69	5.64E-08	1.65E-06
Alanine	-3.26	0.0014	0.0066	SM(41:1)	-5.16	7.20E-07	1.38E-05
Ornitine	-2.73	0.0071	0.0252	SM(43:1)	-4.53	1.14E-05	0.0002
Serine	2.51	0.0131	0.0408	SM(38:1)	-4.48	1.41E-05	0.0002
Phosphatidylcholines (39)				SM(42:1)	-4.44	1.64E-05	0.0002
PC(44:1)	-5.84	2.73E-08	1.17E-06	SM(37:1)	-4.02	8.98E-05	0.0006
PC(46:2)	-5.72	4.87E-08	1.65E-06	SM(39:1)	-3.82	0.0002	0.0012
PC(39:4)	5.18	6.19E-07	1.32E-05	SM(36:2)	-3.74	0.0003	0.0016
PC(38:3)	-4.69	5.83E-06	8.67E-05	SM(39:2)	-2.62	0.0095	0.0311

PC(32:1)	-4.54	1.09E-05	0.0002	SM(40:4)	-2.59	0.0105	0.0331
PC(32:6)	-4.47	1.45E-05	0.0002	SM(32:1)	-2.47	0.0144	0.0421
PC(34:4)	-4.25	3.56E-05	0.0003	SM(44:1)	-2.40	0.0173	0.0489
PC(40:6)	-4.22	3.95E-05	0.00033	Natural lipids (28)			
PC(36:1)	-3.82	0.0002	0.0012	TG(48:1)	-4.48	1.41E-05	0.0002
PC(38:4)	-3.46	0.0007	0.0037	TG(49:1)	-4.30	2.98E-05	0.0003
PC(35:2)	3.24	0.0014	0.0068	TG(51:2)	-4.19	4.52E-05	0.0004
PC(32:0)	-3.13	0.0021	0.0091	TG(50:3)	-3.88	0.0001	0.001
PC(42:4)	-2.90	0.0042	0.0166	TG(52:3)	-3.54	0.0005	0.003
PC(42:5)	-2.90	0.0042	0.0166	TG(54:4)	-3.42	0.0008	0.0043
PC(43:6)	-2.79	0.0058	0.0219	TG(48:2)	-3.34	0.001	0.0054
PC(40:2)	-2.77	0.0062	0.0228	TG(54:6)	-3.14	0.002	0.0091
PC(40:3)	-2.76	0.0065	0.0237	TG(44:4)	-3.134	0.002	0.0091
PC(42:7)	-2.75	0.0067	0.0241	TG(54:5)	-3.01	0.0023	0.0099
PC(36:4)	-2.63	0.0092	0.0308	TG(50:4)	-2.78	0.006	0.0223
PC(40:4)	-2.51	0.0128	0.0403	TG(49:2)	-2.69	0.008	0.0278
PC(34:1)	-2.50	0.0134	0.041	TG(56:8)	-2.67	0.0084	0.0286
PC(36:3)	-2.47	0.0144	0.0421	TG(53:5)	-2.63	0.0093	0.0308
PC(40:7)	-2.42	0.0166	0.0472	CE(16:1)	-3.52	0.0006	0.0031
PC(O-44:6)	5.60	8.86E-08	2.33E-06	DG(34:1)	-5.75	4.23E-08	1.61E-06
PC(O-34:3)	4.50	1.26E-05	0.0002	DG-O(34:1)	-4.46	1.50E-05	0.0002
PC(O-30:0)	4.2	4.37E-05	0.0004	DG(36:2)	-4.35	2.36E-05	0.0002
PC(O-36:2)	4.17	4.85E-05	0.0004	DG(41:1)	-4.11	0.0001	0.0005
PC(O-42:6)	3.188	0.001716	0.008	DG(39:0)	-3.49	0.0006	0.0035
PC(O-40:4)	3.1778	0.001774	0.0082	DG(34:3)	-2.68	0.0081	0.0279
PC(O-42:4)	3.0025	0.003098	0.0129	DG(36:3)	-2.48	0.0141	0.0419

PC(O-36:6)	2.8683	0.004669	0.0181	TG(50:1)	-5.89	2.12E-08	1.17E-06
PC(O-42:5)	2.63	0.009362	0.0308	TG(56:6)	-5.90	2.00E-08	1.17E-06
PC(O-34:1)	2.46	0.014905	0.0432	TG(50:2)	-5.27	4.31E-07	9.83E-06
LPC(18:2)	4.77	3.99E-06	6.21E-05	TG(52:2)	-5.15	7.28E-07	1.38E-05
LPC(18:1)	4.40	1.98E-05	0.000211	TG(54:3)	-4.84	2.94E-06	5.04E-05
LPC(20:4)	4.13	5.77E-05	0.000439	TG(56:7)	-4.77	4.00E-06	6.21E-05
LPC(17:0)	3.56	0.000481	0.002885	Biogenic Amines (5)			
LPC(O-16:1)	2.98	0.003223	0.013278	Kynurenine	-4.09	6.67E-05	0.0005
LPC(15:0)	2.48	0.013809	0.041793	Spermidine	-3.38	0.000892	0.0047
				ADMA	-2.94	0.003674	0.0148
				Taurine	-2.82	0.005323	0.0203
				Sarcosine	-2.50	0.013257	0.0408

a NAFLD served as the reference group.

Abbreviations: FDR, false discovery rate; SM, sphingomyelin; PC, phosphatidylcholine; Cer, ceramide; CE, cholesterol ester; DG, diglyceride; TG, triglyceride

Supplementary table 2. Different metabolomic profiles by each genetic variant with raw p values

TM6SF2 rs58542926 (CC vs CT/TT)				*PNPLA3* rs738409 (CC/CG vs GG)			
Metabolites	t value[a]	P-value	FDR	Metabolites	t value[b]	P-value	FDR
PC(32:0)	3.061	0.003	0.043	Kynurenine	-2.407	0.017	0.579
PC(36:3)	3.379	0.001	0.041	PC(40:4)	-2.462	0.015	0.579
PC(36:4)	3.230	0.001	0.041	PC(42:7)	-2.039	0.043	0.741
PC(40:7)	3.203	0.002	0.041	TG(56:6)	-2.734	0.007	0.579
PC(38:4)	2.969	0.003	0.048				
PC(34:1)	2.807	0.006	0.053	*SAMM50* rs2073080 (CC/CT vs TT)			
PC(35:2)	2.650	0.009	0.059	Metabolites	t value[c]	P-value	FDR
PC(42:7)	2.571	0.011	0.070	Glutamine	2.111	0.036	0.611
PC(40:6)	2.490	0.014	0.082	Glutamate	-2.917	0.004	0.407
PC(40:3)	2.396	0.018	0.096	PC(40:4)	-2.285	0.024	0.611
PC(40:4)	2.309	0.022	0.097	PC(39:4)	2.210	0.029	0.611
PC(34:4)	2.266	0.025	0.104	PC(42:7)	-2.175	0.031	0.611
PC(42:4)	2.195	0.030	0.115	TG(56:6)	-2.448	0.015	0.611
PC(38:3)	2.153	0.033	0.122	Cer(42:1)	-1.983	0.049	0.498
PC(40:2)	2.076	0.039	0.124				
PC(32:1)	2.000	0.047	0.136	*SAMM50* rs3761472 (AA/AG vs GG)			
PC(O-30:0)	2.770	0.006	0.053	Metabolites	t value[d]	P-value	FDR
PC(O-34:3)	2.367	0.019	0.096	Glutamine	2.214	0.028	0.475
PC(O-36:6)	2.217	0.028	0.113	Glutamate	-3.182	0.002	0.177
LPC(17:0)	2.726	0.007	0.055	PC(32:6)	-2.001	0.047	0.476
LPC(15:0)	2.350	0.020	0.096	PC(39:4)	2.241	0.026	0.475

LPC(O-16:1)	2.686	0.008	0.058	PC(40:4)	-2.391	0.018	0.475
Cer(42:1)	2.129	0.035	0.122	PC(42:7)	-2.352	0.020	0.475
SM(32:1)	3.066	0.003	0.043	SM(38:3)	-2.150	0.033	0.476
SM(38:1)	2.935	0.004	0.048	DG(41:1)	-1.994	0.048	0.476
SM(39:1)	2.899	0.004	0.048	TG(56:6)	-2.734	0.007	0.475
SM(41:1)	2.320	0.022	0.097				
SM(42:1)	2.120	0.035	0.122				
SM(39:2)	2.080	0.039	0.124				
SM(40:4)	2.061	0.041	0.125				
CE(16:1)	2.782	0.006	0.053				
TG(50:4)	3.228	0.002	0.041				
TG(50:3)	2.349	0.020	0.096				
TG(56:6)	2.111	0.036	0.122				
TG(51:2)	2.000	0.047	0.136				

a. The *TM6SF2* rs58542926 genetic variant has CC, CT and TT genotypes; the CT/TT genotype served as the reference group.

b. The *PNPLA3* rs738409 genetic variant has CC, CG, and GG genotypes; the homozygous variants (GG) served as the reference group.

c. The *SAMM50* rs2073080 genetic variant has CC, CT and TT genotypes; the homozygous variant (TT) genotype served as the reference group.

d. The *SAMM50* rs3716172 genetic variant has AA, AG and GG genotypes; the homozygous variant (GG) genotype served as the reference group.

Abbreviations: *TM6SF2,* transmembrane 6 superfamily member 2;

PNPLA3, phospholipase-containing domain 3;

FDR, false discovery rate; PC, phosphatidylcholine; Cer, ceramide; SM, sphingomyelin; DG, diglyceride; TG, triglyceride

Chapter 9. Abstract in Korean

배경

일부 유전자 변이와 다른 대사체 양상이 비알콜성 지방간질환과 관련되어 있다고 보고되고 있지만 유전자 변이와 대사체 분석을 동시에 시행한 연구는 거의 없으며 소아청소년 집단에서는 시행된 연구가 없다.

목적

본 연구의 목적은 소아 비알콜성 지방간 질환에 미치는 유전적 변이의 영향을 파악하고, 더불어 환자와 대조군의 대사체 차이와 유전자변이와 연관성을 분석하여 질병의 병인을 보다 명확히 규명하고자 함이다.

방법

유전자 분석을 위하여 228명의 소아 환자와 225명의 대조군 (과체중 69, 정상체중 156명)이 등록되었고 그중 105명의 환자와 61명의 대조군 (과체중 118명, 정상체중 48명)을 대사체 분석을 위해 포함하였다. 다음과 같은 *PNPLA3* (rs738409), *TM6SF2* (rs58542926) and *SAMM50* (rs2073080, rs3761472) 4개의 유전자 변이를 분석하였다. 혈장 대사체는 Biocrates AbsoluteIDQ p400 kit와 Thermo Q Exactive Plus orbitrap mass spectrometer를 통해 분석하였다. 모든 참가자는 신체계측을 시행하였고 일반 혈액검사와, 간기능검사를 받았으며, 과체중집단에서는 복부둘레와 혈압, 공복 혈당, 인슐린, 지방분석을 시행하였다. 간 초음파를 통해 지방간 유무와 지방간 단계를 평가하였으며 비침습적 섬유화 점수를 따라 섬유화 정도를 평가하였다. 비알콜성 지방간질환에 영향을 미치

는 유전자 변이와 대사체 차이는 모두 일반소아 집단과 과체중집단에서 분석하였다.

결과

PNPPLA3 (rs738409), *TM6SF2* (rs58542926), *SAMM50* (rs2073080, rs3761472) 변이는 일반 소아집단 (승산비: 1.99~ 3.26, P <0.05)과 과체중소아집단 (승산비: 2.22~ 22.94, P <0.05에서 독립적으로 NAFLD위험을 높였다. 다른 독립적인 위험인자는 체질량지수-Z 점수와, 남성이었으며, 과체중집단에서는 공복 인슐린이 추가적인 위험 인자였다. 이 유전자변이들은 나이, 성별, 체질량지수-Z점수와 관련없이 알라닌 아미노 전이효소와, 초음파상의 지방간 정도, 섬유화 수치를 증가시켰다. 이러한 유전적영향은 전체 소아 집단보다 과체중집단에서 더욱 컸다. 과체중집단과, 정상체중 집단 모두에서 비알콜성 지방간질환 환자는 대조군에 비해 더 높은 혈장 류신, 아이소류신, 발린 같은 branched chain 아미노산과, 글루타메이트, 타이로신, 포스파티딜콜린, 스핑고마이엘린, 다이글리세라이드, 중성지방 농도를 보였으며 이것들중 branched chain 아미노산, 글루타메이트 등은 HOMA-IR과 양성상관관계를 보였다. 글라이신, 글루타민, 세린 농도는 비알콜성 지방간환자에서 낮았으며 글라이신, 글루타민은 HOMA-IR과 음의 상관관계를 보여 환자와 대조군간의 대사체 차이는 인슐린저항성과 연관이 있음을 시사하였다. *TM6SF2* 변이가 있는 환자에서는 변이가 없는 환자보다 더 낮은 혈장 포스파티딜콜린, 스핑고마이엘린, 중성지방 농도를

나타냈으며 이것은 비알콜성지방간 유무와 반대되는 분포였다.

결론

본 연구는 소아청소년 연령에서 최초로 다양한 유전자 변이와 혈장 대사체를 동시에 분석한 연구이다. *PNPLA3, TM6SF2, SAMM50* 은 소아청소년 일반집단과 과체중집단 모두에서 비알콜성 지방간질환의 발생과 중증도에 독립적으로 영향을 미쳤으며 그 영향은 과체중집단에서 컸다. 101개의 대사체가 비알콜성 지방간질환과 대조군 사이에 차이가 있었으며 그중 49개는 인슐린 저항성과 관련이 있었다. *TM6SF2* 변이를 가진 사람은 혈장 지방농도가 더 낮았으며, 이것은 일반적인 NAFLD환자와는 반대되는 양상이었으며 다른 유전자 변이에 따른 대사체 변화는 후속 연구가 필요하겠다

주요어: 비알콜성지방간질환; *Patatin-like phospholipase domain-containing 3; Transmembrane 6 superfamily member 2; Samm50;* 유전체학; 대사체학

학번: 2017-34184

Chapter 10. Acknowledgements

This work was supported by a National Research Foundation of Korea (NRF) grant from the Korean government (MSIT, Ministry of Science and ICT) (No. NRF-2019R1F1A1057588).

Metabolomic experiments and raw data generation were conducted in the Department of Clinical Pharmacology and Therapeutics, Seoul National University College of Medicine and Hospital, Seoul, Korea.