Variants in *MARC1* and *HSD17B13* reduce severity of NAFLD in children, perturb phospholipid metabolism, and suppress fibrotic pathways

Supplementary Methods

Hepatic proteomic analysis

A subset of 70 liver tissue specimens were homogenized under denaturing conditions with a FastPrep in a buffer containing 3 M GdmCL, 5 mM TCEP, 20 mM CAA, and 50 mM Tris pH 8.5, followed by sonication for 15 min and boiled at 95 °C for 15 min. The lysates were digested and purified using the preOmics in-stage tip kit (iST kit 96x, Martinsried, Germany). Samples were eluted sequentially in three fractions using the SDB-RPS-1 and -2 buffers[1] and the elution buffer provided by preOmics for subsequent analysis on a nano LC-MS/MS.

LC−MS/MS was carried out by nanoflow reverse-phase liquid chromatography (Dionex Ultimate 3000, Thermo Scientific, Waltham, MA) coupled online to a Q-Exactive HF Orbitrap mass spectrometer (Thermo Scientific, Waltham, MA). The LC separation was performed using a PicoFrit analytical column (75 µm ID × 55 cm long, 15 µm Tip ID (New Objectives, Woburn, MA) packed in-house with 3-µm C18 resin (Reprosil-AQ Pur, Dr. Maisch, Ammerbuch-Entringen, Germany), as reported previously (Gielisch and Meierhofer 2015). Peptides were eluted using a gradient from 3.8 to 50 % solvent B in solvent A over 121 min at 266 nL per minute flow rate. Solvent A was 0.1 % formic acid and solvent B was 79.9 % acetonitrile, 20 % H_2O, 0.1 % formic acid. Nanoelectrospray was generated by applying 3.5 kV. A cycle of one full Fourier transformation scan mass spectrum (300−1750 m/z, resolution of 60,000 at m/z 200, AGC target $1e^6$) was followed by 12 data-dependent MS/MS scans scans (resolution of 30,000, AGC target $5e^5$) with a normalized collision energy of 25 eV. In order to avoid repeated sequencing of the same peptides, a dynamic exclusion window of 30 seconds was used. Additionally, only peptide charge states between two and eight were sequenced.

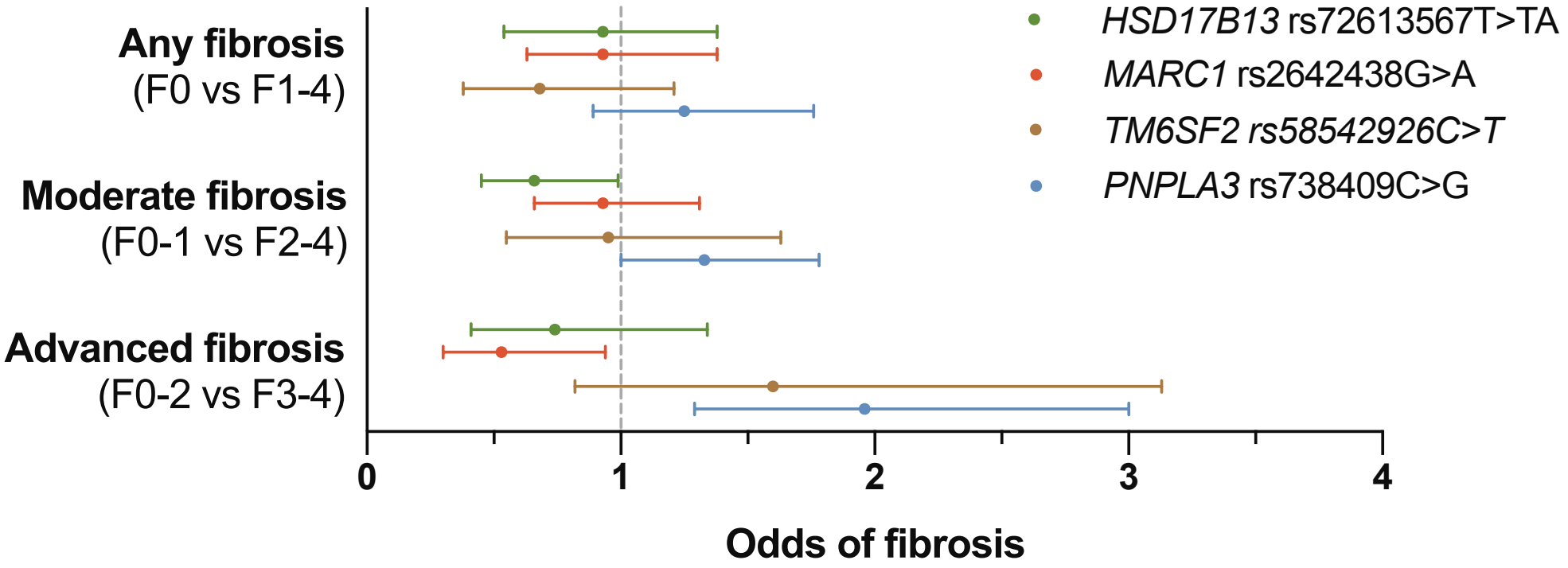
Raw MS data were processed with MaxQuant software (v1.6.0.1)[2] and searched against the human proteome database UniProtKB with 70,941 entries, released in 01/2017. Parameters of MaxQuant database searching were as follows: a false discovery rate (FDR) of 0.01 for proteins and peptides, a minimum peptide length of 7 amino acids, a mass tolerance of 4.5 ppm for precursor and 20 ppm for fragment ions. A maximum of two missed cleavages was allowed for the tryptic digest. Cysteine carbamidomethylation was set as fixed modification, while N-terminal acetylation and methionine oxidation were set as variable modifications.
Proteomics data were logarithmically transformed and then standardised to mean = 0 and standard deviation = 1. Only proteins with detectable abundance of >70% in each group were included in analyses, imputation from normal distribution was used for missing (or zero) values. Association between individual proteins and genotype was assessed according to an additive genetic model using linear regression, with proteins as dependent variable and genotype (coded as 0, 1, or 2 effect alleles), age, and sex and independent variables. Beta regression coefficient therefore represents the change in log-standard deviations of protein abundance per effect allele, adjusted for age and sex.

In addition, a two-sample t-test was performed comparing homozygous wild-type against homozygous mutants (e.g. rs72613567 T/T vs. TA/TA. Multiple test correction was done by Benjamini-Hochberg with an FDR of 0.05 by using Perseus (v1.6.0.2)[3].

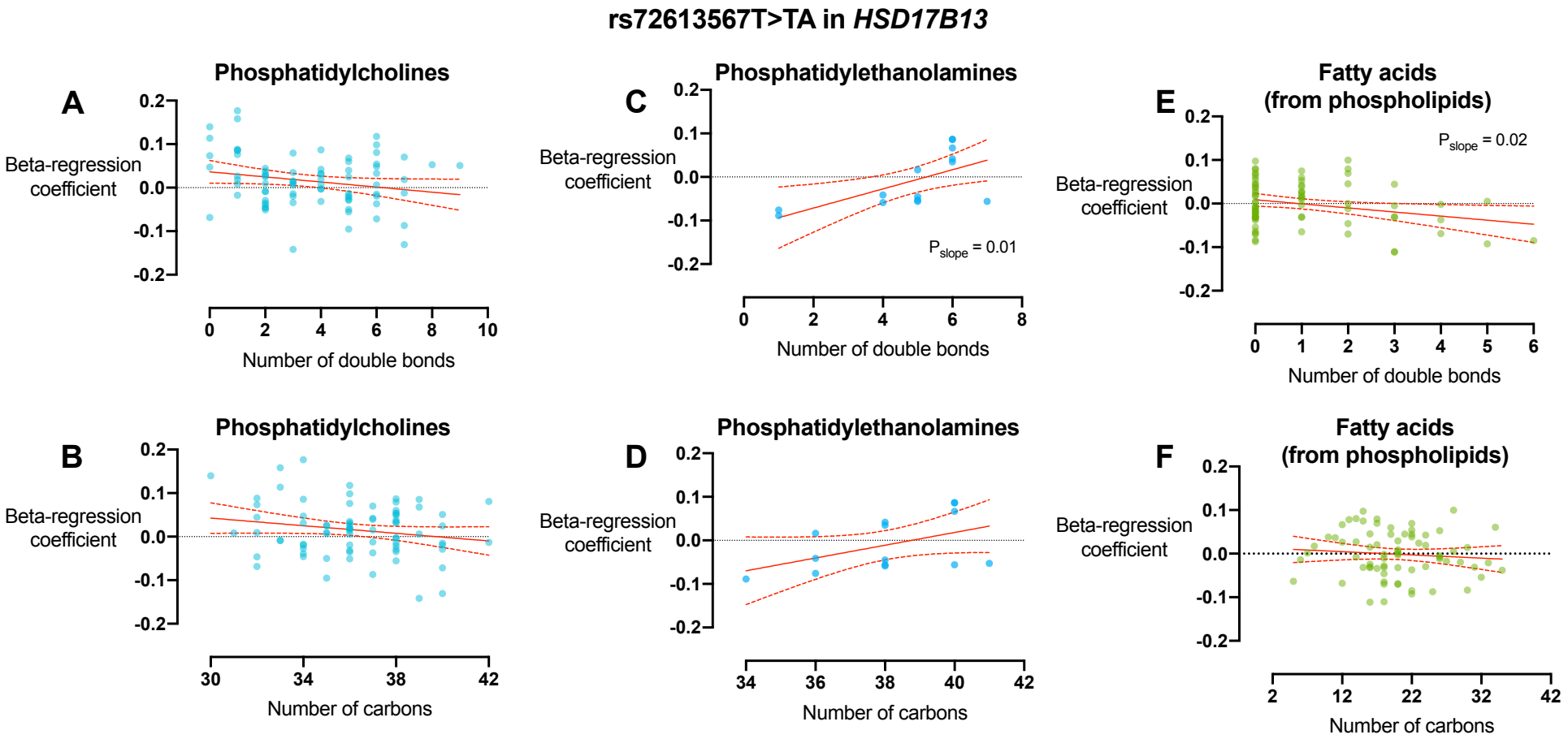
For comprehensive proteome data analyses, gene set enrichment analysis (GSEA, v4.0.0)[4] was applied in order to see if *a priori* defined sets of proteins show statistically significant, concordant differences according to genotype. GSEA was performed using the preranked gene list derived from linear regression analysis. GSEA default settings were used. As recommended for hypothesis discovery, cut off for significantly regulated pathways was set to a *P*-value < 0.05 and an FDR Q-value < 0.25. GSEA results were then passed to Cytoscape[5,6] for plotting of enrichment maps by clustering similar gene sets, using Enrichment Map[7]. Groups of gene sets were annotated using AutoAnnotate[8] with WordCloud[9].

Volcano blot was performed using an FDR Q-value of 0.05 and an artificial within groups variance (s0) of 0.1.

Supplementary Figures

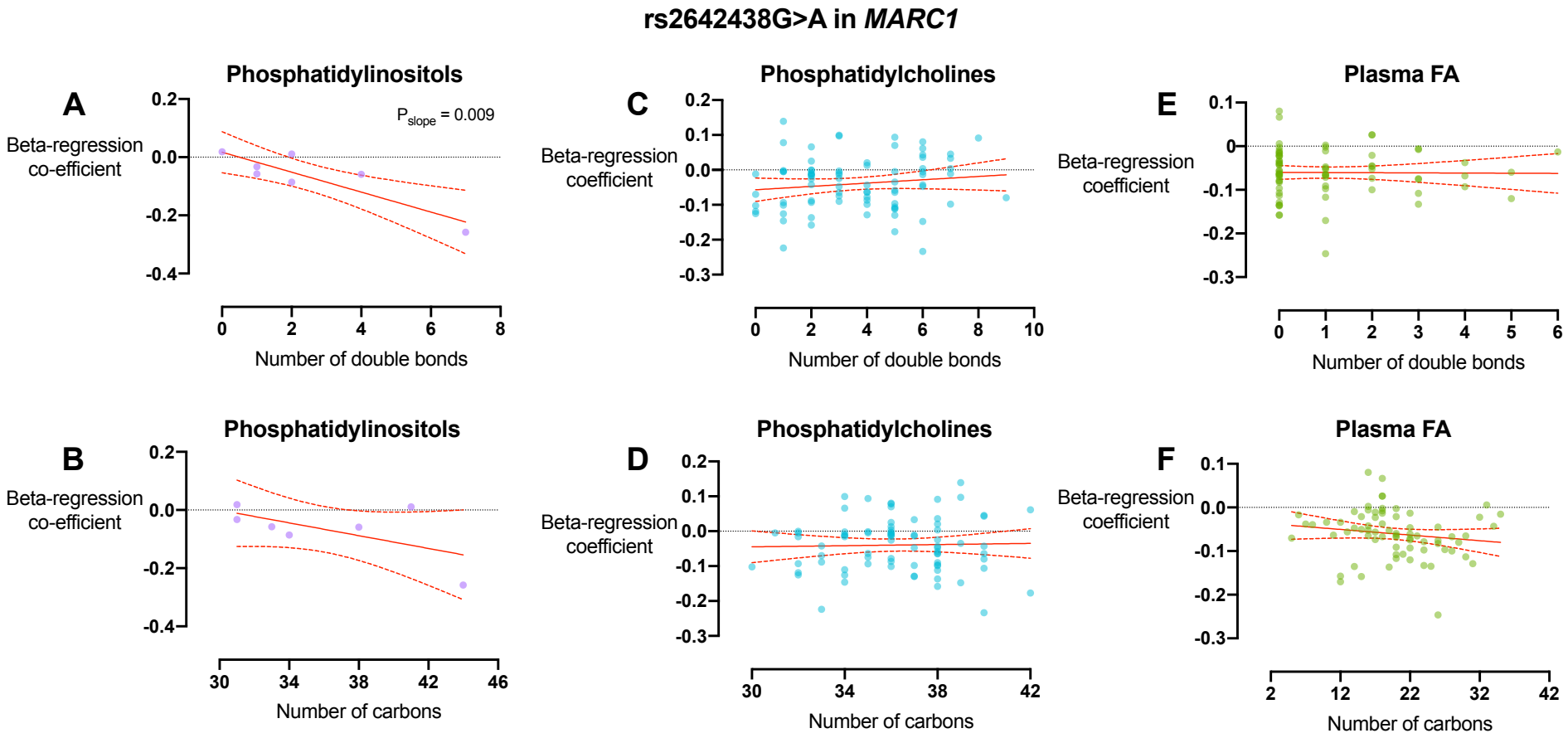


Supplementary Figure 1.
Odds ratios for the presence of any, moderate or advanced fibrosis using an additive genetic model. Data from 394 children with biopsy-proven NAFLD and liver histology. *'PNPLA3'* refers to 738409C>G, *'MARC1'* refers to rs2642438G>A, and *'HSD17B13'* refers to rs72613567T>TA.



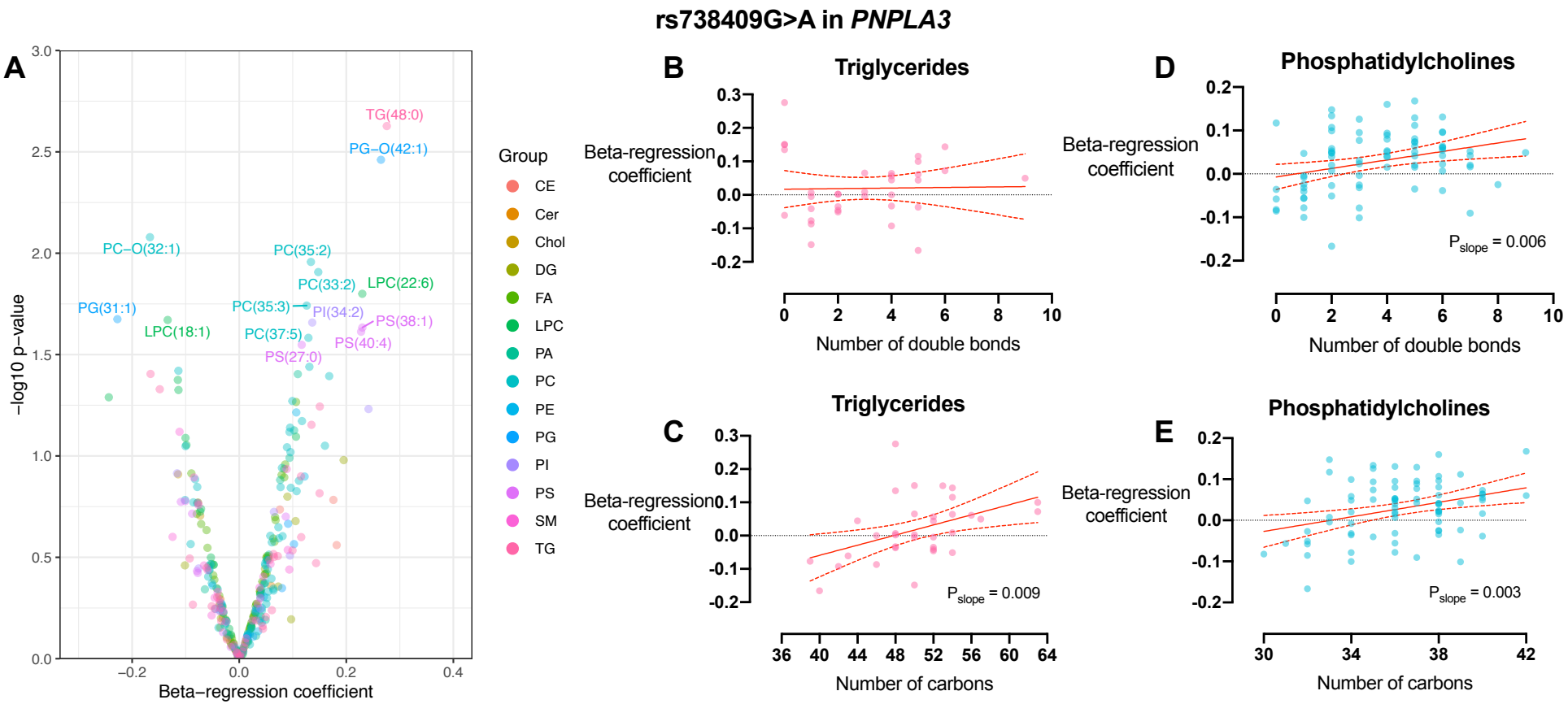
Supplementary Figure 2.

Characteristics of plasma lipid species associated with rs72613567T>TA in *HSD17B13*. There was no trend in the association between rs72613567T>TA and number of double bonds (A) or carbons (B) in phosphatidylcholines (PC). Whereas the variant was positively associated with polyunsaturated phosphatidylethanolamines (PE) and negatively associated with monounsaturated PE (C), but this was not significantly influenced by chain length (D). Similarly rs72613567T>TA was negatively associated with polyunsaturated fatty acids (FA, derived from plasma phospholipids, E), but there was no trend with carbon chain length (F). Beta-regression coefficient of association was calculated by linear regression between genotype (coding T/T=0, T/TA=1, TA/TA=2) and logarithmically-transformed lipid abundance, adjusted for age and sex. Simple linear regression with 95% confidence intervals are shown in red. Data from 129 children with biopsy-confirmed NAFLD.



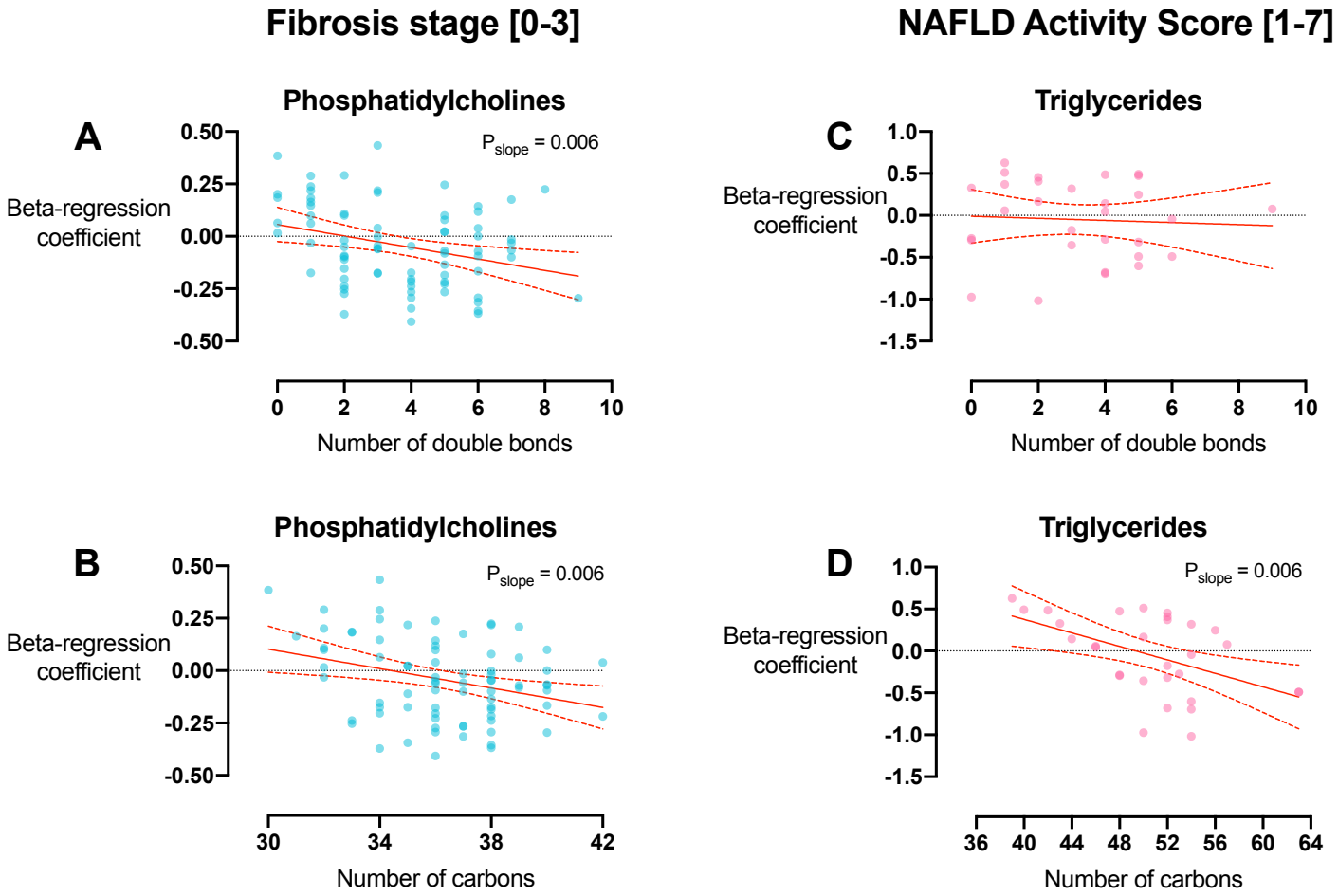
Supplementary Figure 3.

Characteristics of plasma lipid species associated with rs2642438G>A in *MARC1*. rs2642438G>A was negatively associated with polyunsaturated phosphatidylinositols (PI) and positively associated with unsaturated PI (A), but there was no relationship with carbon chain length (B). There was also no relationship between the variant and number of double bonds or chain length of PC (C & D) or plasma fatty acids from phospholipids (E & F). Beta-regression coefficient of association was calculated by linear regression between genotype (coding GG=0, GA=1, AA=2) and logarithmically-transformed lipid abundance, adjusted for age and sex. Simple linear regression with 95% confidence intervals are shown in red. Data from 129 children with biopsy-confirmed NAFLD.



Supplementary Figure 4.

Plasma lipid species associated with rs738409C>G in *PNPLA3*. Volcano plot demonstrating the association (as beta-regression coefficient) between plasma lipid species and rs738409C>G, where beta-regression coefficient was calculated by linear regression between genotype (coding CC=0, CG=1, GG=2) and logarithmically-transformed lipid abundance, adjusted for age and sex. The variant was positively associated with very long chain triglycerides (TG) (C) but there was no relationship with the number of double bonds (B). rs738409C>G was also more positively associated with polyunsaturated (D) and long-chain (E) PC. Simple linear regression with 95% confidence intervals are shown in red. Data from 129 children with biopsy-confirmed NAFLD.



Supplementary Figure 5.

Characteristics of plasma lipid species associated with histological severity of NAFLD in children. Fibrosis stage [0-3] was negatively associated with number of double bonds (A) and carbon chain length (B) in PC. NAFLD Activity Score was negatively associated with very long-chain TG that showed no relationship with number of double bonds © and positively associated with short chain TG (D), and . Beta-regression coefficient of association was calculated by linear regression between histological feature and logarithmically-transformed lipid abundance, adjusted for age and sex. Simple linear regression with 95% confidence intervals are shown in red.

Supplementary Tables

Odds Ratios for the presence of NAFLD.

SNP	Genotype	Controls	Cases	Genotypic Odds	*P* value	Allelic Odds	*P* value
HSD17B13						0.70 (0.28-0.88)	**0.002**
	T/T	206	394	1			
	T/TA	143	168	0.60 (0.45-0.81)	**0.001**		
	TA/TA	21	28	0.80 (0.59-1.09)	0.159		
	Trend			0.70 (0.56-0.87)	**0.002**		
	recessive			0.79 (0.44-1.44)	0.444		
	dominant			0.61 (0.47-0.81)	**0.001**		
MARC1						1.00 (0.81-1.24)	>0.999
	GG	199	324	1			
	GA	151	227	0.92 (0.70-1.22)	0.574		
	AA	20	39	1.13 (0.84-1.51)	0.419		
	Trend			1.02 (0.82-1.27)	0.889		
	recessive			1.31 (0.74-2.34)	0.352		
	dominant			0.96 (0.74-1.26)	0.781		
PNPLA3						1.84 (1.32-2.58)	**4.0×10^{-4}**
	CC	48	139	1			
	CG	42	172	1.41 (0.87.-2.3)	0.168		
	GG	7	82	2.02 (1.30-3.15)	**0.002**		
	Trend			1.79 (1.26-2.53)	**0.001**		
	recessive			3.54 (1.55-8.06)	**0.003**		
	dominant			1.78 (1.11-2.84)	**0.017**		
TM6SF2						1.14 (0.62 – 2.10)	0.761
	CC	84	337	1			
	CT	11	52	1.28 (0.63 – 2.62)	0.494		
	TT	1	4	1.04 (0.34 – 3.17)	0.949		
	Trend			1.22 (0.65 – 2.26)	0.535		
	dominant			1.27 (0.64 – 2.52)	0.497		

Supplementary Table 1.

P values were calculated by binary logistic regression with correction for age and sex (Wald-test) for genotype and chi-square (two-sided Fisher´s exact test) for allelic odds ratios.

Clinical and Laboratory Characteristics of NAFLD patients stratified by *HSD17B13* genotype.

Variable	T/T (N = 394)	T/TA and TA/TA (N = 196)	*P* value	*Q* value
Age (years)	13.0 (11.3 – 15.4)	13.0 (11.0 – 15.2)	0.182	0.824
Male sex, n (%)	241 (61.2)	126 (64.3)	0.462	0.924
BMI z-score	2.4 (1.8 – 3.0)	2.3 (1.8 – 3.1)	0.982	0.982
Obesity, n (%)	258 (67.0)	129 (65.8)	0.741	0.982
ALT (U/l)	49 (32 – 81)	45 (28 – 71)	0.209	0.824
AST (U/l)	39 (29 – 55)	37 (26 – 50)	0.264	0.824
GGT (U/l)	25 (16 – 38)	21 (16 – 31)	0.190	0.824
Cholesterol (mg/dl)	160 (140 – 184)	161 (139 – 186)	0.910	0.982
LDL (mg/dl)	97 (81 – 112)	97 (78 – 113)	0.925	0.982
HDL (mg/dl)	43 (38 – 50)	44 (39 – 51)	0.924	0.982
Triglycerides (mg/dl)	98 (72 – 141)	99 (68 – 144)	0.759	0.982
HOMA	3.7 (2.6 – 5.9)	3.5 (2.3 – 5.3)	0.308	0.824
MARC1 genotype, n (%)				
GG / GA / AA	217 (55.1)/152 (38.6)/25 (6.3)	107 (54.6)/75 (38.3)/14 (7.1)	0.751	0.982
PNPLA3 genotype, n (%)				
CC / CG / GG	91 (34.2)/114 (42.9)/61 (22.9)	48 (37.8.4)/58 (45.7)/21 (16.5)	0.353	0.824

Supplementary Table 2.

Data represent frequencies (%) or median (interquartile range) as appropriate. For clinical characteristics, *P*-values were calculated using Mann-Whitney U test for continuous traits and Chi-square test for categorical traits. For plasma markers, *P*-values were calculated using linear regression with correction for age, sex. For genotypes, *P*-values were calculated using binary logistic regression with correction for age and sex (Wald test). FDR correction (*Q* value) for multiple comparisons was calculated using the Benjamini and Hochberg method.

Clinical and Laboratory Characteristics of NAFLD patients stratified by *MARC1* genotype.

Variable	GG (N = 324)	GA and AA (N = 266)	*P* value	*Q* value
Age (years)	13.0 (11.0 – 15.1)	13.0 (11.5 – 15.5)	0.518	0.806
Male sex, n (%)	200 (61.7)	167 (62.8)	0.793	0.854
BMI z-score	2.4 (1.8 – 3.2)	2.3 (1.8 – 2.9)	0.209	0.806
Obesity, n (%)	215 (67.8)	171 (65.0)	0.476	0.806
ALT (U/l)	46 (30 – 78)	49 (31 – 79)	0.422	0.806
AST (U/l)	38 (28 – 55)	39 (27 – 53)	0.744	0.854
GGT (U/l)	23 (16 – 34)	24 (16 – 38)	0.161	0.806
Cholesterol (mg/dl)	162 (140 – 185)	158 (138 – 184)	0.331	0.806
LDL (mg/dl)	97 (81 – 113)	97 (81 – 112)	0.456	0.806
HDL (mg/dl)	44 (38 – 51)	43 (38 – 50)	0.251	0.806
Triglycerides (mg/dl)	98 (68 – 144)	99 (75 – 141)	0.720	0.854
HOMA	3.8 (2.5 – 5.5)	3.6 (2.5 – 5.9)	0.679	0.854
HSD17B13 genotype, n (%)				
TT / TTA / TATA	217 (67.0)/92 (28.4)/15 (4.6)	177 (66.5)/76 (28.6)/13 (4.9)	0.883	0.883
PNPLA3 genotype, n (%)				
CC / CG / GG	64 (30.5)/103 (49.0)/43 (20.5)	75 (41.0)/69 (37.7)/39 (21.3)	0.147	0.806

Supplementary Table 3.

Data represent frequencies (%) or median (interquartile range) as appropriate. For clinical characteristics, *P*-values were calculated using Mann-Whitney U test for continuous traits and Chi-square test for categorical traits. For plasma markers, *P*-values were calculated using linear regression with correction for age, sex. For genotypes, *P*-values were calculated using binary logistic regression with correction for age and sex (Wald test). FDR correction (*Q* value) for multiple comparisons was calculated using the Benjamini and Hochberg method.

Clinical and Laboratory Characteristics of patients with liver biopsies.

Variable	NAFLD (N = 590)	NAFLD with liver biopsy (N = 394)
Age (years)	13.0 (11.0 – 15.4)	13 (11.0 – 14.6)
Male sex, n (%)	386 (66.6)	240 (60.9)
BMI z-score	2.4 (1.8 – 3.0)	2.1 (1.6 – 2.6)
Obesity, n (%)	387 (66.5)	209 (53.0)
ALT (U/l)	47 (31 – 78)	49 (31 – 80)
AST (U/l)	38 (28 – 54)	40 (28 – 56)
GGT (U/l)	23 (16 – 36)	22 (14 – 39)
Cholesterol (mg/dl)	160 (139 – 185)	159 (138 – 184)
LDL (mg/dl)	97 (81 – 112)	98 (82 – 112)
HDL (mg/dl)	43 (38 – 50)	43 (38 – 50)
Triglycerides (mg/dl)	98 (71 – 141)	98 (72 – 144)
HOMA	3.7 (2.5 – 5.6)	3.6 (2.5 – 5.4)
HSD17B13 genotype, n (%)		
TT / TTA / TATA	394 (66.8)/168 (28.5)/28 (4.7)	263 (66.8)/116 (29.4)/15 (3.8)
MARC1 genotype, n (%)		
GG / GA / AA	324 (54.9)/227 (38.5)/39 (6.6)	204 (51.8)/168 (42.6)/22 (5.6)
PNPLA3 genotype, n (%)		
CC / CG / GG	139 (35.4)/172 (43.8)/82 (20.9)	134 (35.6)/162 (41.1)/80 (21.3)

Supplementary Table 4.

Data represent frequencies (%) or median (interquartile range) as appropriate. BMI, body mass index; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma glutamyl transferase; LDL, low density lipoprotein ; HDL, high density lipoprotein; HOMA, homeostatic model assessment of insulin resistance.

Associations of genetic variants with histologic features of disease severity.

Histologic Trait	HSD17B13			MARC1			PNPLA3			TM6SF2		
n (% of genotype)	TT	TTA	TATA	GG	GA	AA	CC	CG	GG	CC	CT	TT
Steatosis												
1	56 (21.3)	29 (25.0)	5 (33.3)	39 (19.1)	41 (24.4)	10 (45.5)	39 (29.1)	34 (21.0)	13 (16.3)	78 (24.2)	7 (14.0)	1 (25.0)
2	141 (53.6)	44 (37.9)	2 (13.3)	100 (49.0)	78 (46.4)	9 (40.9)	68 (50.7)	79 (48.8)	30 (37.5)	148 (46.0)	26 (52.0)	3 (75.0)
3	66 (25.1)	43 (37.1)	8 (53.3)	65 (31.9)	49 (29.2)	3 (13.6)	27 (20.1)	49 (30.2)	37 (46.3)	96 (29.8)	17 (34.0)	0 (0.0)
P value, univariate		0.151			**0.018**			**9.783×10^{-5}**			0.524	
multivariate		0.386			**0.022**			**3.195×10^{-5}**			0.322	
Fibrosis												
0	56 (21.3)	30 (26.1)	4 (26.7)	46 (22.5)	38 (22.8)	6 (27.3)	36 (26.9)	34 (21.0)	15 (19.0)	70 (21.7)	13 (26.5)	2 (50.0)
1	96 (36.5)	50 (43.5)	6 (40.0)	77 (37.7)	67 (40.1)	8 (36.4)	56 (41.8)	62 (38.3)	29 (36.7)	128 (39.8)	19 (38.8)	0 (0.0)
2	75 (28.5)	19 (16.5)	5 (33.3)	48 (23.5)	44 (26.3)	7 (31.8)	32 (23.9)	44 (27.2)	18 (22.8)	85 (26.4)	8 (16.3)	1 (25.0)
3	35 (13.3)	15 (13.0)	0 (0.0)	32 (15.7)	17 (10.2)	1 (4.5)	10 (7.5)	21 (13.0)	16 (20.3)	37 (11.5)	9 (18.4)	1 (25.0)
4	1 (0.4)	1 (0.9)	0 (0.0)	1 (0.5)	1 (0.6)	0 (0.0)	0 (0.0)	1 (0.6)	1 (1.3)	2 (0.6)	0 (0.0)	0 (0.0)
P value, univariate		0.075			0.362			**0.012**			0.661	
multivariate		0.134			0.394			**0.018**			0.955	
Lobular I.												
0	50 (19.0)	22 (19.0)	5 (33.3)	39 (19.1)	33 (19.6)	5 (22.7)	30 (22.4)	30 (18.5)	11 (13.8)	58 (18.0)	11 (22.0)	2 (50.0)
1	132 (50.2)	64 (55.2)	6 (40.0)	104 (51.0)	85 (50.6)	13 (59.1)	70 (52.2)	80 (49.4)	41 (51.2)	166 (51.6)	25 (50.0)	0 (0.0)
2	81 (30.8)	30 (25.9)	4 (26.7)	61 (29.9)	50 (29.8)	4 (18.2)	34 (25.4)	52 (32.1)	28 (35.0)	98 (30.4)	14 (28.0)	2 (50.0)
P value, univariate		0.284			0.503			0.056			0.513	
multivariate		0.104			0.470			**0.020**			0.690	

Portal I.												
0	65 (24.7)	44 (38.9)	8 (53.3)	65 (32.5)	46 (28.0)	6 (27.3)	41 (30.6)	52 (33.3)	18 (23.1)	91 (28.9)	18 (36.7)	2 (50.9)
1	160 (60.8)	59 (52.2)	7 (46.7)	107 (53.5)	106 (64.6)	13 (59.1)	79 (59.0)	88 (56.4)	47 (60.3)	186 (59.0)	26 (53.1)	2 (50.0)
2	33 (12.5)	10 (8.8)	0 (0.0)	28 (14.0)	12 (7.3)	3 (13.6)	14 (10.4)	16 (10.3)	13 (16.7)	38 (12.1)	5 (10.2)	0 (0.0)
P value,univariate		**7.282×10^{-4}**			0.884			0.202			0.156	
multivariate		**1.542×10^{-4}**			0.851			0.234			0.182	
Ballooning												
0	76 (29.3)	39 (34.2)	6 (40.0)	63 (31.2)	51 (31.1)	7 (31.8)	41 (30.6)	50 (31.8)	21 (26.6)	94 (29.7)	16 (32.7)	2 (50.0)
1	110 (42.5)	48 (42.1)	3 (20.0)	83 (41.1)	69 (42.1)	9 (40.9)	59 (44.0)	55 (35.0)	38 (48.1)	131 (40.7)	19 (38.8)	2 (50.0)
2	73 (28.2)	27 (23.7)	6 (40.0)	56 (27.7)	44 (26.8)	6 (27.3)	34 (25.4)	52 (33.1)	20 (25.3)	92 (29.0)	14 (28.6)	0 (0.0)
P value, univariate		0.422			0.917			0.646			0.380	
multivariate		0.241			0.971			0.710			0.236	

Supplementary Table 5.

Data represent frequencies *P*-values were calculated using univariate or multivariate logistic regression with correction for age, sex, BMI z-score and HOMA.

Odds Ratios for the presence of fibrosis.

SNP	Fibrosis	Odds ratio, additive model	*P* value	Odds ratio, dominant model	*P* value
HSD17B13					
	Any	0.82 (0.54 - 1.23)	0.332	0.77 (0.47 - 1.27)	0.309
	Moderate	0.66 (0.45 - 0.99)	**0.043**	0.61 (0.39 - 0.97)	**0.037**
	Advanced	0.74 (0.41 - 1.34)	0.322	0.82 (0.43 - 1.59)	0.562
MARC1					
	Any	0.93 (0.63 - 1.38)	0.731	0.96 (0.60 - 1.53)	0.857
	Moderate	0.93 (0.66 - 1.31)	0.673	0.90 (0.60 - 1.37)	0.632
	Advanced	0.53 (0.30 - 0.94)	**0.028**	0.50 (0.27 - 0.95)	**0.033**
PNPLA3					
	Any	1.25 (0.89 - 1.76)	0.194	1.40 (0.85 - 2.32)	0.185
	Moderate	1.33 (1.00 - 1.78)	**0.049**	1.58 (1.00 - 2.48)	**0.048**
	Advanced	1.96 (1.29 - 3.00)	**0.002**	2.59 (1.23 - 5.44)	**0.012**
TM6SF2					
	Any	0.68 (0.38 - 1.21)	0.190	0.70 (0.36 - 1.34)	0.276
	Moderate	0.95 (0.55 - 1.63)	0.845	0.89 (0.49 - 1.64)	0.714
	Advanced	1.60 (0.82 - 3.13)	0.171	1.69 (0.78 - 3.63)	0.183

Supplementary Table 6.

P values were calculated by binary logistic regression with correction for age and sex.

Supplementary Table 7.

Clinical and Laboratory Characteristics of patients with liver tissue proteomic profiles.

Variable	NAFLD (N = 590)	NAFLD with liver proteomics (N = 70)
Age (years)	13.0 (11.0 – 15.4)	14.0 (13.0 – 16.0)
Male sex, n (%)	386 (66.6)	54 (77.1)
BMI z-score	2.4 (1.8 – 3.0)	2.9 (2.5 – 3.2)
Obesity, n (%)	387 (66.5)	60 (85.7)
ALT (U/l)	47 (31 – 78)	88 (62 – 125)
AST (U/l)	38 (28 – 54)	52 (39 – 69)
GGT (U/l)	23 (16 – 36)	40 (30 – 56)
Cholesterol (mg/dl)	160 (139 – 185)	163 (146 – 181)
LDL (mg/dl)	97 (81 – 112)	104 (90 – 119)
HDL (mg/dl)	43 (38 – 50)	42 (36 – 51)
Triglycerides (mg/dl)	98 (71 – 141)	118 (93 – 168)
HOMA	3.7 (2.5 – 5.6)	5.9 (4.4 – 9.0)
HSD17B13 genotype, n (%)		
TT / TTA / TATA	394 (66.8)/168 (28.5)/28 (4.7)	50 (71.4)/15 (21.4)/5 (7.1)
MARC1 genotype, n (%)		
GG / GA / AA	324 (54.9)/227 (38.5)/39 (6.6)	41 (58.6)/23 (32.9)/6 (8.6)
PNPLA3 genotype, n (%)		
CC / CG / GG	139 (35.4)/172 (43.8)/82 (20.9)	20 (29.0)/31 (44.9)/18 (26.1)

Supplementary Table 7.

Data represent frequencies (%) or median (interquartile range) as appropriate. BMI, body mass index; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma glutamyl transferase; LDL, low density lipoprotein ; HDL, high density lipoprotein; HOMA, homeostatic model assessment of insulin resistance .

Supplementary Table 8 [Excel spreadsheet].
Hepatic proteomics results for rs72613567T>TA in *HSD17B13* and rs2642438G>A in *MARC1*. Association between proteins and variants was calculated using linear regression (for an additive model coding number of effect alleles as 0, 1, or 2) adjusted for age and sex. Beta regression coefficient therefore represents the change in log-standard deviations of protein abundance per effect allele.

Supplementary Table 9 [Excel spreadsheet].
Gene set enrichment analysis (GSEA) results for rs72613567T>TA in *HSD17B13* and rs2642438G>A in *MARC1* for gene sets with nominal P-value < 0.05 and FDR Q-value <0.25.

Predicted consequence of p.Ala165Thr in *MARC1* (from rs2642438G>A) using three *in silico* analysis tools.

Tool	Result for *MARC1* p.Ala165Thr	Explanatory notes
Prediction of impact of missense variants on protein function		
SNPs&GO[10,11]	PhD-SNP: Neutral (Prob=0.483, RI=0) PANTHER: Disease (Prob=0.804, RI=6) SNPs&GO: Neutral (Prob=0.455, RI=1)	Several models are integrated to give the overall SNPs&GO output. A probability of >0.5 is predicted as 'Disease'. https://snps.biofold.org/snps-and-go/
Align-GVGD[12–14]	GV=0.0 GD=58.02 Prediction=Class C55	GV is a measure of biochemical variation of the mutation. GD is a measure of the difference in properties of mutation. There are seven classifiers, where C65 is the most likely to interfere with protein function, C55 is the second most likely, and C0 is the least likely. http://agvgd.hci.utah.edu/
MutPred2[15]	Score = 0.745 Affected motifs: Loss of Helix (Prob=0.27, p=0.05); Altered Metal binding (Prob=0.26, p=9.6e-03); Gain of Relative solvent accessibility (Prob=0.25, p=0.03); Gain of Allosteric site at W168 (Prob=0.21,p=0.03); Gain of Catalytic site at W168 (Prob=0.18, p=0.01); Gain of Disulfide linkage at C161 (Prob=0.15, p=0.03); Loss of Pyrrolidone carboxylic acid at Q167 (Prob=0.10, p=0.01)	Score of >0.5 indicates pathogenicity. MutPred2 also predicts the probability of structural & functional properties and generates a p-value for each to occur compared to the probability of those motifs being altered by benign mutations. A probability of >0.25 is suggested as a threshold for implicating a particular mechanism of pathogenicity, interpreted in combination with the p-value. http://mutpred.mutdb.org/index.html
Prediction of impact on protein stability		
I-Mutant3.0[16–18]	ΔΔG Value Prediction: -0.63 kcal/mol SVM2 Prediction Effect: Decrease, RI=7 SVM3 Prediction Effect: Large Decrease, RI=3	A negative ΔΔG indicates a decrease in the stability of protein tertiary structure. The tool can either give a binary classification (SMV2) of increase/decrease; or a ternary classification (SMV3) or increase/neutral/decrease. http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi
DUET[19]	mCSM Predicted Stability Change (ΔΔG):-1.901 kcal/mol (Destabilizing) SDM Predicted Stability Change (ΔΔG): -2.54 Kcal/mol (Destabilizing) DUET Predicted Stability Change (ΔΔG): -2.083 Kcal/mol (Destabilizing)	A tool that combines two previously published approaches (SDM and mCSM) into a single estimate of protein stability, expressed as ΔΔG. This is calculated using the known crystalline structure of MARC1 (6fw2 on PDBe). http://biosig.unimelb.edu.au/duet/stability
CUPSAT[20–22]	Overall stability (ΔΔG):-3.74 kcal/mol (Destabilizing), with unfavourable torsion	A tool that combines the physical properties of amino acids with the known crystalline structure of MARC1 (6fw2 on PDBe) to predict the impact on protein stability. http://cupsat.tu-bs.de/

Supplementary Table 10.

GD, Grantham Difference; GV, Grantham Variation; mCSM, mutation Cutoff Scanning Matrix; PDBe, Protein Data Bank in Europe; Prob, probability; RI, reliability index; SDM, Site Directed Mutator;

Clinical and Laboratory Characteristics of patients with plasma lipidomics.

Variable	NAFLD (N = 590)	NAFLD with plasma lipidomics (N = 129)
Age (years)	13.0 (11.0 – 15.4)	12.4 (10.4 – 13.4)
Male sex, n (%)	386 (66.6)	67 (52.0)
BMI z-score	2.4 (1.8 – 3.0)	2.08 (1.8 – 2.6)
Obesity, n (%)	387 (66.5)	74 (57.3)
ALT (U/l)	47 (31 – 78)	60 (41 – 80)
AST (U/l)	38 (28 – 54)	44 (33 – 56)
Cholesterol (mg/dl)	160 (139 – 185)	157 (140 – 190)
LDL (mg/dl)	97 (81 – 112)	99 (90 – 105)
HDL (mg/dl)	43 (38 – 50)	43 (38 – 47)
Triglycerides (mg/dl)	98 (71 – 141)	112 (82 – 155)
HOMA	3.7 (2.5 – 5.6)	3.0 (2.1 – 4.1)
HSD17B13 genotype, n (%)		
TT / TTA / TATA	394 (66.8)/168 (28.5)/28 (4.7)	87 (67.4)/39 (30.2)/3 (2.3)
MARC1 genotype, n (%)		
GG / GA / AA	324 (54.9)/227 (38.5)/39 (6.6)	62 (48.1)/60 (46.5)/7 (5.4)
PNPLA3 genotype, n (%)		
CC / CG / GG	139 (35.4)/172 (43.8)/82 (20.9)	47 (36.4)/63 (48.8)/19 (14.7)

Supplementary Table 11.

Data represent frequencies (%) or median (interquartile range) as appropriate. BMI, body mass index; ALT, alanine aminotransferase; AST, aspartate aminotransferase; LDL, low density lipoprotein ; HDL, high density lipoprotein; HOMA, homeostatic model assessment of insulin resistance

Supplementary References

1. Kulak NA, Pichler G, Paron I, et al. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nat Methods 2014;11:319–324.
2. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 2008;26:1367–1372.
3. Tyanova S, Temu T, Sinitcyn P, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 2016;13:731–740.
4. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–15550.
5. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–2504.
6. Smoot ME, Ono K, Ruscheinski J, et al. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 2011;27:431–432.
7. Merico D, Isserlin R, Stueker O, et al. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 2010;5:e13984.
8. Kucera M, Isserlin R, Arkhangorodsky A, et al. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. F1000Res 2016;5:1717.
9. Oesper L, Merico D, Isserlin R, et al. WordCloud: a Cytoscape plugin to create a visual semantic summary of networks. Source Code Biol Med 2011;6:7.
10. Capriotti E, Calabrese R, Fariselli P, et al. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. BMC Genomics 2013;14 Suppl 3:S6.
11. Calabrese R, Capriotti E, Fariselli P, et al. Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat 2009;30:1237–1244.
12. Tavtigian SV, Deffenbaugh AM, Yin L, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J Med Genet 2006;43:295–305.
13. Tavtigian SV, Byrnes GB, Goldgar DE, et al. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. Hum Mutat 2008;29:1342–1354.
14. Mathe E, Olivier M, Kato S, et al. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. Nucleic Acids Res 2006;34:1317–1325.
15. Pejaver V, Urresti J, Lugo-Martinez J, et al. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. bioRxiv 2017:134981. Available at: https://www.biorxiv.org/content/10.1101/134981v1 [Accessed May 7, 2020].
16. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 2005;33:W306–10.
17. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 2006;22:2729–2734.

18. Capriotti E, Fariselli P, Calabrese R, et al. Predicting protein stability changes from sequences using support vector machines. Bioinformatics 2005;21 Suppl 2:ii54–8.
19. Pires DEV, Ascher DB, Blundell TL. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic Acids Res 2014;42:W314–9.
20. Parthiban V, Gromiha MM, Schomburg D. CUPSAT: prediction of protein stability upon point mutations. Nucleic Acids Res 2006;34:W239–42.
21. Parthiban V, Gromiha MM, Hoppe C, et al. Structural analysis and prediction of protein mutant stability using distance and torsion potentials: role of secondary structure and solvent accessibility. Proteins 2007;66:41–52.
22. Parthiban V, Gromiha MM, Abhinandan M, et al. Computational modeling of protein mutant stability: analysis and optimization of statistical potentials and structural features reveal insights into prediction model development. BMC Struct Biol 2007;7:54.