Contribution of a Genetic Risk Score to Ethnic Differences in Fatty Liver Disease

Maddie J. Kubiliun, MD, Jonathan C. Cohen, PhD, Helen H. Hobbs, MD, and Julia Kozlitina, PhD

Table of Contents

Supporting Information

Matched Case-Control Cohort

To ensure that genetic analyses were not confounded by differences between DHS participants and FLD patients with respect to age, sex and ethnic composition, we also performed a sensitivity analysis by selecting a sample of sex-, age-, and ethnicity-matched controls. To supplement the relatively small number of Whites and Hispanics in the Dallas Heart Study sample, additional controls were obtained from the Dallas Biobank (DB). The DB is a repository of genomic DNA and plasma samples obtained from individuals in North-Central Texas.[1] Study participants provided written informed consent for inclusion in the repository, and were over 18 years of age. Whites in the DB had completed a preventive medicine examination at the Cooper Clinic in Dallas, TX between 2008 and 2012 and were enrolled into the study after signing an informed consent approved by the Cooper Research Institute IRB. DB participants with elevated levels of serum ALT level were excluded. Controls were matched to FLD patients 1:1 for sex, ethnicity, and age (using a caliper of 0.5 SD units).

The matched case-control analysis included 1162 FLD patients, for whom we were able to identify a match (424 Whites, 57 Blacks, 646 Hispanic, and 35 other ethnicities), and an equal number of matched controls (712 from the DHS and 450 from the DB).

Genotype QC

Genotypes underwent standard quality control filtering (call rate >95%, Hardy-Weinberg equilibrium p-value >0.001 among controls). The allele frequencies of all variants among self-reported White, Black and Hispanic controls were concordant with frequencies reported among corresponding ethnic groups in gnomAD (https://gnomad.broadinstitute.org) and the 1000 Genomes Project (https://www.internationalgenome.org). In addition, in a subset of DHS participants (>75%) and UTSW-FLD participants (~35%), genotypes for the above variants were assayed using an independent method (Illumina Exome Chip and TaqMan assay in DHS, and

whole-exome sequencing in UTSW-FLD). Among these participants, cross-platform genotype concordance was >99.5%.

Sensitivity Analysis – GRS based on Weights from External Cohorts

To correct for possible bias resulting from using internally estimated weights, we performed an additional sensitivity analysis, in which the weights were estimated externally, based on data from previous studies. For *PNPLA3*-rs738409, the estimates of association with different stages of NAFLD and ALD have ranged between OR=1.5 and OR=3.26 per each risk allele, with most studies reporting odds ratios >2.[2-5] For *TM6SF2*-rs58542926, effect estimates have ranged between OR=1.5 and OR=2.1.[2,4,5] For *TMC4*/*MBOAT7*-rs641738 and *GCKR*-rs1260326, the reported odds ratios have ranged between OR=1.2 and 2.1.[10,4] For *HSD17B13*-rs72613567, heterozygous odds ratios have been in the range 0.56-0.84.[5,6] We therefore used the mid- to lower-bound values of these ranges as conservative estimates of the effects of these variants. We could not identify independent estimates of the effect of *HSD17B13*-rs80182459. However, given that both *HSD17B13*-rs80182459 and -rs72613567 result in a loss of function, we assumed that the effect of rs80182459 is similar to that of rs72613567. Similarly, we identified only one independent report of the association of *PNPLA3*-rs6006460 with steatosis.[7] In this report and in our own data,[8] the effect of *PNPLA3*-rs6006460 was about 0.5-0.7 times the magnitude of the effect of *PNPLA3*-rs738409, but in the opposite direction. We therefore estimated the effect of this variant to be -0.6 * log(effect of *PNPLA3*-rs738409). The resulting score was calculated as follows:

$GRS_{\text{external weights}}$ = ln(2) *PNPLA3*-rs738409 + ln(1.8) *TM6SF2*-rs58542926 + log(1.2) *TMC4*/*MBOAT7*-rs641738 + log(1.2) *GCKR*-rs1260326 + ln(0.7) *HSD17B13*-rs72613567 + ln(0.7) *HSD17B13*-rs80182459 - 0.6 * ln(2) *PNPLA3*-rs6006460.

Table S1 Baseline Characteristics of NAFLD and ALD Participants in UTSW-FLD Cohort

Characteristic	No. of people with data	NAFLD (n=733)	ALD (n=461)	P-value
Age, years, mean ± SD	1194	55 ± 12.7	52.6 ± 10.3	<0.001
Female, N (%)	1194	543 (74.1)	118 (25.6)	<0.001
Race/ethnicity, N (%)				
Non-Hispanic Black	1194	35 (4.8)	22 (4.8)	1
Non-Hispanic White	1194	249 (34)	180 (39)	0.083
Hispanic	1194	418 (57)	250 (54.2)	0.369
Other	1194	31 (4.2)	9 (2)	0.033
BMI, kg/m^2, mean ± SD	1194	34.2 ± 7.2	29.6 ± 6.5	<0.001
Glucose, mg/dL, median (IQR)	1028	116 (96 - 161)	106 (91 - 136)	0.015
Type 2 Diabetes	1194	386 (52.7)	118 (25.6)	<0.001
Systolic BP, mm Hg, mean ± SD	1176	128.4 ± 16.5	122.5 ± 16.3	<0.001
Diastolic BP, mm Hg, mean ± SD	1176	72.5 ± 11.5	69.9 ± 11.2	<0.001
Total cholesterol, mg/dL, mean ± SD	849	168.9 ± 48.5	149.3 ± 56.4	<0.001
LDL-cholesterol, mg/dL, mean ± SD	835	92.7 ± 37.5	82.4 ± 40.8	<0.001
HDL-cholesterol, mg/dL, mean ± SD	843	44.6 ± 15.6	40.9 ± 23.2	0.081
Triglycerides, mg/dL, median (IQR)	851	129 (93 - 195.5)	94 (69 - 140.2)	<0.001
ALT (U), median (IQR)	1177	43 (28 - 69)	31 (22 - 46)	<0.001
AST (U), median (IQR)	1174	42 (31 - 61)	48 (33 - 73)	0.0028
AST/ALT, median (IQR)	1173	0.96 (0.72 - 1.34)	1.57 (1.2 - 2.1)	<0.001

UTSW-FLD subjects stratified based on etiology of fatty liver disease (FLD), either nonalcoholic fatty liver disease (NAFLD) or alcoholic fatty liver disease (ALD).
BP, blood pressure; TG, triglyceride; ALT, alanine aminotransaminase; AST, aspartate aminotransferase

Table S2: Characteristics of FLD Patients Stratified by Disease Stage

Characteristic	Total	Steatosis/ Steatohepatitis	Cirrhosis	HCC	P-value
N	1194	446	684	64	
Age, years, mean ± SD	54.1 ± 11.9	49.6 ± 12.2	56.2 ± 10.7	63 ± 10.2	<0.001
Female, N (%)	661 (55.4)	317 (71.1)	327 (47.8)	17 (26.6)	<0.001
Race/ethnicity, N (%)					
Non-Hispanic White	429 (35.9)	128 (28.7)	274 (40.1)	27 (42.2)	0.15
Non-Hispanic Black	57 (4.8)	32 (7.2)	24 (3.5)	1 (1.6)	0.046
Hispanic	668 (55.9)	261 (58.5)	372 (54.4)	35 (54.7)	0.42
Other	40 (3.4)	25 (5.6)	14 (2)	1 (1.6)	<0.001
BMI, kg/m^2, mean ± SD	32.4 ± 7.3	33.7 ± 6.8	31.8 ± 7.5	30.2 ± 6.4	<0.001
Glucose, mg/dL, median (IQR)	111 (94 - 150)	105 (93 - 136)	115 (95 - 158)	115 (99 - 143.8)	0.051
Type 2 Diabetes	504 (42.2)	163 (36.5)	306 (44.7)	35 (54.7)	0.020
Systolic BP, mm Hg, mean ± SD	126.1 ± 16.7	128.1 ± 14.9	124.7 ± 17.6	126.8 ± 17.2	<0.001
Diastolic BP, mm Hg, mean ± SD	71.5 ± 11.5	76.3 ± 10.2	68.9 ± 11.2	66.4 ± 11.3	<0.001
Total cholesterol, mg/dL, mean ± SD	162.1 ± 52.2	186 ± 43	144.9 ± 51.9	150.3 ± 47.5	<0.001
LDL-cholesterol, mg/dL, mean ± SD	89.2 ± 39	105.4 ± 35.9	78.1 ± 37.8	76.1 ± 27.7	<0.001
HDL-cholesterol, mg/dL, mean ± SD	43.4 ± 18.6	44.3 ± 13.4	42.3 ± 21.7	47.3 ± 19.6	<0.001
Triglycerides, mg/dL, median (IQR)	117 (82 - 179)	152 (110 - 234)	99 (71 - 142)	87 (71 - 131)	<0.001
ALT (U), median (IQR)	37 (25 - 60)	54 (32 - 84.5)	31 (23 - 46)	38 (25.5 - 47)	<0.001
AST (U), median (IQR)	44 (31.2 - 66)	39 (28 - 65)	46 (33.5 - 67)	55 (41 - 73)	<0.001
AST/ALT, median (IQR)	1.18 (0.82 - 1.69)	0.78 (0.62 - 0.98)	1.47 (1.12 - 1.92)	1.48 (1.18 - 1.88)	<0.001

For continuous characteristics, P-values were calculated using linear regression models, adjusted for age, gender, ethnicity, and BMI, where appropriate. For categorical variables, P-values were calculated using logistic regression with the same covariates. Disease stage was entered into the model as a categorical variable and coded using dummy variables.

Table S3. Baseline Characteristics of Controls Stratified by Ethnicity

Characteristic	White (1)	Black (2)	Hispanic (3)	P-value (2 vs 1)	P-value (3 vs 1)	P-value (3 vs 2)
N	863	1767	417			
Age, years, mean ± SD	49.8 ± 11.6	47.5 ± 11.6	42.9 ± 10.9	**<0.001**	**<0.001**	**<0.001**
Female, N (%)	512 (59.3)	1081 (61.2)	246 (59)	0.37	0.95	0.44
BMI, kg/m^2, mean ± SD	28.4 ± 6.5	31.8 ± 8.3	29.3 ± 6.2	**<0.001**	**0.0011**	**<0.001**
Obese, N (%)	273 (31.6)	901 (51)	154 (36.9)	**<0.001**	**0.012**	**<0.001**
Glucose, mg/dL, median (IQR)	93 (87 - 99)	91 (84 - 101)	94 (87 - 102)	**0.0058**	**<0.001**	**<0.001**
Type 2 Diabetes	69 (8)	277 (15.7)	38 (9.1)	**<0.001**	**0.017**	0.38
Systolic BP, mm Hg, mean ± SD	124.6 ± 17.5	134.2 ± 20.7	121.4 ± 18.6	**<0.001**	0.86	**<0.001**
Diastolic BP, mm Hg, mean ± SD	76.7 ± 8.2	81.9 ± 9.9	75.6 ± 9.1	**<0.001**	0.095	**<0.001**
ALT (U), median (IQR)	18 (14 - 23)	16 (12 - 21)	18 (14 - 23)	**<0.001**	0.55	**<0.001**
AST (U), median (IQR)	20 (17 - 24)	19 (16 - 23)	20 (17 - 24)	0.10	0.43	**0.034**
Net number of risk alleles, mean ± SD	1.6 ± 1.35	0.55 ± 1.26	1.9 ± 1.42	**<0.0001**	**0.0003**	**<0.0001**
GRS, mean ± SD	0.56 ± 0.66	-0.15 ± 0.91	0.91 ± 0.85	**<0.0001**	**<0.0001**	**<0.0001**

Obesity is defined as BMI ≥30 kg/m^2. BP, blood pressure; ALT, alanine aminotransaminase; AST, aspartate aminotransferase

Table S4: Association of Individual Variants with FLD in the Primary Cohort

			Cases (UTSW-FLD)					Controls						
			Genotype frequency			Allele	P	Genotype frequency			Allele	P	OR	P-value
Gene variant	SNP	Allele	REF	HET	HOM	frequency	(HWE)	REF	HET	HOM	frequency	(HWE)	(95% CI)	
All groups														
PNPLA3 I148M	rs738409	G	276	470	446	0.5713	-	2021	936	163	0.2022	-	2.6 (2.3 - 2.93)	1.1E-57
TM6SF2 E167K	rs58542926	T	999	183	10	0.0852	-	2867	248	4	0.0410	-	1.88 (1.5 - 2.36)	6.3E-08
GCKR P446L	rs1260326	T	486	541	161	0.3632	-	1874	1045	201	0.2319	-	1.23 (1.09 - 1.38)	8.3E-04
TMC4/MBOAT7	rs641738	T	431	560	196	0.4010	-	1302	1409	397	0.3544	-	1.12 (1 - 1.25)	0.060
PNPLA3 S453I	rs6006460	T	1175	8	0	0.0034	-	2737	364	19	0.0644	-	0.26 (0.11 - 0.51)	1.9E-05
HSD17B13 A192fs	rs80182459	-	1148	31	1	0.0140	-	2483	566	71	0.1135	-	0.44 (0.28 - 0.66)	5.1E-05
HSD17B13 splice	rs72613567	TA	913	239	30	0.1265	-	2392	633	95	0.1319	-	0.69 (0.59 - 0.81)	2.9E-06
White														
PNPLA3 I148M	rs738409	G	163	184	81	0.4042	0.027	526	300	37	0.2167	0.55	2.32 (1.9 - 2.85)	4.4E-17
TM6SF2 E167K	rs58542926	T	330	90	9	0.1259	0.38	761	100	2	0.0603	0.76	2.09 (1.53 - 2.88)	4.8E-06
GCKR P446L	rs1260326	T	126	220	82	0.4486	0.44	320	422	121	0.3847	0.35	1.37 (1.13 - 1.65)	0.0010
TMC4/MBOAT7	rs641738	T	125	206	96	0.4660	0.56	297	414	148	0.4133	0.89	1.17 (0.97 - 1.4)	0.097
PNPLA3 S453I	rs6006460	T	427	1	0	0.0012	1	862	1	0	0.0006	1	-	-
HSD17B13 A192fs	rs80182459	-	423	1	0	0.0012	1	861	2	0	0.0012	1	-	-
HSD17B13 splice	rs72613567	TA	264	134	28	0.2230	0.068	462	332	69	0.2723	0.39	0.85 (0.69 - 1.04)	0.11
Black														
PNPLA3 I148M	rs738409	G	36	20	1	0.1930	0.67	1317	416	34	0.1370	0.84	1.44 (0.87 - 2.29)	0.15
TM6SF2 E167K	rs58542926	T	54	3	0	0.0263	1	1655	110	2	0.0323	0.70	0.77 (0.19 - 2.1)	0.64
GCKR P446L	rs1260326	T	48	8	1	0.0877	0.35	1322	417	28	0.1338	0.54	0.61 (0.3 - 1.14)	0.13
TMC4/MBOAT7	rs641738	T	20	31	6	0.3772	0.40	812	750	198	0.3256	0.21	1.24 (0.84 - 1.82)	0.27
PNPLA3 S453I	rs6006460	T	50	6	0	0.0536	1	1403	347	17	0.1078	0.46	0.45 (0.17 - 0.94)	0.032
HSD17B13 A192fs	rs80182459	-	40	14	1	0.1455	1	1164	533	70	0.1904	0.35	0.72 (0.4 - 1.2)	0.22
HSD17B13 splice	rs72613567	TA	49	8	0	0.0702	1	1577	181	9	0.0563	0.17	1.26 (0.55 - 2.47)	0.56

Hispanic

PNPLA3 I148M	rs738409	G	63	250	354	0.7181	0.056	136	191	90	0.4448	0.14	2.98 (2.41 - 3.72)	8.7E-26
TM6SF2 E167K	rs58542926	T	584	81	1	0.0623	0.50	386	30	0	0.0361	1	1.76 (1.08 - 2.92)	0.023
GCKR P446L	rs1260326	T	299	290	74	0.3303	0.79	197	178	42	0.3141	0.82	1.18 (0.95 - 1.46)	0.13
TMC4/MBOAT7	rs641738	T	277	304	82	0.3529	1	173	198	45	0.3462	0.33	1.05 (0.85 - 1.31)	0.64
PNPLA3 S453I	rs6006460	T	658	1	0	0.0008	1	406	10	1	0.0144	0.077	0.06 (0 - 0.39)	0.0010
HSD17B13 A192fs	rs80182459	-	645	16	0	0.0121	1	392	24	1	0.0312	0.33	0.31 (0.14 - 0.66)	0.0020
HSD17B13 splice	rs72613567	TA	573	85	2	0.0674	0.76	308	95	14	0.1475	0.076	0.47 (0.33 - 0.65)	4.9E-06

P-values for deviation from Hardy-Weinberg equilibrium (HWE) were calculated using exact test. P-values for association between variant alleles and FLD status were calculated using logistic regression adjusted for age, gender, BMI, and ethnicity (in the all groups analysis).
*For variants with fewer than 5 carriers in cases and controls, unadjusted odds ratios and p-values were calculated using Fisher's exact.

Table S5: Distribution of Genetic Risk Score in the Primary Cohort Stratified by Ethnicity

	Mean ± SD			
Ethnicity	UTSW-FLD	Controls	OR (95% CI)	P-value
Net risk allele number				
All	2.55 ± 1.4	1.39 ± 1.46	1.55 (1.46 - 1.65)	7.3E-47
White	2.43 ± 1.46	1.6 ± 1.35	1.47 (1.34 - 1.63)	7.8E-16
Black	0.83 ± 1.41	0.55 ± 1.26	1.18 (0.96 - 1.47)	0.12
Hispanic	2.76 ± 1.25	1.9 ± 1.42	1.70 (1.51 - 1.93)	2.1E-20
GRS				
All	1.31 ± 0.85	0.44 ± 0.96	2.49 (2.24 - 2.78)	7.0E-72
White	1.05 ± 0.81	0.53 ± 0.66	2.36 (1.97 - 2.83)	2.6E-22
Black	0.11 ± 0.97	-0.22 ± 0.95	1.42 (1.07 - 1.90)	0.015
Hispanic	1.59 ± 0.71	0.9 ± 0.88	2.91 (2.39 - 3.57)	4.4E-31

Table S6. Baseline Characteristics of UTSW-FLD Patients and Age-, Gender- and Ethnic-Matched Controls.

Characteristic	No. of people with data	UTSW-FLD (n=1162)	Control* (n=1162)	P-value
Age, years, mean ± SD	2324	53.8 ± 11.9	53.4 ± 12	0.42
Female, n (%)	2324	654 (56.3)	654 (56.3)	1
Race/ethnicity, N (%)				
Non-Hispanic Black	2324	424 (36.5)	424 (36.5)	1
Non-Hispanic White	2324	57 (4.9)	57 (4.9)	1
Hispanic	2324	646 (55.6)	646 (55.6)	1
BMI, kg/m^2, mean ± SD	2324	32.5 ± 7.4	29 ± 6.2	<0.001
Glucose, mg/dL, median (IQR)	2160	111 (94 - 149)	93 (85 - 102)	<0.001
Type 2 Diabetes, N (%)	2323	491 (42.3)	134 (11.5)	<0.001
Systolic BP, mm Hg, mean ± SD	2302	125.9 ± 16.5	126.5 ± 19.6	0.0036
Diastolic BP, mm Hg, mean ± SD	2302	71.6 ± 11.5	76.4 ± 8.9	<0.001
Total cholesterol, mg/dL, mean ± SD	1990	162.2 ± 52.2	186.5 ± 39.5	<0.001
LDL-cholesterol, mg/dL, mean ± SD	1961	89.5 ± 38.9	108.8 ± 34.8	<0.001
HDL-cholesterol, mg/dL, mean ± SD	1984	43.2 ± 18.6	51.4 ± 14.9	<0.001
Triglycerides, mg/dL, median (IQR)	1990	117 (82 - 178)	111 (77 - 162)	0.30
ALT (U), median (IQR)	2305	37 (25 - 60)	19 (14 - 25)	<0.001
AST (U), median (IQR)	2302	45 (32 - 67)	20 (17 - 25)	<0.001

UTSW-FLD subjects matched to controls for age, gender, and ethnicity. Exact matching for gender and ethnicity, and matching within a caliper (<0.5 SD) for age. BP, blood pressure; TG, triglyceride; ALT, alanine aminotransaminase; AST, aspartate aminotransferase.

Table S7: Association of Individual Variants with FLD in the Matched Case-Control Cohort

			Cases (UTSW-FLD)					Controls						
			Genotype frequency			Allele	P	Genotype frequency			Allele	P	OR	P-value
Gene variant	SNP	Allele	REF	HET	HOM	frequency	(HWE)	REF	HET	HOM	frequency	(HWE)	(95% CI)	
All groups														
PNPLA3 I148M	rs738409	G	267	455	438	0.5737	-	522	481	158	0.3432	-	2.65 (2.31 - 3.06)	2.5E-46
TM6SF2 E167K	rs58542926	T	972	178	10	0.0853	-	1061	98	1	0.0431	-	2.07 (1.58 - 2.72)	7.4E-08
GCKR P446L	rs1260326	T	474	523	159	0.3638	-	509	520	133	0.3382	-	1.12 (0.98 - 1.28)	0.10
TMC4/MBOAT7	rs641738	T	417	545	193	0.4030	-	467	526	161	0.3674	-	1.17 (1.02 - 1.33)	0.020
PNPLA3 S453I	rs6006460	T	1143	8	0	0.0035	-	1050	24	2	0.0130	-	0.32 (0.13 - 0.7)	0.0033
HSD17B13 A192fs	rs80182459	-	1116	31	1	0.0144	-	1107	53	2	0.0245	-	0.53 (0.32 - 0.86)	0.010
HSD17B13 splice	rs72613567	TA	890	231	29	0.1257	-	819	293	47	0.1670	-	0.78 (0.65 - 0.94)	0.0075
White														
PNPLA3 I148M	rs738409	G	161	182	80	0.4043	0.034	266	143	14	0.2021	0.37	2.44 (1.91 - 3.13)	2.0E-13
TM6SF2 E167K	rs58542926	T	326	89	9	0.1262	0.37	376	46	0	0.0545	0.62	2.38 (1.6 - 3.57)	1.2E-05
GCKR P446L	rs1260326	T	123	219	81	0.4504	0.38	153	208	63	0.3939	0.61	1.32 (1.05 - 1.65)	0.015
TMC4/MBOAT7	rs641738	T	125	202	95	0.4645	0.43	146	196	78	0.4190	0.42	1.2 (0.97 - 1.49)	0.10
PNPLA3 S453I	rs6006460	T	422	1	0	0.0012	1	338	1	0	0.0015	1	0.8 (0.01 - 63.03)	1
HSD17B13 A192fs	rs80182459	-	418	1	0	0.0012	1	423	1	0	0.0012	1	1.01 (0.01 - 79.57)	1
HSD17B13 splice	rs72613567	TA	261	132	28	0.2233	0.050	226	163	35	0.2748	0.47	0.85 (0.67 - 1.08)	0.19
Black														
PNPLA3 I148M	rs738409	G	36	20	1	0.1930	0.67	42	15	0	0.1316	0.58	1.86 (0.85 - 4.23)	0.12
TM6SF2 E167K	rs58542926	T	54	3	0	0.0263	1	54	3	0	0.0263	1	1 (0.13 - 7.8)	1
GCKR P446L	rs1260326	T	48	8	1	0.0877	0.35	42	12	3	0.1579	0.12	0.6 (0.26 - 1.3)	0.20
TMC4/MBOAT7	rs641738	T	20	31	6	0.3772	0.40	28	19	7	0.3056	0.21	1.3 (0.73 - 2.36)	0.38
PNPLA3 S453I	rs6006460	T	50	6	0	0.0536	1	47	9	1	0.0965	0.41	0.58 (0.19 - 1.62)	0.31
HSD17B13 A192fs	rs80182459	-	40	14	1	0.1455	1	38	17	2	0.1842	1	0.7 (0.32 - 1.48)	0.36
HSD17B13 splice	rs72613567	TA	49	8	0	0.0702	1	52	5	0	0.0439	1	2.27 (0.68 - 8.22)	0.18

Hispanic														
PNPLA3 I148M	rs738409	G	59	238	348	0.7240	0.061	191	312	143	0.4628	0.48	2.76 (2.31 - 3.31)	2.7E-31
TM6SF2 E167K	rs58542926	T	566	77	1	0.0613	0.50	599	46	1	0.0372	0.60	1.82 (1.23 - 2.73)	0.0029
GCKR P446L	rs1260326	T	290	277	74	0.3315	0.53	296	289	61	0.3181	0.47	1.04 (0.87 - 1.25)	0.64
TMC4/MBOAT7	rs641738	T	264	296	81	0.3573	0.93	284	289	72	0.3357	0.93	1.12 (0.94 - 1.33)	0.22
PNPLA3 S453I	rs6006460	T	636	1	0	0.0008	1	634	11	0	0.0085	1	0.09 (0 - 0.63)	0.0061
HSD17B13 A192fs	rs80182459	-	623	16	0	0.0125	1	617	29	0	0.0224	1	0.56 (0.28 - 1.06)	0.076
HSD17B13 splice	rs72613567	TA	557	80	1	0.0643	0.50	516	117	10	0.1065	0.30	0.63 (0.46 - 0.85)	0.0027

P-values for deviation from Hardy-Weinberg equilibrium (HWE) were calculated using exact test. P-values for association between variant alleles and FLD status were calculated using logistic regression adjusted for age, gender, BMI, and ethnicity (in the all groups analysis).
*For variants with fewer than 5 carriers in either cases or controls, unadjusted odds ratios and p-values were calculated using Fisher's exact.

Table S8: Distribution of Genetic Risk Score in the Matched Case-Control Cohort Stratified by Ethnicity.

Ethnicity	Mean ± SD		OR (95% CI)	P-value
	UTSW-FLD	Matched Controls		
Net risk allele number				
All	2.56 ± 1.40	1.78 ± 1.42	1.51 (1.41 - 1.63)	7.0E-32
White	2.43 ± 1.47	1.54 ± 1.35	1.50 (1.33 - 1.70)	1.5E-11
Black	0.83 ± 1.41	0.59 ± 1.25	1.15 (0.85 - 1.57)	0.38
Hispanic	2.79 ± 1.23	2.04 ± 1.38	1.53 (1.39 - 1.69)	2.6E-19
GRS				
All	1.36 ± 0.84	0.80 ± 0.83	2.17 (1.96 - 2.42)	4.3E-53
White	1.10 ± 0.82	0.54 ± 0.63	2.18 (1.79 - 2.67)	1.8E-16
Black	0.23 ± 0.87	-0.05 ± 0.86	1.40 (0.96 - 2.11)	0.079
Hispanic	1.63 ± 0.71	1.03 ± 0.81	2.23 (1.94 - 2.57)	4.6E-34

Table S9: Association with FLD by Quintiles of Genetic Risk Score Stratified by Ethnicity in the Matched Case-Control Cohort

		N		OR	P-value	P-trend
Ethnicity	Quintile	Cases	Controls	(95% CI)		
ALL	1	98	245	1	-	
	2	111	184	1.63 (1.13 - 2.35)	0.0093	
	3	204	258	2.43 (1.74 - 3.41)	1.9E-07	
	4	195	171	3.62 (2.55 - 5.19)	1.2E-12	
	5	517	208	8.7 (6.25 - 12.23)	1.1E-36	9.9E-44
White	1	37	68	1	-	
	2	45	69	1.05 (0.57 - 1.94)	0.88	
	3	54	69	1.42 (0.78 - 2.6)	0.25	
	4	76	63	2.05 (1.15 - 3.69)	0.015	
	5	201	67	4.57 (2.69 - 7.9)	3.0E-08	2.1E-11
Black	1	8	15	1	-	
	2	6	7	1.34 (0.3 - 5.92)	0.70	
	3	16	13	2.28 (0.71 - 7.8)	0.17	
	4	6	8	1.45 (0.35 - 6.04)	0.60	
	5	18	11	3.22 (1.01 - 11)	0.053	0.064
Hispanic	1	29	131	1	-	
	2	75	149	2.05 (1.23 - 3.49)	0.0064	
	3	79	116	2.82 (1.69 - 4.8)	1.0E-04	
	4	181	127	6.16 (3.83 - 10.17)	2.7E-13	
	5	260	118	8.72 (5.45 - 14.32)	1.2E-18	1.3E-27

Association was tested using logistic regression adjusted for age, gender, BMI, ethnicity (in the all groups analysis), and T2D.

Table S10: Distribution of Genetic Risk Score Based on External Weights in the Primary Cohort Stratified by Ethnicity

	Mean ± SD			
Ethnicity	UTSW-FLD	Controls	OR (95% CI)	P-value
GRS				
All	1.07 ± 0.64	0.47 ± 0.61	2.17 (1.98 - 2.38)	1.5E-70
White	0.88 ± 0.65	0.47 ± 0.54	1.98 (1.72 - 2.3)	3.4E-22
Black	0.28 ± 0.56	0.13 ± 0.51	1.35 (0.99 - 1.82)	0.0597
Hispanic	1.26 ± 0.55	0.76 ± 0.62	2.51 (2.12 - 3)	1.4E-30

Table S11: Association with FLD by Quintiles of Genetic Risk Score Based on External Weights in the Primary Cohort Stratified by Ethnicity

		N		OR	P-value	P-trend
Ethnicity	Quintile	Cases	Controls	(95% CI)		
ALL	1	72	1018	1	-	
	2	60	564	0.83 (0.56 - 1.24)	0.37	
	3	139	612	1.49 (1.06 - 2.1)	0.023	
	4	340	641	2.75 (2.02 - 3.78)	2.0E-10	
	5	546	272	7.43 (5.43 - 10.28)	5.8E-35	3.9E-54
White	1	46	222	1	-	
	2	37	122	1.18 (0.69 - 2.01)	0.54	
	3	66	204	1.3 (0.83 - 2.06)	0.26	
	4	79	142	2.6 (1.65 - 4.11)	4.0E-05	
	5	190	169	4.47 (2.98 - 6.8)	1.0E-12	1.4E-16
Black	1	8	356	1	-	
	2	13	443	1.3 (0.54 - 3.34)	0.57	
	3	8	341	1.06 (0.38 - 2.92)	0.92	
	4	7	280	1.17 (0.4 - 3.33)	0.77	
	5	18	340	2.29 (1.01 - 5.7)	0.057	0.0647
Hispanic	1	28	95	1	-	
	2	51	73	2.69 (1.41 - 5.2)	0.0030	
	3	113	94	4.06 (2.27 - 7.45)	3.6E-06	
	4	116	70	7.16 (3.95 - 13.36)	2.2E-10	
	5	338	83	15.99 (9.13 - 28.94)	3.8E-21	6.1E-27

Association was tested using logistic regression adjusted for age, gender, BMI, ethnicity (in the all groups analysis), and T2D.

Table S12: Distribution of Genetic Risk Score without PNPLA3 I148M Variant, Stratified by Ethnicity

	Mean ± SD			
Ethnicity	UTSW-FLD	Controls	OR (95% CI)	P-value
Net risk allele number				
All	1.41 ± 1.2	0.84 ± 1.22	1.31 (1.22 - 1.4)	1.7E-14
White	1.63 ± 1.3	1.17 ± 1.2	1.3 (1.17 - 1.44)	1.4E-06
Black	0.43 ± 1.19	0.28 ± 1.15	1.12 (0.88 - 1.43)	0.34
Hispanic	1.33 ± 1.07	1.01 ± 1.13	1.35 (1.18 - 1.55)	7.9E-06
GRS				
All	0.22 ± 0.4	-0.08 ± 0.61	1.83 (1.61 - 2.09)	3.4E-22
White	0.28 ± 0.43	0.12 ± 0.36	1.76 (1.43 - 2.18)	1.2E-07
Black	-0.28 ± 0.73	-0.48 ± 0.78	1.27 (1 - 1.64)	0.048
Hispanic	0.22 ± 0.32	0.05 ± 0.45	2.24 (1.72 - 2.97)	3.6E-10

Table S13: Association with FLD by Quintiles of Genetic Risk Score without PNPLA3 I148M Variant, Stratified by Ethnicity

		N		OR	P-value	P-trend
Ethnicity	Quintile	Cases	Controls	(95% CI)		
ALL	1	73	1011	1	-	
	2	246	692	2.04 (1.49 - 2.84)	1.45E-05	
	3	180	457	2.27 (1.62 - 3.2)	2.46E-06	
	4	332	606	2.24 (1.64 - 3.09)	7.03E-07	
	5	327	341	3.84 (2.78 - 5.35)	9.08E-16	7.3E-14
White	1	68	192	1	-	
	2	51	156	0.93 (0.59 - 1.47)	0.763	
	3	82	188	1.12 (0.74 - 1.71)	0.588	
	4	74	170	1.08 (0.7 - 1.65)	0.727	
	5	144	153	2.3 (1.55 - 3.42)	3.9E-05	2.8E-05
Black	1	6	365	1	-	
	2	8	353	1.42 (0.49 - 4.39)	0.521	
	3	21	463	2.97 (1.25 - 8.21)	0.0213	
	4	11	266	2.62 (0.97 - 7.73)	0.0636	
	5	8	313	1.61 (0.55 - 4.98)	0.385	0.19
Hispanic	1	57	95	1	-	
	2	219	132	2.73 (1.73 - 4.34)	1.8E-05	
	3	94	51	2.6 (1.5 - 4.55)	7.1E-04	
	4	127	68	3.06 (1.83 - 5.17)	2.2E-05	
	5	149	69	3.8 (2.29 - 6.36)	2.8E-07	1.5E-05

Association was tested using logistic regression adjusted for age, gender, BMI, ethnicity (in the all groups analysis), and T2D.

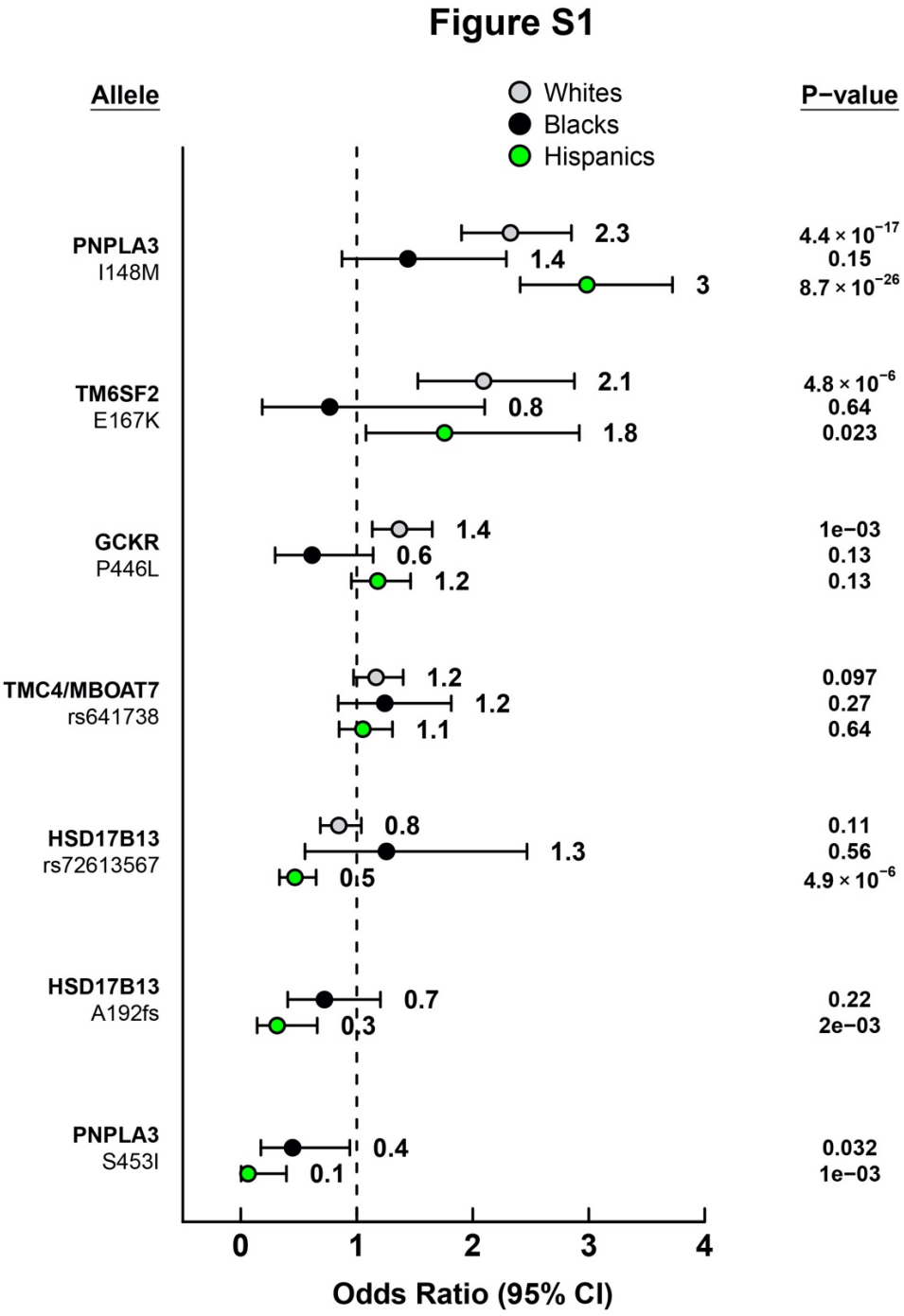

Fig S1. Ethnicity specific comparison of risk allele frequencies between UTSW-FLD Cohort participants and controls

Odds ratios were calculated by logistic regression analysis. Error bars are 95% confidence intervals.

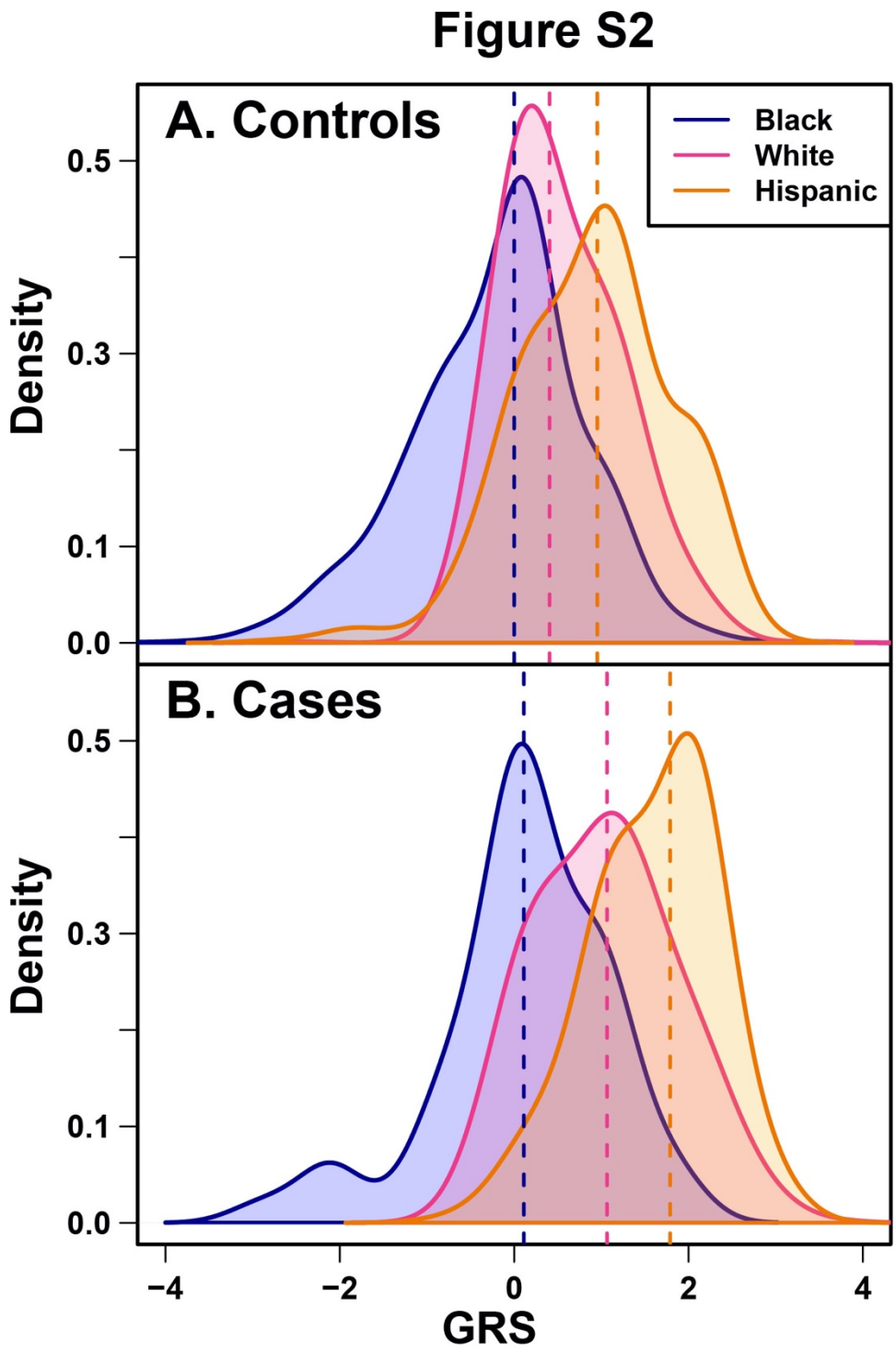

Fig S2. Distribution of GRS among cases and controls stratified by ethnicity.

Dashed lines denote the means.

References

1. 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*. Apr 2014;46(4):352-6. doi:10.1038/ng.2901
2. Kozlitina J. Genetic Risk Factors and Disease Modifiers of Nonalcoholic Steatohepatitis. *Gastroenterol Clin North Am*. Mar 2020;49(1):25-44. doi:10.1016/j.gtc.2019.09.001
3. Tian C, Stokowski RP, Kershenobich D, Ballinger DG, Hinds DA. Variant in PNPLA3 is associated with alcoholic liver disease. *Nat Genet*. Jan 2010;42(1):21-3. doi:10.1038/ng.488
4. Buch S, Stickel F, Trepo E, et al. A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. *Nat Genet*. Dec 2015;47(12):1443-8. doi:10.1038/ng.3417
5. Gellert-Kristensen H, Richardson TG, Davey Smith G, Nordestgaard BG, Tybjaerg-Hansen A, Stender S. Combined Effect of PNPLA3, TM6SF2, and HSD17B13 Variants on Risk of Cirrhosis and Hepatocellular Carcinoma in the General Population. *Hepatology*. Sep 2020;72(3):845-856. doi:10.1002/hep.31238
6. Abul-Husn NS, Cheng X, Li AH, et al. A Protein-Truncating HSD17B13 Variant and Protection from Chronic Liver Disease. *N Engl J Med*. Mar 22 2018;378(12):1096-1106. doi:10.1056/NEJMoa1712191
7. Palmer ND, Kahali B, Kuppa A, et al. Allele-specific variation at APOE increases nonalcoholic fatty liver disease and obesity but decreases risk of Alzheimer's disease and myocardial infarction. *Hum Mol Genet*. Jul 9 2021;30(15):1443-1456. doi:10.1093/hmg/ddab096
8. Romeo S, Kozlitina J, Xing C, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet*. Dec 2008;40(12):1461-5. doi:10.1038/ng.257