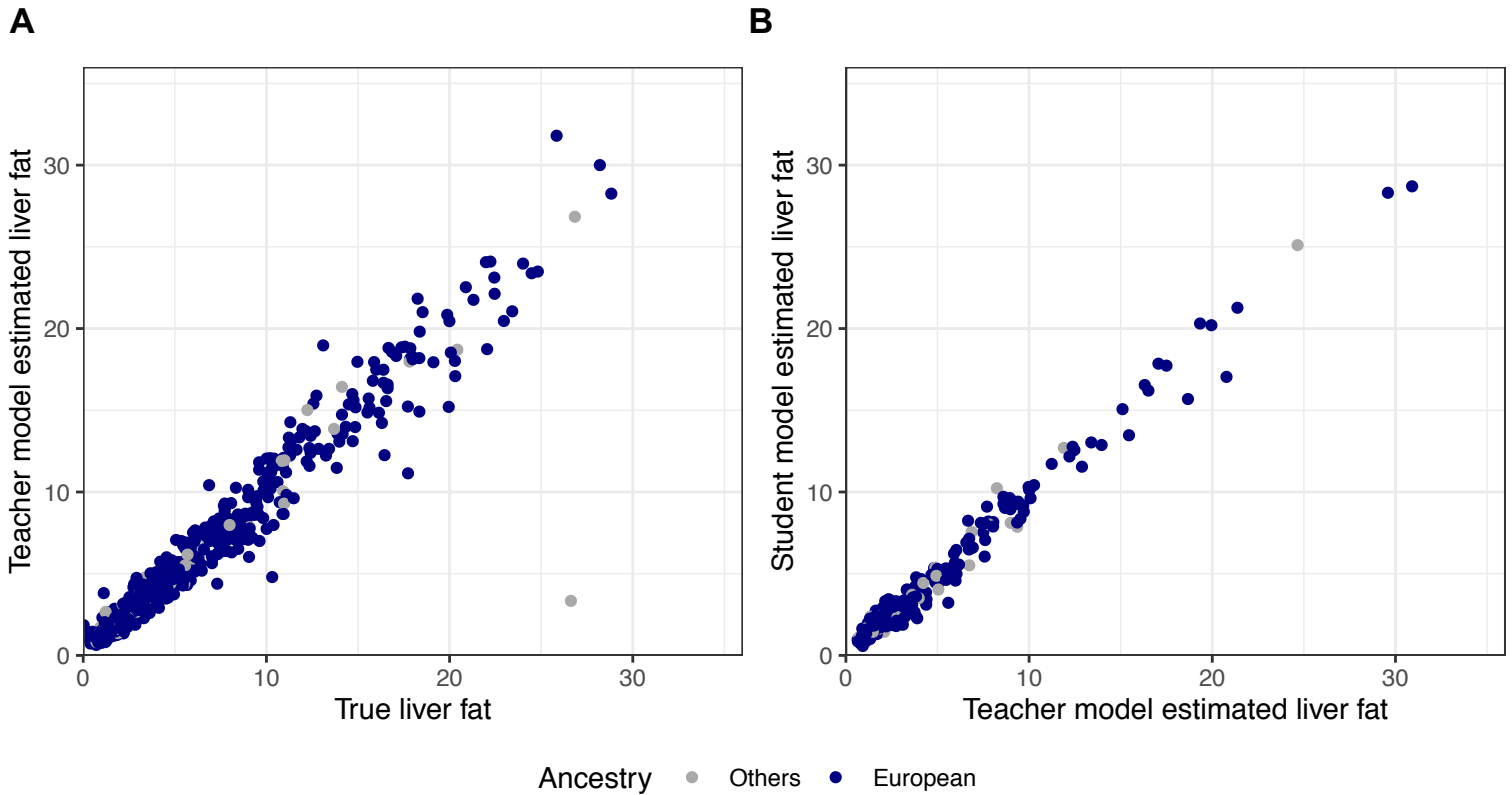
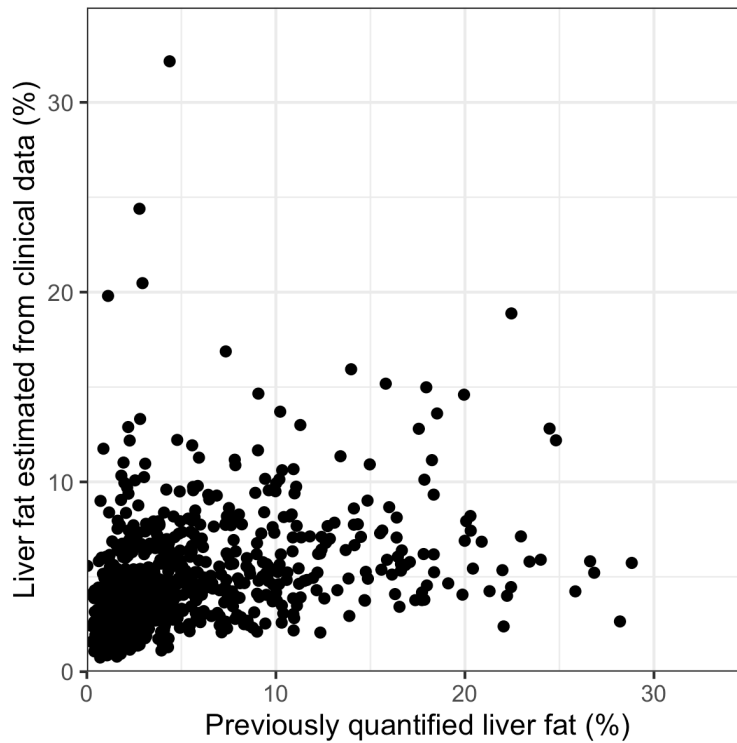


# Supplementary Figure 1



**Supplemental Figure 1. Comparison of truth data with teacher-model inferences, and teacher-model inferences with student-model inferences. A)** In a held-out set of 1,214 participants with truth data, the Pearson correlation between the previously-quantified liver fat ('True liver fat') and liver fat inferred from the ML teacher model was 0.974 (95% CI 0.971-0.977;  $P < 2.4 \times 10^{-784}$ ). **B)** In a separate held-out set of 390 samples with both gradient-echo and IDEAL imaging who were not used for model creation, the Pearson correlation between the teacher model inferred liver fat and the student model inferred liver fat was 0.992 (95% CI 0.990-0.993;  $P = 3.1 \times 10^{-351}$ ). Blue points indicate European ancestries (whether the sample was included in the GWAS, as defined in the methods), gray points indicate all other ancestries.

## Supplementary Figure 2



**Supplemental Figure 2. Prediction of liver fat using a variable beta regression model of clinical and anthropometric measurements.** In a held-out set of 1,196 participants of all ancestries with truth data, the Pearson correlation between the previously-quantified liver fat and liver fat estimated from clinical data was 0.453 (95% CI 0.407-0.497;  $P=9.9 \times 10^{-62}$ ). Measurements that were at least nominally associated with liver fat in univariable analysis and therefore included in the beta regression model were: BMI, waist circumference, hip circumference, total body fat mass, total body fat percent, age at baseline, sex, height, weight, trunk fat mass, trunk fat percent, WHR, LDL cholesterol, total cholesterol, HDL cholesterol, triglycerides, systolic blood pressure, alkaline phosphatase, alanine aminotransferase (ALT), aspartate aminotransferase (AST), ALT/AST, gamma glutamyltransferase, hemoglobin A1c, random glucose, and C-reactive protein. Lipid measures were adjusted for lipid-lowering medication use and blood pressure was adjusted for anti-hypertensive medication use, as previously described ((Ehret et al., 2016; Patel et al., 2020).

**Supplemental Table 1. Baseline characteristics of participants in UK Biobank who underwent abdominal imaging.**

	<b>Overall (N=36,703)</b>
Female	19049 (51.9%)
Age at enrollment, years	54.9 (7.47)
Age at imaging, years	64.2 (7.56)
Race	
White	35572 (96.9%)
Black	214 (0.6%)
South Asian	313 (0.9%)
Other Asian	165 (0.4%)
Multiple or other	336 (0.9%)
Coronary artery disease	1283 (3.5%)
Diabetes	1956 (5.3%)
Obese or severely obese	6495 (17.7%)
Hypertension	11085 (30.2%)
Excessive alcohol intake	2015 (5.5%)
Medications	
Blood pressure medications	4940 (13.5%)
Lipid-lowering therapy	5552 (15.1%)
Anthropometric data	
Weight, kg	76.8 (14.8)
Waist-to-hip ratio	0.859 (0.0871)
Body mass index, kg/m <sup>2</sup>	26.6 (4.19)
Body fat, %	30.0 (8.17)
Systolic blood pressure	135 (17.6)
Weekly drinks, American standard	5.48 (6.37)
Liver-associated biomarkers	
Alanine aminotransferase (IU/L)	23.0 (13.9)
Aspartate aminotransferase (IU/L)	25.8 (10.5)
Gamma glutamyltransferase (IU/L)	33.7 (33.9)
Lipid biomarkers	
Total cholesterol, mg/dl	221 (42.0)
LDL cholesterol, mg/dl	138 (32.1)
HDL cholesterol, mg/dL	57.0 (14.5)
Triglycerides, mg/dL	123 [88-179]
Glycemic biomarkers	
Glycated hemoglobin, %	5.36 (0.475)
Glucose, mg/dL	89.9 (17.5)
Liver fat at imaging, %	2.22 [1.57-4.17]

Values correspond to number (%), mean (standard deviation), or median [interquartile range].

**Supplementary Table 2. Definitions of disease in each cohort.**

Cohort	Disease	Definition
UK Biobank	NAFLD	Hospitalization due to nonalcoholic fatty liver (ICD10 K76.0) or other specified inflammatory liver diseases including nonalcoholic steatohepatitis (ICD10 K75.8)
UK Biobank	NASH	Hospitalization due to other specified inflammatory liver diseases including nonalcoholic steatohepatitis (ICD10 K75.8)
UK Biobank	NAFLD/NASH	Hospitalization due to nonalcoholic fatty liver (ICD10 K76.0) or other specified inflammatory liver diseases including nonalcoholic steatohepatitis (ICD10 K75.8)
UK Biobank	Cirrhosis	Hospitalization or death due to (ICD10; ICD9 codes): cirrhosis (K74.6; 5715), alcoholic cirrhosis or liver damage (K70.3; 5712, 5713), oesophageal varices (I85.0,I85.9; 4560, 4561) or portal hypertension (K76.6; 5723)
UK Biobank	Hepatocellular Carcinoma	Hospitalization due to liver cell carcinoma (ICD10 C22.0)
UK Biobank	Coronary Artery Disease	Self report of heart attack diagnosed by doctor (df-6150), self-reported myocardial infarction, coronary angioplasty or coronary artery bypass graft or triple heart bypass during verbal interview with trained nurse (df-20002, df-20004), hospitalization or death due to myocardial infarction (ICD10 I21, I21.0-4, I21.9, I22, I22.0, I22.1, I22.8, I22.9, I23, I23.0-6, I23.8; ICD9 410, 4109, 412, 4129), ischemic heart disease ICD10 (I24, I24.0, I24.1, I24.8, I24.9, I25.2; ICD9 411, 4119); operative procedures: replacement of coronary artery(ies)(K40, K40.1-4, K40.8-9, K41, K41.1-4, K41.8-9, K42,K42.1-4,K42.8-9, K43, K43.1-4, K43.8-9, K44, K44.1-4, K44.8-9), connection of thoracic artery to coronary artery (K45.1-K45.6, K45.8-9), other bypass of coronary artery(ies) (K46, K46.1-K46.6, K46.8-9) percutaneous transluminal balloon angioplasty of coronary artery(ies) (K49.1-4,K49.8-9), transluminal operations on coronary artery (K50.1-2, K50.4), percutaneous transluminal balloon angioplasty and insertion of stent into coronary artery (K75.1-4,K75.8-9)
UK Biobank	Diabetes	Self report of diabetes diagnosed by doctor (df-2443), self-reported diabetes during verbal interview with trained nurse (df-20002), started insulin within one year diagnosis of diabetes, self-reported insulin medication, hospitalization or death due to diabetes (ICD10 E10, E10.0-9, E11, E11.0-9, E12, E12.1, E12.8, E12.9, E13, E13.1-9, E14, E14.0-9; ICD9 2500, 25001, 25009, 2501, 25011, 25019, 2503, 2504, 2505, 25099)
UK Biobank	Hypertension	Self-report of high blood pressure diagnosed by doctor (df-6150), self-reported hypertension, or essential hypertension during verbal interview with trained nurse (df-20002); hospitalization or death due to (ICD10; ICD9 codes) essential hypertension (I10; 401, 4010, 4011, 4019), hypertensive heart disease (I11, I11.0, I11.9, I13, I13.0-13.2; 402, 4020, 4021, 4029, 404, 4040, 4041, 4049), hypertensive renal disease (I12, I12.0, I12.9, I13, I13.0-13.2; 403, 4030, 4031, 4039, 404, 4040, 4041, 4049), secondary hypertension or renovascular hypertension (I15, I15.0-15.2, I15.8, I15.9; 405, 4050, 4051, 4059)
Mass General Brigham Biobank	NAFLD/NASH	Hospitalization due to nonalcoholic fatty liver (ICD10 K76.0) or nonalcoholic steatohepatitis (ICD10 K75.81)

**Supplementary Table 3. Effects of variants previously associated with quantitative liver fat in common variant association study, and cirrhosis-associated variants nominated to affect liver fat, on liver fat in 32,976 individuals in UK Biobank.**

Variant	Chr.	Position (hg19)	Effect Allele	Other Allele	Allele Freq.	Effect on liver fat (SD)	P value	Locus	Previous reference(s)
rs1260326	2	27730940	T	C	0.39	0.04	4.10E-07	<i>GCKR</i>	(Palmer et al., 2013; Parisinos et al., 2020)
rs780094	2	27741237	T	C	0.38	0.04	3.50E-06	<i>GCKR</i>	(Speliotes et al., 2011; Palmer et al., 2013)
rs2228603	19	19329924	C	T	0.92	-0.22	1.80E-51	<i>NCAN</i>	(Speliotes et al., 2011; Palmer et al., 2013)
rs12137855	1	219448378	C	T	0.79	0.01	0.46	<i>LYPLAL1</i>	(Speliotes et al., 2011; Palmer et al., 2013)
rs4240624	8	9184231	G	A	0.09	-0.02	0.10	<i>PPP1R3B</i>	(Speliotes et al., 2011; Palmer et al., 2013)
rs641738	19	54676763	C	T	0.56	-0.03	8.80E-06	<i>MBOAT7-TMC4</i>	(Buch et al., 2015; Mancina et al., 2016)
rs62305723	4	88231429	G	A	0.93	0.01	0.59	<i>HSD17B13</i>	(Ma et al., 2019)
rs6834314	4	88213808	A	G	0.72	0.01	0.32	<i>HSD17B13</i>	(Abul-Husn et al., 2018; Ma et al., 2019; Gellert-Kristensen et al., 2020)
rs72613567	4	88231392	T	TA	0.73	0.01	0.40	<i>HSD17B13</i>	(Abul-Husn et al., 2018; Ma et al., 2019; Gellert-Kristensen et al., 2020)
rs11597086	10	101953705	A	C	0.55	0.02	0.01	<i>CHUK</i>	(Feitosa et al., 2013)
rs10883451	10	101924418	T	C	0.50	0.02	0.02	<i>ERLIN1</i>	(Feitosa et al., 2013)

Effects on inverse-normal transformed liver fat are given in units of standard deviation (SD).

**Supplementary Table 4. Effects of CVAS top variants after adjusting for alcohol consumption.**

Lead Variant	Chr.	Position (hg19)	Nearest Gene	Effect Allele	Other Allele	Effect Allele Freq.	Unadjusted for alcohol consumption (original CVAS, n=32976)			Former and excessive alcohol consumers excluded (n=30197)			Former alcohol consumers excluded, adjusted for number of weekly drinks (n=32043)		
							Effect on liver fat (SD)	SE (SD)	P value	Effect on liver fat (SD)	SE (SD)	P value	Effect on liver fat (SD)	SE (SD)	P value
<b>Newly-Identified Variants</b>															
rs2642438	1	220970028	<i>MARC1</i>	G	A	0.70	0.052	0.008	1.70E-09	0.046	0.009	4.20E-07	0.050	0.009	8.70E-09
rs1229984	4	100239319	<i>ADH1B</i>	C	T	0.98	0.158	0.025	7.00E-10	0.147	0.026	1.40E-08	0.153	0.026	3.40E-09
rs112875651	8	126506694	<i>TRIB1</i>	G	A	0.61	0.050	0.008	3.80E-10	0.052	0.008	3.50E-10	0.053	0.008	3.70E-11
rs2250802	10	113921354	<i>GPAM</i>	G	A	0.27	0.054	0.009	1.40E-09	0.056	0.009	2.10E-09	0.056	0.009	3.70E-10
rs56252442	19	18229208	<i>MAST3</i>	T	G	0.25	0.049	0.009	2.70E-08	0.051	0.009	3.20E-08	0.050	0.009	1.70E-08
<b>Previously-Reported Variants</b>															
rs58542926	19	19379549	<i>TM6SF2</i>	T	C	0.07	0.289	0.015	2.80E-85	0.283	0.015	1.60E-75	0.288	0.015	6.10E-83
rs429358	19	45411941	<i>APOE</i>	T	C	0.85	0.121	0.011	1.50E-29	0.114	0.011	4.80E-24	0.120	0.011	3.90E-28
rs738408	22	44324730	<i>PNPLA3</i>	T	C	0.21	0.195	0.009	5.30E-95	0.191	0.010	3.10E-84	0.194	0.010	1.10E-92

Excessive alcohol consumption was defined as >14 weekly drinks for women and >21 weekly drinkers for men. Chr., chromosome; Freq., frequency; SD, standard deviation; SE, standard error.

**Supplementary Table 5. Replication of eight loci associated with liver fat in the discovery CVAS of liver fat in UK Biobank in two additional cohorts.**

						UK Biobank Discovery CVAS				Framingham Heart Study Replication						MESA Replication				Combined Framingham + MESA Replication		
Lead Variant	Chr.	Position (hg19)	Nearest Gene	Effect Allele	Other Allele	Effect Allele	Effect on MRI liver fat (SD)	SE (SD)	P value	Effect Allele	Other Allele	Effect Allele	Effect on CT liver fat (SD)	SE (SD)	P value	Effect Allele	Effect on CT liver fat (SD)	SE (SD)	P value	Effect on CT liver fat (SD)	SE (SD)	P value
						Freq.				Freq.				Freq.				Freq.				
<b>Newly-identified variants</b>																						
rs2642438	1	220970028	MARC1	G	A	0.70	0.052	0.008	1.70E-09	G	A	0.71	0.088	0.028	1.60E-03	G	0.062	0.026	0.017	0.074	0.019	1.04E-04
rs1229984	4	100239319	ADH1B	C	T	0.98	0.158	0.025	7.00E-10	C	T	0.95	0.101	0.057	0.081	C	0.034	0.040	0.394	0.056	0.033	0.088
rs112875651	8	126506694	TRIB1	G	A	0.61	0.050	0.008	3.80E-10	G	A	0.60	0.070	0.026	7.70E-03	G	0.044	0.023	0.059	0.055	0.017	1.40E-03
rs2250802	10	113921354	GPAM	G	A	0.27	0.054	0.009	1.40E-09	G	A	0.30	0.058	0.028	0.037	G	0.031	0.025	0.218	0.043	0.019	0.021
rs56252442	19	18229208	MAST3	T	G	0.25	0.049	0.009	2.70E-08	T	G	0.19	0.065	0.032	0.041	T	0.017	0.026	0.501	0.036	0.020	0.070
<b>Previously-identified variants</b>																						
rs429358	19	45411941	APOE	T	C	0.85	0.121	0.011	1.50E-29	T	C	0.97	-0.119	0.075	0.110	T	0.056	0.029	0.054	0.033	0.027	0.219
rs58542926	19	19379549	TM6SF2	T	C	0.07	0.289	0.015	2.80E-85	T	C	0.05	0.326	0.056	6.20E-09	T	0.229	0.045	3.36E-07	0.267	0.035	2.25E-14
rs738408	22	44324730	PNPLA3	T	C	0.21	0.195	0.009	5.30E-95	T	C	0.24	0.278	0.030	2.50E-20	T	0.294	0.024	1.35E-33	0.288	0.019	9.60E-53

Liver fat was measured by magnetic resonance imaging (MRI) in UK Biobank and by computed tomography (CT) imaging in the Framingham Heart Study (n=3284) and Multi-Ethnic Study of Atherosclerosis (MESA, n=4195). Effects of Framingham and MESA were combined via fixed-effects inverse variance weighted meta-analysis. Chr., chromosome; Freq., frequency; SD, standard deviation; SE, standard error.

**Supplementary Table 6. Effects of 8 common loci associated with quantitative liver fat on blood biomarkers of liver injury.**

Lead Variant	Chr.	Position (hg19)	Nearest Gene	Effect Allele	Other Allele	Effect Allele Freq.	Effect on liver fat (SD)	SE (SD)	P value	Effect on ALT (U/L)	SE (U/L)	P value	Effect on AST (U/L)	SE (U/L)	P value
<b>Newly-Identified Variants</b>															
rs2642438	1	220970028	<i>MARC1</i>	G	A	0.70	0.052	0.008	2E-09	0.47	0.036	1E-39	0.16	0.028	5E-09
rs1229984	4	100239319	<i>ADH1B</i>	C	T	0.98	0.158	0.025	7E-10	0.41	0.112	2E-04	0.32	0.086	2E-04
rs112875651	8	126506694	<i>TRIB1</i>	G	A	0.61	0.050	0.008	4E-10	0.56	0.034	2E-61	0.19	0.026	2E-13
rs2250802	10	113921354	<i>GPAM</i>	G	A	0.27	0.054	0.009	1E-09	0.41	0.037	3E-29	0.16	0.028	3E-08
rs56252442	19	18229208	<i>MAST3</i>	T	G	0.25	0.049	0.009	3E-08	0.29	0.038	1E-14	0.03	0.029	0.342
<b>Previously-Reported Variants</b>															
rs58542926	19	19379549	<i>TM6SF2</i>	T	C	0.07	0.289	0.015	3E-85	1.28	0.062	2E-94	0.67	0.048	2E-44
rs429358	19	45411941	<i>APOE</i>	T	C	0.85	0.121	0.011	2E-29	0.66	0.045	7E-49	0.07	0.035	0.045
rs738408	22	44324730	<i>PNPLA3</i>	T	C	0.21	0.195	0.009	5E-95	1.68	0.040	<1E-300	1.06	0.031	8E-262

Effects on alanine aminotransferase (ALT, n=345,930 with ALT measurement) and aspartate aminotransferase (AST, n=344,799 with AST measurement) were measured in the n=362,910 subset of the UK Biobank who did not undergo abdominal imaging. Chr., chromosome; Freq., frequency; SD, standard deviation; SE, standard error.



**Supplementary Table 7. Effects of 8 common loci associated with quantitative liver fat on clinical diagnoses of liver disease.**

Lead Variant	Chr.	Position (hg19)	Nearest Gene	Effect Allele	Other Allele	Effect Allele Freq.	UK Biobank			Mass General Brigham Biobank			Combined UK Biobank + Mass General Brigham Biobank		
							Effect on NAFLD/NASH diagnosis (OR)	SE (logOR)	P value	Effect on NAFLD/NASH diagnosis (OR)	SE (logOR)	P value	Effect on NAFLD/NASH diagnosis (OR)	SE (logOR)	P value
<b>Newly-identified variants</b>															
rs2642438	1	220970028	MARC1	G	A	0.70	1.14	0.034	1.75E-04	1.06	0.027	0.03	1.09	0.021	4.44E-05
rs1229984	4	100239319	ADH1B	C	T	0.98	1.33	0.118	0.015	1.26	0.052	6.69E-06	1.28	0.048	3.36E-07
rs112875651	8	126506694	TRIB1	G	A	0.61	1.10	0.032	1.71E-03	1.07	0.025	6.44E-03	1.08	0.020	4.35E-05
rs2250802	10	113921354	GPAM	G	A	0.27	1.08	0.033	0.016	1.07	0.026	7.51E-03	1.08	0.021	3.24E-04
rs56252442	19	18229208	MAST3	T	G	0.25	1.00	0.035	0.991	1.04	0.028	0.20	1.02	0.022	0.32
<b>Previously-identified variants</b>															
rs429358	19	45411941	APOE	T	C	0.85	1.24	0.045	1.39E-06	1.18	0.036	5.24E-06	1.20	0.028	4.83E-11
rs58542926	19	19379549	TM6SF2	T	C	0.07	1.41	0.050	5.86E-12	1.38	0.044	7.76E-13	1.39	0.033	3.13E-23
rs738408	22	44324730	PNPLA3	T	C	0.21	1.52	0.033	6.00E-37	1.38	0.027	8.52E-32	1.43	0.021	9.80E-66

Effects on NAFLD/NASH in a subset of the UK Biobank who did not undergo abdominal imaging (2,225 cases vs 360,685 controls) and NAFLD/NASH in Mass General Brigham Biobank (4,129 cases vs 26,444 controls) were measured. Effects of UK Biobank and Mass General Brigham Biobank were combined via fixed-effects inverse variance weighted meta-analysis. Chr., chromosome; Freq., frequency; OR, odds ratio; SE, standard error.

**Supplementary Table 8. Rare inactivating variants in *APOB* are associated with decreased circulating lipids but increased markers of liver injury.**

	<i>APOB</i> inactivating variant			<i>MTTP</i> inactivating variant		
	Controls (N=42196)	carriers (N=64) <sup>a</sup>	Adjusted P	Controls (N=42214) <sup>b</sup>	carriers (N=46) <sup>c</sup>	Adjusted P
Apolipoprotein B, mean (SD), g/L	1.03 (0.235)	0.619 (0.283)	1.77E-29	1.03 (0.236)	1.05 (0.212)	0.66
LDL cholesterol, mean (SD), mg/dL	146 (32.8)	81.2 (37.1)	4.48E-56	146 (32.9)	154 (33.8)	0.13
Triglycerides, mean (SD), mg/dL	157 (93.6)	94.5 (75.6)	8.86E-08	157 (93.6)	161 (59.6)	0.91
Alanine aminotransferase, mean (SD), U/L	23.5 (13.5)	30.7 (18.1)	1.73E-05	23.5 (13.5)	27.7 (17.4)	0.05
Aspartate aminotransferase, mean (SD), U/L	26.3 (10.3)	29.9 (12.5)	6.45E-03	26.3 (10.3)	28.4 (11.2)	0.21
Coronary artery disease, N (%)	2727 (6.5%)	1 (1.6%)	0.15	2726 (6.5%)	2 (4.3%)	0.45

<sup>a</sup>Apolipoprotein B values were missing in 25/64 inactivating variant carriers (39%), all other measurements had <=6% missing in either group.

<sup>b</sup>Measurements were missing in 6.2-6.5% of controls.

<sup>c</sup>Measurements were missing in 4/46 inactivating variant carriers (8.7%).