Cell Genomics, Volume *1*

Supplemental information

Machine learning enables new insights into genetic contributions to liver fat accumulation

Mary E. Haas, James P. Pirruccello, Samuel N. Friedman, Minxian Wang, Connor A. Emdin, Veeral H. Ajmera, Tracey G. Simon, Julian R. Homburger, Xiuqing Guo, Matthew Budoff, Kathleen E. Corey, Alicia Y. Zhou, Anthony Philippakis, Patrick T. Ellinor, Rohit Loomba, Puneet Batra, and Amit V. Khera

Table of Contents

Table S1. Baseline characteristics of participants in UK Biobank stratified by inclusion in the imaging substudy, Related to Table 1. 3

Table S2. Definitions of disease in each cohort, Related to STAR methods. 4

Table S3. 95% credible sets after fine-mapping of lead variant at eight common GWAS loci associated with quantitative liver fat, Related to Table 2. 5

Table S4. Effects of eight common variants associated with quantitative liver fat in the n=4,040 individuals with previously-quantified liver fat compared to the n=32,974 individuals with liver fat quantification enabled by machine learning, Related to Table 2. 6

Table S5. Effect of variants previously associated with liver fat, related liver disease, or cirrhosis in common variant association study, on liver fat in 32,974 individuals in UK Biobank, Related to Table 2. 7

Table S6. Effects of eight common liver fat variants on quantitative liver fat after adjusting for alcohol consumption, Related to Table 2. 8

Table S7. Replication of eight common UK Biobank variants associated with quantitative liver fat in two additional liver fat cohorts, Related to Table 2. 9

Table S8. Effects of eight common variants associated with quantitative liver fat on blood biomarkers of liver injury, Related to Table 2. 10

Table S9. Effects of eight common variants associated with quantitative liver fat on clinical diagnoses of liver disease, Related to Table 2. 11

Table S10. Comparison of C statistics for discrimination of liver diseases with addition of an eight-variant polygenic score, Related to Figure 4. 12

Table S11. Rare inactivating variants in *APOB* are associated with decreased circulating lipids but increased markers of liver injury, Related to Figure 5. 13

Table S12. Characteristics of UK Biobank participants with hepatic steatosis stratified by presence of an inactivating variant in *APOB* or *MTTP*, Related to Figure 5. 14

Figure S1. Comparison of previously-quantified liver fat with teacher-model inferences, and teacher-model inferences with student-model inferences, Related to STAR methods. 15

Figure S2. Assessment of machine learning model overfitting via model loss, Related to STAR methods. 16

Figure S3. Prediction of liver fat using a variable dispersion beta regression model of clinical and anthropometric measurements, Related to STAR methods. 17

Figure S4. Quantile-quantile plot for common variant genome-wide association study analysis of liver fat in 32,974 UK Biobank participants, Related to Figure 3. 18

Figure S5. Relationship between percent liver fat and inverse normal transformation of percent liver fat among 32,974 UK Biobank participants, Related to Figures 3-5. 19

Supplemental References 20

Table S1. Baseline characteristics of participants in UK Biobank stratified by inclusion in the imaging substudy, Related to Table 1.

	Overall (N=502521)	Imaged (N=36703)	Not Imaged (N=465818)	P-value
Female	273394 (54.4%)	19049 (51.9%)	254345 (54.6%)	$1.5x10^{-23}$
Age at enrollment, years	56.5 (8.10)	54.9 (7.47)	56.7 (8.13)	$<1x10^{-300}$
Age at imaging, years	NA	64.2 (7.56)	NA	NA
Self-reported ethnicity				
White	472711 (94.1%)	35572 (96.9%)	437139 (93.8%)	$1.2x10^{-114}$
Black	8034 (1.6%)	214 (0.6%)	7820 (1.7%)	$8.8x10^{-59}$
Other Asian	3389 (0.7%)	165 (0.4%)	3224 (0.7%)	$3.7x10^{-8}$
South Asian	8024 (1.6%)	313 (0.9%)	7711 (1.7%)	$1.7x10^{-32}$
Multiple, other or not provided	10363 (2.1%)	439 (1.2%)	9924 (2.1%)	$7.7x10^{-34}$
Coronary artery disease	17404 (3.5%)	1076 (2.9%)	16328 (3.5%)	$7.2x10^{-9}$
Diabetes	27848 (5.5%)	1808 (4.9%)	26040 (5.6%)	$8.5x10^{-8}$
Obese	122252 (24.3%)	6495 (17.7%)	115757 (24.9%)	$6.8x10^{-215}$
Hypertension	147343 (29.3%)	10289 (28.0%)	137054 (29.4%)	$1.8x10^{-8}$
Medications				
Anti-hypertensive therapy	104005 (20.7%)	4940 (13.5%)	99065 (21.3%)	$9.4x10^{-277}$
Lipid-lowering therapy	98894 (19.7%)	5552 (15.1%)	93342 (20.0%)	$6.3x10^{-115}$
Anthropometric data				
Weight, kg	78.1 (15.9)	76.8 (14.8)	78.2 (16.0)	$5.3x10^{-43}$
Waist-to-hip ratio	0.87 (0.09)	0.86 (0.09)	0.87 (0.09)	$1.7x10^{-142}$
Body-mass index, kg/m^2	27.4 (4.80)	26.6 (4.19)	27.5 (4.84)	$7.0x10^{-277}$
Body fat, %	31.5 (8.55)	30.0 (8.17)	31.6 (8.57)	$7.7x10^{-258}$
Estimated untreated systolic blood pressure, mmHg	141 (20.7)	137 (19.3)	141 (20.8)	$6.8x10^{-292}$
Alcohol consumption				
Weekly drinks, U.S. standard	4.84 (6.74)	5.48 (6.37)	4.79 (6.76)	$1.5x10^{-264}$
Weekly drinks, U.K. standard	8.47 (11.8)	9.58 (11.1)	8.38 (11.8)	$1.5x10^{-264}$
Excessive alcohol intake, U.S.	26408 (5.3%)	2015 (5.5%)	24393 (5.2%)	0.036
Excessive alcohol intake, U.K.	105842 (21.1%)	9066 (24.7%)	96776 (20.8%)	$1.5x10^{-70}$
Liver-associated biomarker concentrations				
Alanine aminotransferase, IU/L	23.5 (14.2)	23.0 (13.9)	23.6 (14.2)	$1.7x10^{-18}$
Aspartate aminotransferase, IU/L	26.2 (10.7)	25.8 (10.5)	26.3 (10.7)	$2.5x10^{-17}$
Gamma glutamyltransferase, IU/L	37.4 (42.1)	33.7 (33.9)	37.7 (42.7)	$2.2x10^{-119}$
Estimated untreated lipid concentrations				
Total cholesterol, mg/dL	228 (42.4)	227 (40.7)	228 (42.5)	$2.7x10^{-8}$
LDL cholesterol, mg/dL	146 (33.3)	144 (32.0)	146 (33.4)	$1.8x10^{-20}$
HDL cholesterol, mg/dL	56.0 (14.8)	57.0 (14.5)	55.9 (14.8)	$2.2x10^{-45}$
Triglycerides, mg/dL	135 [94-197]	126 [89-184]	136 [95-199]	$8.4x10^{-127}$
Glycemic biomarker concentrations				
Glycated hemoglobin, %	5.46 (0.620)	5.36 (0.475)	5.47 (0.629)	$2.3x10^{-260}$
Random glucose, mg/dL	92.3 (22.4)	89.9 (17.5)	92.5 (22.7)	$1.9x10^{-93}$

Values correspond to number (%), mean (standard deviation), or median [interquartile range]. P-values correspond to chi-squared test or Wilcoxon rank sum for categorical and continuous variables, respectively, for imaged compared to not imaged. Obesity was defined as body-mass index ≥ 30 kg/m^2 [1]; excessive alcohol intake, U.S. was defined as alcohol intake exceeding American Association for the Study of Liver Disease guidelines for NAFLD definition [2]; excessive alcohol intake, U.K. was defined as alcohol intake exceeding the UK Chief Medical Officers recommendations [3]. Diseases were defined as prevalent at time of initial assessment. Estimated untreated lipid measures and blood pressure were according to previously described adjustments [4,5]. NA, Not applicable.

Table S2. Definitions of disease in each cohort, Related to STAR methods.

Cohort	Disease	Definition
UK Biobank	NAFLD	Hospitalization due to nonalcoholic fatty liver (ICD10 K76.0)
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), esophageal 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	Hepatitis B/C	Hospitalization due to hepatitis B or C (ICD10 B18.0-.2) or self-reported hepatitis B or C during verbal interview with trained nurse (df-20002)
UK Biobank	Coronary Artery Disease	Self-report of heart attack diagnosed by doctor (df-6150), self-reported heart attack, coronary angioplasty, coronary artery bypass grafts, 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-.2, 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-K46.5, K46.8-.9), endarterectomy of coronary artery (K47.1), 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), adjudicated report of myocardial infarction from self-report, hospitalization, or death (df-42001)
UK Biobank	Diabetes	Self-reported diabetes during verbal interview with trained nurse (df-20002), 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, N08.3, O24.0-.3; ICD9 2500, 25000, 25001, 25009, 25010, 25011, 25019, 2502, 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)

NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis.

Table S3. 95% credible sets after fine-mapping of lead variant at eight common GWAS loci associated with quantitative liver fat, Related to Table 2.

Variant	Chr.	Position (hg19)	Nearest Gene	Variant Consequence	Effect Allele	Other Allele	Effect Allele Freq.	Effect on Liver Fat (Beta, SD)	P-value	PIP	Lead Variant at Locus?
rs2642438	1	220970028	*MTARC1*	Missense (p.T165A)	G	A	0.70	0.052	1.7E-09	0.91	Yes
rs867772	1	220972343	*MTARC1*	Intronic	G	A	0.69	0.047	3.3E-08	0.05	No
rs1229984	4	100239319	*ADH1B*	Missense (p.H48R)	C	T	0.98	0.158	7.0E-10	0.9997	Yes
rs112875651	8	126506694	*TRIB1*	Intergenic	G	A	0.61	0.050	3.8E-10	0.77	Yes
rs28601761	8	126500031	*TRIB1*	Intergenic	C	G	0.58	0.047	2.0E-09	0.16	No
rs2001844	8	126478745	*TRIB1*	Intergenic	A	G	0.52	0.037	1.5E-06	0.03	No
rs2792751	10	113940329	*GPAM*	Missense (p.V43I)	T	C	0.27	0.053	2.1E-09	0.26	No
rs2254537	10	113917085	*GPAM*	Synonymous	T	A	0.27	0.053	2.6E-09	0.22	No
rs10787429	10	113949664	*GPAM*	Intronic; Non-coding	T	C	0.27	0.054	1.8E-09	0.20	No
rs77987196	10	113933006	*GPAM*	Intronic	ATT	A	0.27	0.052	5.1E-09	0.12	No
rs2297991	10	113913222	*GPAM*	3' UTR	T	C	0.28	0.050	1.7E-08	0.03	No
rs2250802	10	113921354	*GPAM*	Intronic	G	A	0.27	0.054	1.4E-09	0.02	Yes
rs2803619	10	113934384	*GPAM*	Intronic	G	C	0.27	0.053	2.2E-09	0.01	No
rs2803611	10	113922728	*GPAM*	Intronic	A	G	0.27	0.053	2.3E-09	0.01	No
rs2803608	10	113916302	*GPAM*	Intronic	T	C	0.27	0.053	2.5E-09	0.01	No
rs2254532	10	113916835	*GPAM*	Intronic	A	C	0.27	0.053	2.6E-09	0.01	No
rs2792736	10	113921159	*GPAM*	Intronic	T	A	0.27	0.053	2.7E-09	0.009	No
rs4918722	10	113947040	*GPAM*	Upstream	C	T	0.27	0.053	2.6E-09	0.009	No
rs2792759	10	113936855	*GPAM*	Intronic	C	T	0.27	0.053	3.2E-09	0.008	No
rs2803621	10	113939584	*GPAM*	Intronic	G	A	0.27	0.053	3.6E-09	0.007	No
rs1129555	10	113910721	*GPAM*	3' UTR	A	G	0.27	0.053	3.5E-09	0.007	No
rs2792735	10	113921825	*GPAM*	Intronic	G	A	0.28	0.052	3.2E-09	0.007	No
rs2803609	10	113919124	*GPAM*	Intronic	A	G	0.27	0.052	4.0E-09	0.007	No
rs429358	19	45411941	*APOE*	Missense (p.R130C)	T	C	0.85	0.121	1.5E-29	1.00	Yes
rs56252442	19	18229208	*MAST3*	Intronic	T	G	0.25	0.049	2.7E-08	0.67	Yes
rs60146811	19	18221213	*MAST3*	Intronic	A	G	0.27	0.044	4.4E-07	0.07	No
rs67234314	19	18218610	*MAST3*	Intronic	T	A	0.27	0.044	3.6E-07	0.05	No
rs885683	19	18244690	*MAST3*	Intronic	A	G	0.25	0.047	9.6E-08	0.03	No
rs746721254	19	18220370	*MAST3*	Intronic	A	AGAGT	0.27	0.045	1.7E-07	0.03	No
rs72999466	19	18248499	*MAST3*	Intronic	T	C	0.26	0.046	2.8E-07	0.03	No
rs56345159	19	18235873	*MAST3*	Intronic	T	C	0.25	0.048	1.1E-07	0.03	No
rs874628	19	18304700	*MPV17L2*	Missense (p.M72V)	G	A	0.28	0.040	2.8E-06	0.03	No
rs11554159	19	18285944	*IFI30*	Missense (p.R76Q)	A	G	0.27	0.042	3.2E-06	0.02	No
rs58542926	19	19379549	*TM6SF2*	Missense (p.E167K)	T	C	0.07	0.289	2.8E-85	0.99	Yes
rs738409	22	44324727	*PNPLA3*	Missense (p.I148M)	G	C	0.21	0.195	5.6E-95	0.94	No
rs738408	22	44324730	*PNPLA3*	Synonymous	T	C	0.21	0.195	5.3E-95	0.06	Yes

Chr., chromosome; Freq., frequency; PIP, posterior inclusion probability; GWAS, genome-wide association study; Effect on liver fat (Beta, SD), effect of variant on inverse normal transformed liver fat in standard deviation (SD) units.

Table S4. Effects of eight common variants associated with quantitative liver fat in the n=4,040 individuals with previously-quantified liver fat compared to the n=32,974 individuals with liver fat quantification enabled by machine learning, Related to Table 2.

						Variant effects in 4,040 individuals with previously-quantified liver fat			Variant effects in machine-learning expansion to 32,974 individuals with liver fat		
Lead Variant	Chr.	Position (hg19)	Nearest Gene	Effect Allele	Other Allele	Effect on Liver Fat (Beta, SD)	SE	P-value	Effect on Liver Fat (Beta, SD)	SE	P-value
Newly-identified variants											
rs2642438	1	220970028	*MTARC1*	G	A	0.03	0.024	0.160	0.05	0.008	1.7E-09
rs1229984	4	100239319	*ADH1B*	C	T	0.16	0.073	0.026	0.16	0.025	7.0E-10
rs112875651	8	126506694	*TRIB1*	G	A	0.08	0.023	3.6E-04	0.05	0.008	3.8E-10
rs2250802	10	113921354	*GPAM*	G	A	0.07	0.025	5.9E-03	0.05	0.009	1.4E-09
rs56252442	19	18229208	*MAST3*	T	G	0.06	0.025	0.012	0.05	0.009	2.7E-08
Previously-identified variants											
rs58542926	19	19379549	*TM6SF2*	T	C	0.35	0.042	9.3E-17	0.29	0.015	2.8E-85
rs429358	19	45411941	*APOE*	T	C	0.14	0.031	7.1E-06	0.12	0.011	1.5E-29
rs738409	22	44324727	*PNPLA3*	G	C	0.25	0.027	2.3E-20	0.19	0.009	5.6E-95

Chr., chromosome; SE, standard error; Effect on liver fat (Beta, SD), effect of variant on inverse normal transformed liver fat in standard deviation (SD) units.

Table S5. Effect of variants previously associated with liver fat, related liver disease, or cirrhosis in common variant genome-wide association study, on liver fat in 32,974 individuals in UK Biobank, Related to Table 2.

Variant	Chr.	Position (hg19)	Effect Allele	Other Allele	Effect Allele Freq.	Effect on liver fat (Beta, SD)	P-value	Locus	Previous reference(s)
rs12077210	1	65894160	C	T	0.96	-0.03	0.11	*LEPR*	[6]
rs12137855	1	219448378	C	T	0.79	0.01	0.46	*LYPLAL1*	[7,8]
rs1260326	2	27730940	T	C	0.39	0.04	4.10E-07	*GCKR*	[8–10]
rs780094	2	27741237	T	C	0.38	0.04	3.50E-06	*GCKR*	[7,8]
rs6834314	4	88213808	A	G	0.72	0.01	0.32	*HSD17B13*	[11–13]
rs72613567	4	88231392	T	TA	0.73	0.01	0.40	*HSD17B13*	[11–13]
rs62305723	4	88231429	G	A	0.93	0.01	0.59	*HSD17B13*	[12]
rs11134977	5	175904141	T	C	0.55	-0.01	0.09	*FAF2*	[14]
rs4240624	8	9184231	G	A	0.09	-0.02	0.10	*PPP1R3B*	[7,8]
rs10883451	10	101924418	T	C	0.50	0.02	0.02	*ERLIN1*	[15]
rs11597086	10	101953705	A	C	0.55	0.02	0.01	*CHUK*	[15]
rs62021874	15	55874043	C	T	0.93	-0.02	0.29	*PYGO1*	[6]
rs2228603	19	19329924	C	T	0.92	-0.22	1.80E-51	*NCAN*	[7,8]
rs641738	19	54676763	C	T	0.56	-0.03	8.80E-06	*MBOAT7-TMC4*	[16–18]

Chr., chromosome; Freq., frequency; Effect on liver fat (Beta, SD), effect of variant on inverse normal transformed liver fat in standard deviation (SD) units.

Table S6. Effects of eight common liver fat variants on quantitative liver fat after adjusting for alcohol consumption, Related to Table 2.

					Unadjusted for alcohol consumption (original CVAS, n=32974)			Former alcohol consumers excluded, adjusted for number of weekly drinks (n=32062)			Former and excessive alcohol consumers (US guidelines) excluded (n=30216)			Former and excessive alcohol consumers (UK guidelines) excluded (n=23931)		
Lead Variant	Chr.	Position (hg19)	Nearest Gene	Effect Allele	Effect on liver fat (Beta, SD)	SE	P-value	Effect on liver fat (Beta, SD)	SE	P-value	Effect on liver fat (Beta, SD)	SE	P-value	Effect on liver fat (Beta, SD)	SE	P-value
Newly-Identified variants																
rs2642438	1	220970028	*MTARC1*	G	0.052	0.008	1.70E-09	0.050	0.009	8.70E-09	0.046	0.009	4.20E-07	0.052	0.010	3.20E-07
rs1229984	4	100239319	*ADH1B*	C	0.158	0.025	7.00E-10	0.153	0.026	3.40E-09	0.147	0.026	1.40E-08	0.116	0.029	5.30E-05
rs112875651	8	126506694	*TRIB1*	G	0.050	0.008	3.80E-10	0.053	0.008	3.70E-11	0.052	0.008	3.50E-10	0.057	0.009	8.30E-10
rs2250802	10	113921354	*GPAM*	G	0.054	0.009	1.40E-09	0.056	0.009	3.70E-10	0.056	0.009	2.10E-09	0.050	0.010	1.10E-06
rs56252442	19	18229208	*MAST3*	T	0.049	0.009	2.70E-08	0.050	0.009	1.70E-08	0.051	0.009	3.20E-08	0.039	0.011	1.60E-04
Previously-identified variants																
rs58542926	19	19379549	*TM6SF2*	T	0.289	0.015	2.80E-85	0.288	0.015	6.10E-83	0.283	0.015	1.60E-75	0.285	0.017	3.60E-61
rs429358	19	45411941	*APOE*	T	0.121	0.011	1.50E-29	0.120	0.011	3.90E-28	0.114	0.011	4.80E-24	0.102	0.013	3.90E-16
rs738409	22	44324727	*PNPLA3*	G	0.195	0.009	5.60E-95	0.194	0.010	1.10E-92	0.191	0.010	3.20E-84	0.175	0.011	6.10E-56

Chr., chromosome; SE, standard error; Effect on liver fat (Beta, SD), effect of variant on inverse normal transformed liver fat in standard deviation (SD) units. Excessive alcohol intake, U.S. guidelines was defined as alcohol intake exceeding American Association for the Study of Liver Disease guidelines for NAFLD definition [2]; excessive alcohol intake, U.K. guidelines was defined as alcohol intake exceeding the UK Chief Medical Officers recommendations [3].

Table S7. Replication of eight common UK Biobank variants associated with quantitative liver fat in two additional liver fat cohorts, Related to Table 2.

					UK Biobank Discovery CVAS			Framingham Heart Study Replication			MESA Replication			Combined Framingham + MESA Replication			
Lead Variant	Chr.	Position (hg19)	Nearest Gene	Effect Allele	Effect on MRI liver fat (Beta, SD)	SE	P-value	Effect on CT liver fat (Beta, SD)	SE	P-value	Effect on CT liver fat (Beta, SD)	SE	P-value	Effect on CT liver fat (Beta, SD)	SE	P-value	P_{het}-value
Newly-identified variants																	
rs2642438	1	220970028	*MTARC1*	G	0.052	0.008	2E-09	0.088	0.028	2E-03	0.062	0.026	0.017	0.074	0.019	1E-04	0.50
rs1229984	4	100239319	*ADH1B*	C	0.158	0.025	7E-10	0.101	0.057	0.081	0.034	0.040	0.394	0.056	0.033	0.088	0.34
rs112875651	8	126506694	*TRIB1*	G	0.050	0.008	4E-10	0.070	0.026	8E-03	0.044	0.023	0.059	0.055	0.017	1E-03	0.45
rs2250802	10	113921354	*GPAM*	G	0.054	0.009	1E-09	0.058	0.028	0.037	0.031	0.025	0.218	0.043	0.019	0.021	0.47
rs56252442	19	18229208	*MAST3*	T	0.049	0.009	3E-08	0.065	0.032	0.041	0.017	0.026	0.501	0.036	0.020	0.070	0.24
Previously-identified variants																	
rs58542926	19	19379549	*TM6SF2*	T	0.289	0.015	3E-85	0.326	0.056	6E-09	0.229	0.045	3E-07	0.267	0.035	2E-14	0.17
rs429358	19	45411941	*APOE*	T	0.121	0.011	2E-29	-0.119	0.075	0.110	0.056	0.029	0.054	-0.018	0.087	0.837	0.03
rs738409	22	44324727	*PNPLA3*	G	0.195	0.009	6E-95	0.278	0.030	3E-20	0.296	0.024	4E-34	0.289	0.019	3E-53	0.65

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; random-effects models were used when nominal heterogeneity was noted (P_{het} <0.05). Chr., chromosome; SE, standard error; P_{het}, heterogeneity p value; Effect on liver fat (Beta, SD), effect of variant on inverse normal transformed liver fat in standard deviation (SD) units.

Table S8. Effects of eight common variants associated with quantitative liver fat on blood biomarkers of liver injury, Related to Table 2.

Lead Variant	Chr.	Position (hg19)	Nearest Gene	Effect Allele	Other Allele	Effect Allele Freq.	Effect on liver fat (Beta, SD)	SE	P-value	Effect on ALT (U/L)	SE	P-value	Effect on AST (U/L)	SE	P-value
Newly-identified variants															
rs2642438	1	220970028	*MTARC1*	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	*ADH1B*	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	*TRIB1*	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	*GPAM*	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	*MAST3*	T	G	0.25	0.049	0.009	3E-08	0.29	0.038	1E-14	0.03	0.029	0.342
Previously-identified variants															
rs58542926	19	19379549	*TM6SF2*	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	*APOE*	T	C	0.85	0.121	0.011	2E-29	0.66	0.045	7E-49	0.07	0.035	0.045
rs738409	22	44324727	PNPLA3	G	C	0.21	0.195	0.009	6E-95	1.68	0.040	<1E-300	1.06	0.031	1E-261

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 MRI imaging. Chr., chromosome; Freq., frequency; SE, standard error.

Table S9. Effects of eight common variants associated with quantitative liver fat on clinical diagnoses of liver disease, Related to Table 2.

							UK Biobank			Mass General Brigham Biobank			Combined UK Biobank + Mass General Brigham Biobank			
Lead Variant	Chr.	Position (hg19)	Nearest Gene	Effect Allele	Other Allele	Effect Allele Freq.	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	P_{het}-value
Newly-identified variants																
rs2642438	1	220970028	*MTARC1*	G	A	0.70	1.14	0.034	2E-04	1.06	0.027	0.03	1.09	0.021	4E-05	0.12
rs1229984	4	100239319	*ADH1B*	C	T	0.98	1.33	0.118	0.015	1.26	0.052	7E-06	1.28	0.048	3E-07	0.68
rs112875651	8	126506694	*TRIB1*	G	A	0.61	1.10	0.032	2E-03	1.07	0.025	6E-03	1.08	0.020	4E-05	0.46
rs2250802	10	113921354	*GPAM*	G	A	0.27	1.08	0.033	0.016	1.07	0.026	8E-03	1.08	0.021	3E-04	0.82
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	0.42
Previously-identified variants																
rs58542926	19	19379549	*TM6SF2*	T	C	0.07	1.41	0.050	6E-12	1.38	0.044	8E-13	1.39	0.033	3E-23	0.71
rs429358	19	45411941	*APOE*	T	C	0.85	1.24	0.045	1E-06	1.18	0.036	5E-06	1.20	0.028	5E-11	0.37
rs738409	22	44324727	*PNPLA3*	G	C	0.21	1.52	0.033	9E-37	1.38	0.027	7E-32	1.44	0.050	2E-13	0.02

Effects on NAFLD/NASH in a subset of the UK Biobank who did not undergo abdominal MRI 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; random-effects models were used when nominal heterogeneity was noted (P_{het} <0.05). NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; Chr., chromosome; Freq., frequency; OR, odds ratio; SE, standard error; P_{het}, heterogeneity p-value.

Table S10. Comparison of C statistics for discrimination of liver diseases with addition of an eight-variant polygenic score, Related to Figure 4.

Disease	N Events / N Individuals	Model	C-statistic (95%CI)	P-value
Nonalcoholic fatty liver disease	1635 / 361852	Baseline	0.552 (0.538-0.566)	
		Baseline + Polygenic score	0.598 (0.584-0.612)	4.41E-34
Nonalcoholic steatohepatitis	208 / 361852	Baseline	0.603 (0.564-0.642)	
		Baseline + Polygenic score	0.684 (0.648-0.721)	3.44E-16
Cirrhosis	977 / 361852	Baseline	0.674 (0.657-0.691)	
		Baseline + Polygenic score	0.694 (0.677-0.711)	1.75E-30
Hepatocellular Carcinoma	171 / 361852	Baseline	0.750 (0.713-0.786)	
		Baseline + Polygenic score	0.770 (0.731-0.808)	2.55E-15

Baseline model: age + age squared + sex + genotyping array + 10 principal components of ancestry.
P-value, comparison of baseline and baseline + polygenic score models using likelihood ratio test.

Table S11. Rare inactivating variants in *APOB* are associated with decreased circulating lipids but increased markers of liver injury, Related to Figure 5.

	APOB inactivating			*MTTP* inactivating		
	Noncarriers (N=168470)	variant carriers (N=130)	Adjusted P-value	Noncarriers (N=168510)	variant carriers (N=90)	Adjusted P-value
BMI, mean (SD)	27.4 (4.72)	27.4 (4.21)	0.91	27.4 (4.72)	27.3 (4.50)	0.83
Obesity, N (%)	40081 (23.8%)	35 (26.9%)	0.36	40098 (23.8%)	18 (20.0%)	0.42
Apolipoprotein B, mean (SD), g/L	1.03 (0.24)	0.643 (0.29)	5.48E-48	1.03 (0.24)	1.03 (0.25)	0.61
LDL cholesterol, mean (SD), mg/dL	146 (33.0)	81.2 (38.1)	3.74E-113	146 (33.0)	149 (34.3)	0.60
Triglycerides, mean (SD), mg/dL	160 (95.3)	87.2 (68.4)	7.77E-18	160 (95.3)	151 (59.2)	0.37
Alanine aminotransferase, mean (SD), U/L	23.4 (13.8)	31.6 (24.0)	9.62E-13	23.4 (13.8)	24.7 (13.9)	0.26
Aspartate aminotransferase, mean (SD), U/L	26.2 (10.2)	29.9 (14.6)	1.73E-05	26.2 (10.2)	26.9 (9.33)	0.45
Coronary artery disease, N (%)	10071 (6.0%)	2 (1.5%)	0.04	10067 (6.0%)	6 (6.7%)	0.93
Type 2 diabetes, N (%)	13340 (7.9%)	14 (10.8%)	0.24	13341 (7.9%)	13 (14.4%)	0.04

Apolipoprotein B values were missing in 55/130 APOB inactivating variant carriers (42%) and 8796/168470 controls (5%), all other measurements had <3% difference in missingness between groups. Obesity was defined as body-mass index (BMI) ≥ 30 kg/m^2 [1]. LDL cholesterol and triglycerides were adjusted for lipid-lowering medication to estimate untreated values as previously described [5].

Table S12. Characteristics of UK Biobank participants with hepatic steatosis stratified by presence of an inactivating variant in *APOB* or *MTTP*, Related to Figure 5.

	Noncarriers (N=3260)	*APOB* inactivating variant carriers (N=13)	Adjusted P-value	Noncarriers (N=3259)	*MTTP* inactivating variant carriers (N=14)	Adjusted P-value	Noncarriers (N=3246)	Combined *APOB* and *MTTP* inactivating variant carriers (N=27)	Adjusted P-value
BMI, mean (SD)	29.6 (4.3)	27.6 (3.5)	0.10	29.6 (4.3)	28.0 (3.3)	0.21	29.6 (4.3)	27.8 (3.3)	0.04
Obesity, N (%)	1294 (39.7%)	3 (23.1%)	0.23	1294 (39.7%)	3 (21.4%)	0.21	1291 (39.8%)	6 (22.2%)	0.08
Apolipoprotein B, mean (SD), g/L	1.08 (0.24)	0.60 (0.24)	9.9E-07	1.08 (0.24)	1.15 (0.31)	0.24	1.08 (0.24)	0.99 (0.39)	0.09
LDL cholesterol, mean (SD), mg/dL	151 (31.8)	73.2 (34.4)	1.4E-18	150 (32.1)	168 (48.4)	0.05	150 (31.7)	122 (63.6)	3.2E-06
Triglycerides, mean (SD), mg/dL	204 (110)	60.6 (29.6)	2.5E-06	203 (110)	174 (57.3)	0.29	204 (110)	120 (73.5)	5.4E-05
Alanine aminotransferase, mean (SD), U/L	31.1 (18.1)	31.0 (14.8)	0.91	31.1 (18.1)	25.7 (10.6)	0.25	31.1 (18.1)	28.2 (12.8)	0.46
Aspartate aminotransferase, mean (SD), U/L	28.6 (10.8)	26.9 (5.08)	0.60	28.6 (10.8)	28.1 (6.41)	0.80	28.6 (10.9)	27.5 (5.73)	0.58
Coronary artery disease, N (%)	165 (5.1%)	0 (0%)	0.97	165 (5.1%)	0 (0%)	0.97	165 (5.1%)	0 (0%)	0.98
Type 2 diabetes, N (%)	484 (14.8%)	3 (23.1%)	0.40	485 (14.9%)	2 (14.3%)	0.82	482 (14.8%)	5 (18.5%)	0.69
Weekly drinks, U.S. standard	5.90 (7.54)	5.85 (6.23)	0.97	5.90 (7.54)	4.96 (5.59)	0.57	5.90 (7.55)	5.39 (5.81)	0.71
Excessive alcohol intake, U.S., N (%)	232 (7.1%)	0 (0%)	0.97	232 (7.1%)	0 (0%)	0.97	232 (7.1%)	0 (0%)	0.96
Excessive alcohol intake, U.K., N (%)	855 (26.2%)	4 (30.8%)	0.69	855 (26.2%)	4 (28.6%)	0.89	851 (26.2%)	8 (29.6%)	0.71

Apolipoprotein B values were missing in 7/13 APOB inactivating variant carriers (53%) and 199/3260 controls (6%), all other measurements had <3% difference in missingness between groups. Obesity was defined as body-mass index (BMI) ≥ 30 kg/m^2 [1]; excessive alcohol intake, U.S. was defined as alcohol intake exceeding American Association for the Study of Liver Disease guidelines for NAFLD definition [2]; excessive alcohol intake, U.K. was defined as alcohol intake exceeding the UK Chief Medical Officers recommendations [3]. LDL cholesterol and triglycerides were adjusted for lipid-lowering medication to estimate untreated values as previously described [5].

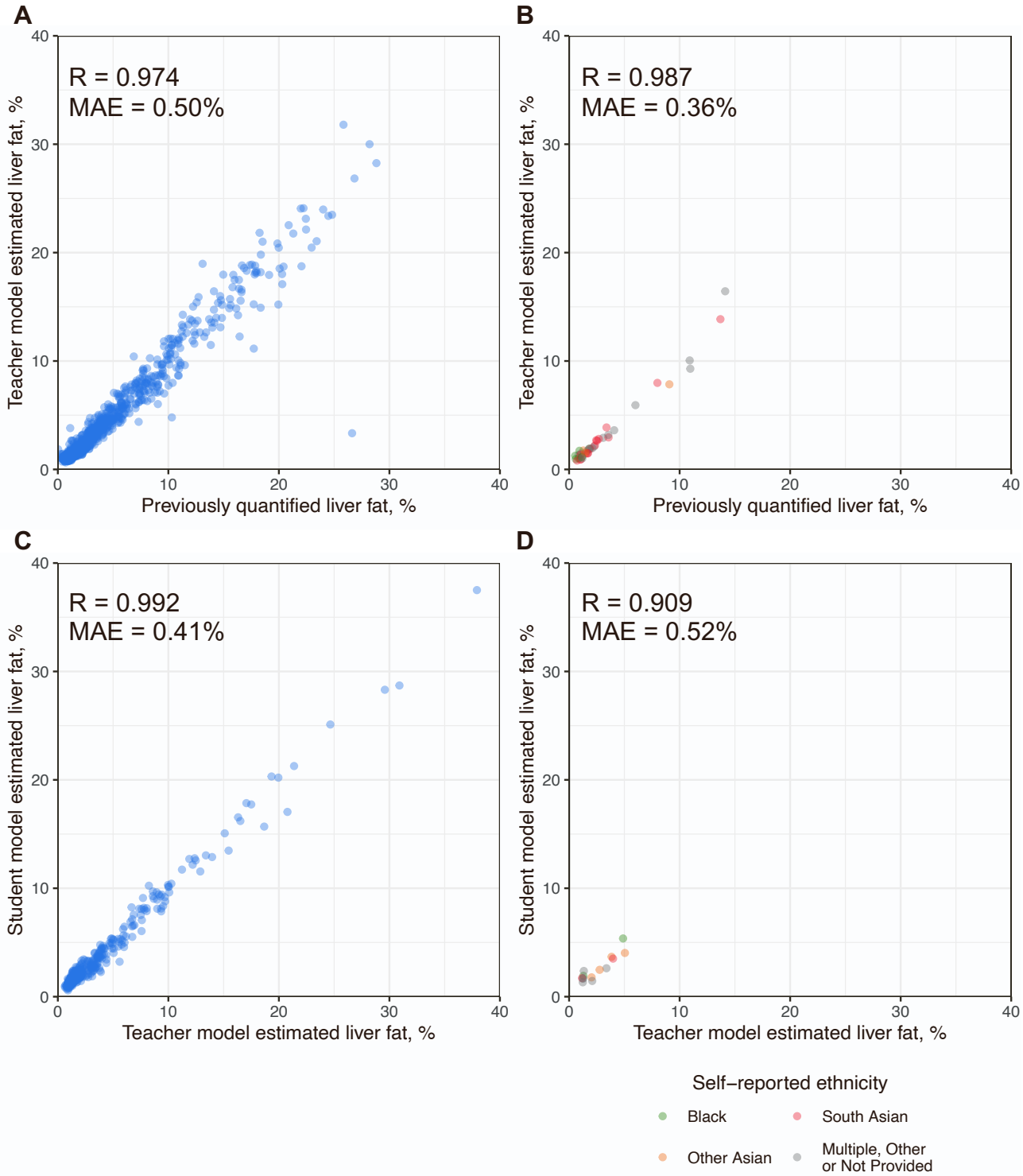

Figure S1. Comparison of previously-quantified liver fat with teacher-model inferences, and teacher-model inferences with student-model inferences, Related to STAR methods. A) In a held-out set of 1,214 participants with previously-quantified liver fat from gradient-echo imaging who were not used for model creation, the Pearson correlation between previously-quantified liver fat and liver fat inferred from the machine learning teacher model was 0.974 (95% CI 0.971-0.977; $P<2.4x10^{-784}$), and the mean absolute error was 0.50% (95%CI, 0.45-0.55%). **B**) The subset of testing individuals in A) with self-reported non-European ethnicity. **C**) In a separate held-out set of 383 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; $P = 3.1 \times 10^{-351}$) and the mean absolute error was 0.41% (95%CI, 0.37-0.46%). **D**) The subset of testing individuals in C) with self-reported non-European ethnicity.

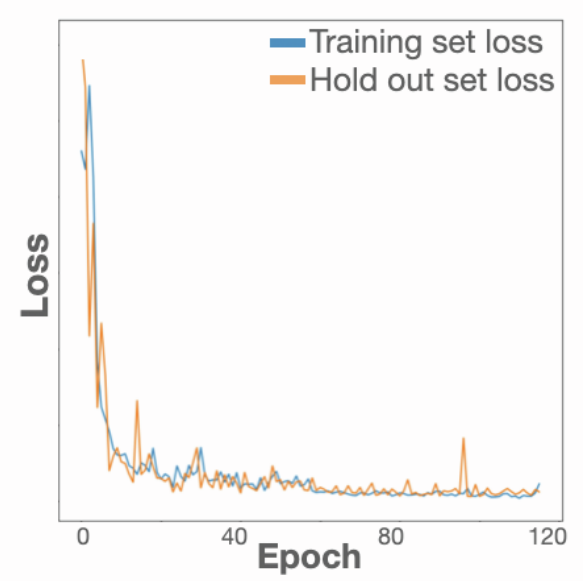

Figure S2. Assessment of machine learning model overfitting via model loss, Related to STAR methods. To determine whether our machine learning model quantifying liver fat was prone to overfitting, we generated learning curves that show model's loss parameters according to epoch on the training set of images from 2,915 individuals (blue line) and on a held-out set of data set of images from 635 individuals (orange line) distinct from the final test set. Each epoch was defined as a full pass over the training set MRIs. Increased loss in the training set is suggestive of model overfitting. By the end of training, we note consistent loss in the training and validation dataset suggestive no evidence of overfitting.

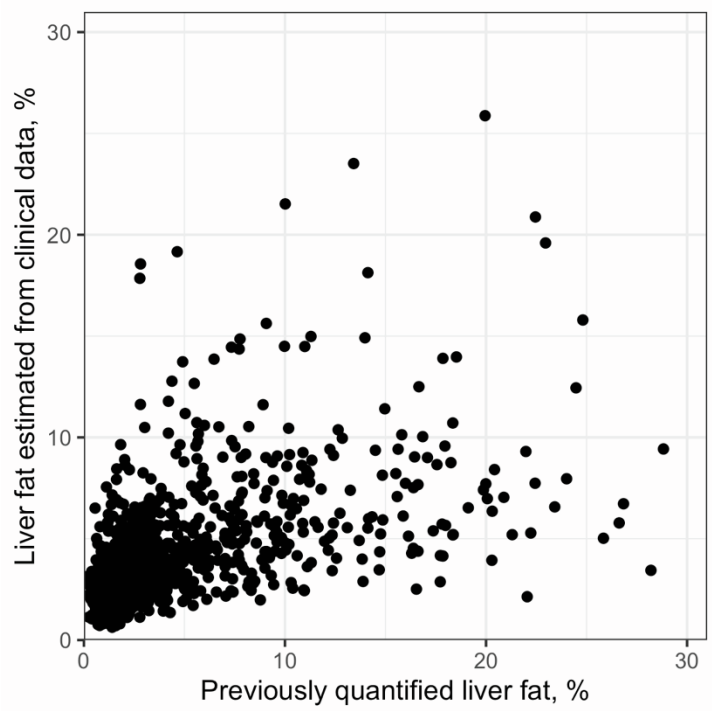

Figure S3. Prediction of liver fat using a variable dispersion beta regression model of clinical and anthropometric measurements, Related to STAR methods. In a held-out set of 1,214 participants with previously-estimated liver fat, the Pearson correlation between the previously-quantified liver fat and liver fat estimated from a beta regression model using clinical and anthropometric measurements was 0.578 (95% CI 0.539-0.614; p-value=3.8×10^{-109}). Measurements that were at least nominally (p-value < 0.05) associated with liver fat in univariable analysis and therefore included in the beta regression model were: body-mass index, waist circumference, hip circumference, total body fat mass, total body fat percent, age at baseline, sex, height, weight, trunk fat mass, trunk fat percent, waist-to-hip ratio, 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 [4,5].

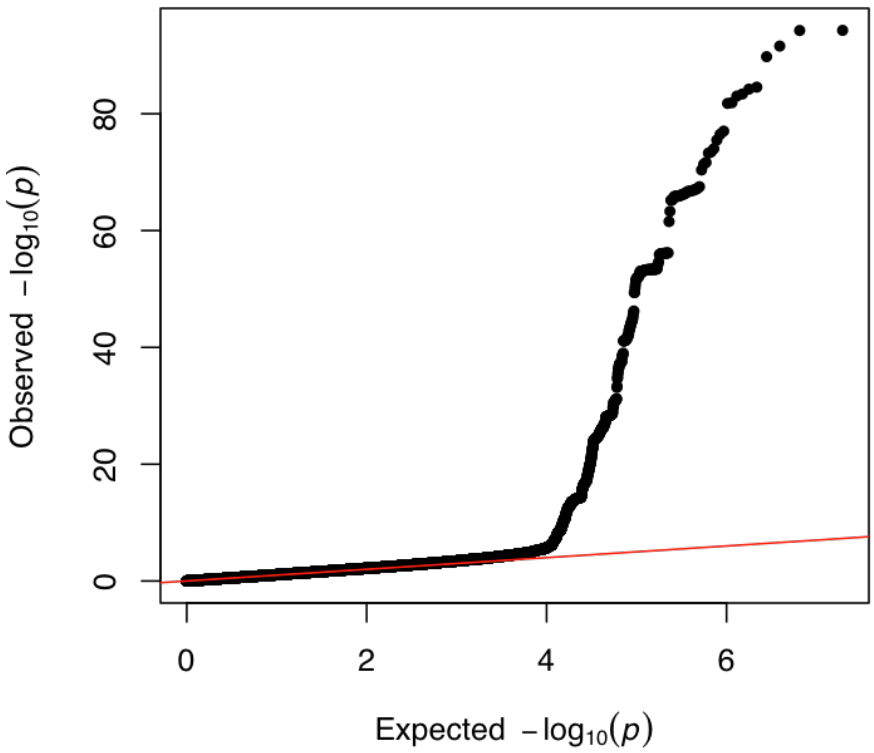

Figure S4. Quantile-quantile plot for common variant genome-wide association study analysis of liver fat in 32,974 UK Biobank participants, Related to Figure 3. Expected p-values from a uniform distribution for each variant in the genome-wide association study (GWAS) are shown on the x-axis; the corresponding observed p-values for each variant are shown on the y-axis.

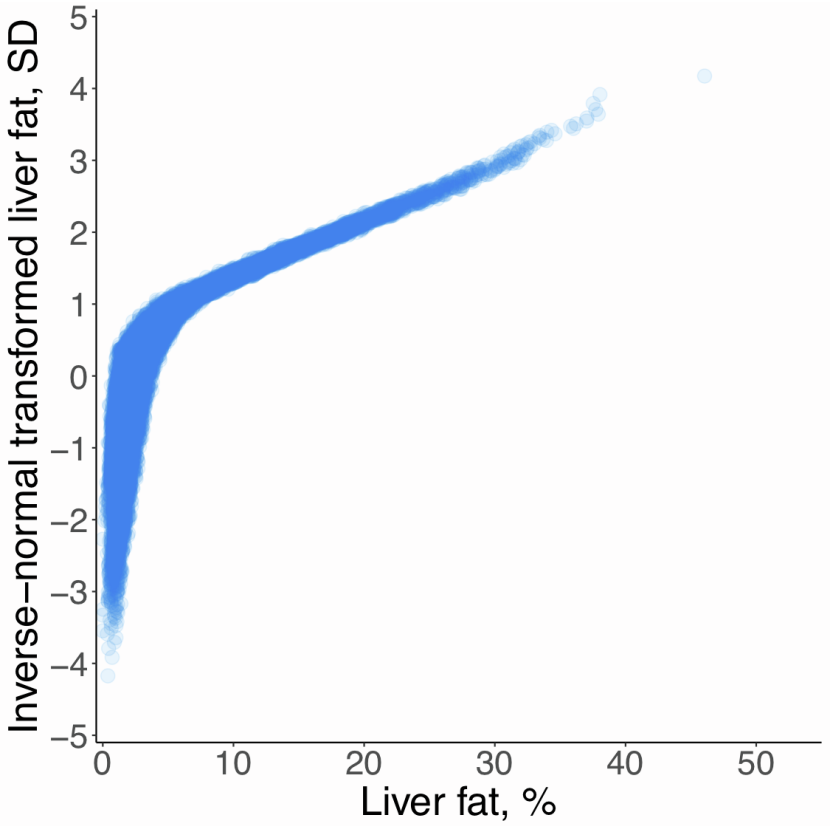

Figure S5. Relationship between percent liver fat and inverse normal transformation of percent liver fat among 32,974 UK Biobank participants, Related to Figures 3-5. Percent liver fat is shown on the x-axis; the inverse-normal transformation of percent liver fat used in common genome-wide and rare variant association studies (GWAS and RVAS) is shown on the y-axis.

Supplemental References

1. NHLBI Expert Panel (1998). Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults--The Evidence Report. Obes. Res. *6 Suppl 2*, 51S-209S.

2. Chalasani, N., Younossi, Z., Lavine, J.E., Charlton, M., Cusi, K., Rinella, M., Harrison, S.A., Brunt, E.M., and Sanyal, A.J. (2018). The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatol. Baltim. Md *67*, 328–357.

3. Department of Health (2016). UK Chief Medical Officers' Low Risk Drinking Guidelines.

4. Ehret, G.B., Ferreira, T., Chasman, D.I., Jackson, A.U., Schmidt, E.M., Johnson, T., Thorleifsson, G., Luan, J., Donnelly, L.A., Kanoni, S., *et al.* (2016). The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. Nat. Genet. *48*, 1171–1184.

5. Patel, A.P., Wang, M., Fahed, A.C., Mason-Suares, H., Brockman, D., Pelletier, R., Amr, S., Machini, K., Hawley, M., Witkowski, L., *et al.* (2020). Association of Rare Pathogenic DNA Variants for Familial Hypercholesterolemia, Hereditary Breast and Ovarian Cancer Syndrome, and Lynch Syndrome With Disease Risk in Adults According to Family History. JAMA Netw. Open *3*, e203959.

6. Anstee, Q.M., Darlay, R., Cockell, S., Meroni, M., Govaere, O., Tiniakos, D., Burt, A.D., Bedossa, P., Palmer, J., Liu, Y.-L., *et al.* (2020). Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort. J. Hepatol. *73*, 505–515.

7. Speliotes, E.K., Yerges-Armstrong, L.M., Wu, J., Hernaez, R., Kim, L.J., Palmer, C.D., Gudnason, V., Eiriksdottir, G., Garcia, M.E., Launer, L.J., *et al.* (2011). Genome-Wide Association Analysis Identifies Variants Associated with Nonalcoholic Fatty Liver Disease That Have Distinct Effects on Metabolic Traits. PLoS Genet. *7*, e1001324-14.

8. Palmer, N.D., Musani, S.K., Yerges-Armstrong, L.M., Feitosa, M.F., Bielak, L.F., Hernaez, R., Kahali, B., Carr, J.J., Harris, T.B., Jhun, M.A., *et al.* (2013). Characterization of European ancestry nonalcoholic fatty liver disease-associated variants in individuals of African and Hispanic descent. Hepatol. Baltim. Md *58*, 966–975.

9. Parisinos, C.A., Wilman, H.R., Thomas, E.L., Kelly, M., Nicholls, R.C., McGonigle, J., Neubauer, S., Hingorani, A.D., Patel, R.S., Hemingway, H., *et al.* (2020). Genome-wide and Mendelian randomisation studies of liver MRI yield insights into the pathogenesis of steatohepatitis. J. Hepatol., 1–41.

10. Palmer, N.D., Kahali, B., Kuppa, A., Chen, Y., Du, X., Feitosa, M.F., Bielak, L.F., O'Connell, J.R., Musani, S.K., Guo, X., *et al.* (2021). 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. *30*, 1443–1456.

11. Abul-Husn, N.S., Cheng, X., Li, A.H., Xin, Y., Schurmann, C., Stevis, P., Liu, Y., Kozlitina, J., Stender, S., Wood, G.C., *et al.* (2018). A Protein-Truncating HSD17B13Variant and Protection from Chronic Liver Disease. N. Engl. J. Med. *378*, 1096–1106.

12. Ma, Y., Belyaeva, O.V., Brown, P.M., Fujita, K., Valles, K., Karki, S., de Boer, Y.S., Koh, C., Chen, Y., Du, X., *et al.* (2019). 17-Beta Hydroxysteroid Dehydrogenase 13 Is a Hepatic Retinol Dehydrogenase Associated With Histological Features of Nonalcoholic Fatty Liver Disease. Hepatol. Baltim. Md *69*, 1504–1519.

13. Gellert-Kristensen, H., Nordestgaard, B.G., Tybjærg-Hansen, A., and Stender, S. (2020). High Risk of Fatty Liver Disease Amplifies the Alanine Transaminase–Lowering Effect of a HSD17B13 Variant. Hepatology *71*, 56–66.

14. Schwantes-An, T.-H., Darlay, R., Mathurin, P., Masson, S., Liangpunsakul, S., Mueller, S., Aithal, G.P., Eyer, F., Gleeson, D., Thompson, A., *et al.* (2020). Genome-wide association study and meta-analysis on alcohol-related liver cirrhosis identifies novel genetic risk factors. Hepatol. Baltim. Md.

15. Feitosa, M.F., Wojczynski, M.K., North, K.E., Zhang, Q., Province, M.A., Carr, J.J., and Borecki, I.B. (2013). The ERLIN1-CHUK-CWF19L1 gene cluster influences liver fat deposition and hepatic inflammation in the NHLBI Family Heart Study. Atherosclerosis *228*, 175–180.

16. Buch, S., Stickel, F., Trépo, E., Way, M., Herrmann, A., Nischalke, H.D., Brosch, M., Rosendahl, J., Berg, T., Ridinger, M., *et al.* (2015). A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. Nat. Genet. *47*, 1443–1448.

17. Mancina, R.M., Dongiovanni, P., Petta, S., Pingitore, P., Meroni, M., Rametta, R., Borén, J., Montalcini, T., Pujia, A., Wiklund, O., *et al.* (2016). The MBOAT7-TMC4 Variant rs641738 Increases Risk of Nonalcoholic Fatty Liver Disease in Individuals of European Descent. Gastroenterology *150*, 1219-1230.e6.

18. Teo, K., Abeysekera, K.W.M., Adams, L., Aigner, E., Anstee, Q.M., Banales, J.M., Banerjee, R., Basu, P., Berg, T., Bhatnagar, P., *et al.* (2021). rs641738C>T near MBOAT7 is associated with liver fat, ALT and fibrosis in NAFLD: A meta-analysis. J. Hepatol. *74*, 20–30.