UCLA
UCLA Previously Published Works

Title

Allele-specific variation at APOE increases nonalcoholic fatty liver disease and obesity but decreases risk of Alzheimer's disease and myocardial infarction

Permalink

https://escholarship.org/uc/item/83d5k879

Journal

Human Molecular Genetics, 30(15)

ISSN

0964-6906

Authors

Palmer, Nicholette D
Kahali, Bratati
Kuppa, Annapurna
et al.

Publication Date

2021-07-09

DOI

10.1093/hmg/ddab096

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Human Molecular Genetics, 2021, Vol. 30, No. 15 1443–1456

https://doi.org/10.1093/hmg/ddab096
Advance Access Publication Date: 15 April 2021
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Allele-specific variation at *APOE* increases nonalcoholic fatty liver disease and obesity but decreases risk of Alzheimer's disease and myocardial infarction

Nicholette D. Palmer[1,†,*], Bratati Kahali[2,3,†], Annapurna Kuppa[3,†], Yanhua Chen[3,†], Xiaomeng Du[3,†], Mary F. Feitosa[4,‡], Lawrence F. Bielak[5,‡], Jeffrey R. O'Connell[6,‡], Solomon K. Musani[7,‡], Xiuqing Guo[8,†], Albert V. Smith[9,‡], Kathleen A. Ryan[6], Gudny Eirksdottir[9], Matthew A. Allison[10], Donald W. Bowden[1], Matthew J. Budoff[11], J. Jeffrey Carr[12], Yii-Der I. Chen[8], Kent D. Taylor[8], Adolfo Correa[7], Breland F. Crudup[7], Brian Halligan[3], Jian Yang[13], Sharon L.R. Kardia[5], Lenore J. Launer[14], Yi-Ping Fu[15,16], Thomas H. Mosley Jr[7], Jill M. Norris[17], James G. Terry[12], Christopher J. O'Donnell[1], Jerome I. Rotter[8], Lynne E. Wagenknecht[19], Vilmundur Gudnason[9,20,¶], Michael A. Province[4,¶], Patricia A. Peyser[5,§] and Elizabeth K. Speliotes[3,*,§]

[1]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA, [2]Centre for Brain Research, Indian Institute of Science, Bangalore, Karnataka, India, [3]Department of Internal Medicine, Division of Gastroenterology and Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA, [4]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA, [5]Department of Epidemiology, University of Michigan, Ann Arbor, MI, USA, [6]Department of Endocrinology, Diabetes, and Nutrition, University of Maryland-Baltimore, Baltimore, MD, USA, [7]Department of Medicine, University of Mississippi, Jackson, MS, USA, [8]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA USA, [9]Icelandic Heart Association, Kopavogur, Iceland, [10]Department of Family Medicine and Public Health, University of California, San Diego, CA, USA, [11]Department of Internal Medicine, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA, [12]Department of Radiology, Vanderbilt University School of Medicine, Nashville, TN, USA, [13]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia, [14]Laboratory of Epidemiology and Population Sciences, National Institute of Aging, Bethesda, MD,

[†]These authors have contributed equally.
[$]These authors have contributed equally.
[¶]These authors have contributed equally.

Received: November 20, 2020. **Revised:** March 19, 2021. **Accepted:** March 31, 2021

© The Author(s) 2021. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

USA, [15]Framingham Heart Study, NHLBI, NIH, Framingham, MA, USA, [16]Office of Biostatistics Research, NHLBI, NIH, Bethesda, MD, USA, [17]Department of Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA, [18]VA Boston Healthcare System, Boston, MA, USA, [19]Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA and [20]Department of Medicine, University of Iceland, Reykjavik 101, Iceland

*To whom correspondence should be addressed at: Department of Biochemistry, Wake Forest School of Medicine, 1 Medical Center Blvd, Winston-Salem, NC 27157, USA. Tel: +336-713-7534; Email: nallred@wakehealth.edu; Divisions of Gastroenterology, and Computational Medicine and Bioinformatics, Department of Internal Medicine, University of Michigan, Ann arbor, MI 48109, USA. Tel: +734-647-2964; Email: espeliot@med.umich.edu

Abstract

Nonalcoholic fatty liver disease (NAFLD) is a leading cause of chronic liver disease and is highly correlated with metabolic disease. NAFLD results from environmental exposures acting on a susceptible polygenic background. This study performed the largest multiethnic investigation of exonic variation associated with NAFLD and correlated metabolic traits and diseases. An exome array meta-analysis was carried out among eight multiethnic population-based cohorts ($n = 16\,492$) with computed tomography (CT) measured hepatic steatosis. A fixed effects meta-analysis identified five exome-wide significant loci ($P < 5.30 \times 10^{-7}$); including a novel signal near *TOMM40/APOE*. Joint analysis of *TOMM40/APOE* variants revealed the *TOMM40* signal was attributed to *APOE* rs429358-T; *APOE* rs7412 was not associated with liver attenuation. Moreover, rs429358-T was associated with higher serum alanine aminotransferase, liver steatosis, cirrhosis, triglycerides and obesity; as well as, lower cholesterol and decreased risk of myocardial infarction and Alzheimer's disease (AD) in phenome-wide association analyses in the Michigan Genomics Initiative, United Kingdom Biobank and/or public datasets. These results implicate *APOE* in imaging-based identification of NAFLD. This association may or may not translate to nonalcoholic steatohepatitis; however, these results indicate a significant association with advanced liver disease and hepatic cirrhosis. These findings highlight allelic heterogeneity at the *APOE* locus and demonstrate an inverse link between NAFLD and AD at the exome level in the largest analysis to date.

Introduction

Nonalcoholic fatty liver disease (NAFLD) is rapidly becoming the most common cause of chronic liver disease worldwide. It is caused by accumulation of fat in the liver (steatosis) and can progress to nonalcoholic steatohepatitis (NASH), fibrosis and cirrhosis, in the absence of excessive alcohol ingestion (1).

NAFLD is suggested to be the hepatic manifestation of metabolic syndrome (MetS), a constellation of risk factors including abdominal adiposity, dyslipidemia, hypertension and impaired glucose tolerance (2,3). MetS increases the likelihood for developing cardiometabolic diseases including type 2 diabetes (T2D) (4) and cardiovascular disease (CVD) (5). It is also associated with Alzheimer's disease (AD) (6). The underlying causes of NAFLD and predisposition to metabolic and other human diseases are still unknown.

NAFLD is partially mediated by genetics. Heritability of computed tomography (CT) measured hepatic steatosis, as measured using liver attenuation, is 20–34% across populations (7). To date, genome-wide association studies (GWAS) have identified common variants associated with NAFLD (8–10). One exome-wide study ($N = 2736$) identified variants in *PNPLA3* and *TM6SF2* associated with hepatic steatosis (11). Taken together, these results do not completely explain the heritability of disease suggesting additional variants have yet to be identified.

To identify potentially functional, coding variants associated with NAFLD, an exome-wide association meta-analysis was performed. Liver attenuation was measured by CT in eight independent population-based cohorts from multiple ancestries. Results were additionally stratified by sex and ancestry. Conditional analyses were performed to examine individual and joint effects at significant loci. To evaluate the effect of NALFD-associated variants on metabolic traits and selected diseases, targeted association analyses were performed in the Michigan Genomics Initiative (MGI) and United Kingdom Biobank (UKBB). These data provide new insight into the genetic architecture of NAFLD and identify molecular links between NAFLD and metabolic disease.

Results

Study sample

The Genetics of Obesity-related Liver Disease (GOLD) Consortium consists of eight cohorts. Study sample characteristics and quality control are shown in Supplementary Material, Table S1. In general, participants were middle to older aged adults, although participants from IRASFS were younger (mean age ~43 versus mean 53–76 years in other cohorts), with a higher proportion of women than men. Among the four ancestries (European, African American, Hispanic American and Chinese American), Hispanic Americans had the lowest mean CT liver attenuation values indicating increased hepatic steatosis. Alcohol intake was generally low across all cohorts (Supplementary Material, Table S1).

Association analysis

Exome-wide association analysis of CT-measured liver attenuation identified 19 variants from five regions that were exome-wide significant, $P < 5.30 \times 10^{-7}$ (Bonferroni correction for 93 621 variants; Supplementary Material, Fig. S1). Analyses revealed a novel exome-wide associated intronic variant in *TOMM40* (rs2075650, $P = 9.66 \times 10^{-8}$). Known exonic associations included missense variants in *PNPLA3* (I148M, rs738409, $P = 1.38 \times 10^{-92}$), *TM6SF2* (E167K, rs58542926, $P = 5.52 \times 10^{-15}$), and *GCKR* (L446P,

Table 1. Variants associated with liver attenuation at exome array-wide significance

SNP ID	Position	EA/OA	EAF	Effect (beta)	Effect (SE)	P-value	N	Gene	Variant annotation
rs738409	22:44324727	G/C	0.3	−0.27	0.01	1.38E-92	16 492	*PNPLA3*	I148M missense
rs4841132	8:9183596	G/A	0.9	−0.22	0.02	2.64E-29	15 744	*LOC157273*	noncoding RNA exonic
rs58542926	19:19379549	T/C	0.1	−0.18	0.02	5.52E-15	16 492	*TM6SF2*	E167K missense
rs1260326	2:27730940	T/C	0.4	−0.08	0.01	8.35E-10	16 488	*GCKR*	L446P missense
rs2075650	19:45395619	A/G	0.9	−0.09	0.02	9.66E-08	16 491	*TOMM40*	Silent Intronic

Chr: Position: chromosome and position, build 37; EA: effect allele; OA: other allele; EAF: effect allele frequency; Effect Beta: effect on inverse normalized liver attenuation by computed tomography; Effect SE: standard error of the effect; P-value: P-value of association in ancestry-combined meta-analysis; N: total number of samples; Gene: gene location of variant; Variant Annotation: protein coding or noncoding changes in transcript of genes caused by the exome-wide variants.

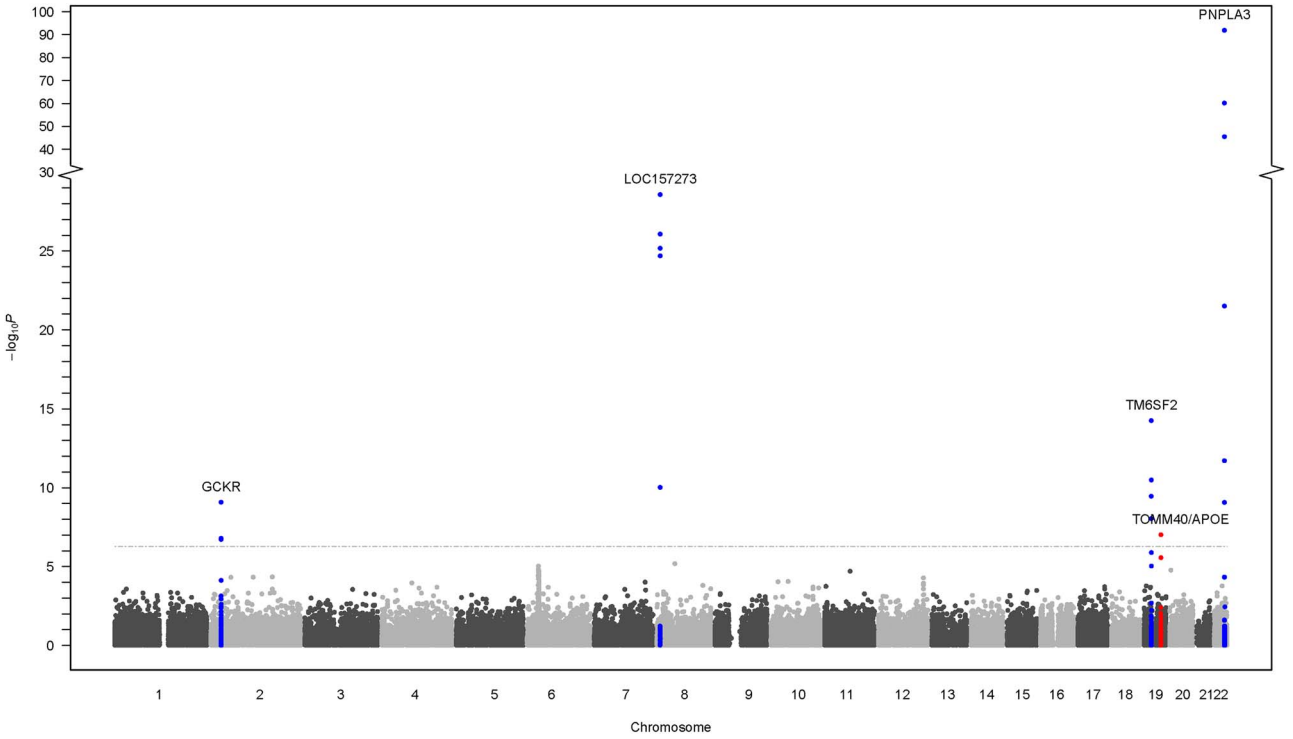

Figure 1. Ancestry and sex combined association meta-analysis plot for CT-measured liver attenuation. Manhattan plot showing previously identified loci in blue and novel loci in red. Loci are labeled with the nearest gene. X-axis: position of SNP on each chromosome. Y-axis: association with CT-measured liver attenuation as $-\log_{10}$P-value. Dotted line: exome-wide significance of 5.34×10^{-7}

rs1260326, $P = 8.35 \times 10^{-10}$) and rs4841132 in the long noncoding RNA *LOC157273* ($P = 2.64 \times 10^{-29}$) (Table 1 and Fig. 1). Variants in these five loci accounted for 4.6% of the variance in liver attenuation, as assessed from a meta-analysis of all individuals across all ancestries. The percent variance ranged from 2.81% for rs738409 in *PNPLA3* to 0.18% for rs2075650 in *TOMM40*. A quantile–quantile plot of NAFLD susceptible variants is shown in Supplementary Material, Fig. S1.

Effects by ancestry

There was no significant heterogeneity of effects across ancestries for the exome-wide significant variants (Fig. 2). Across ancestries, the frequency of the liver attenuation decreasing variants were similar for *LOC157273*, *TOMM40* and *TM6SF2*. The frequency of the liver attenuation decreasing allele in *PNPLA3* (rs738409) was higher in Hispanic Americans (41%) and Chinese Americans (38%) than in European Americans (23%) and African Americans (14%), similar to previous observations (8); whereas the effect was strongest in African Americans (beta = −0.32), followed by European (beta = −0.27), Hispanic (beta = −0.24) and Chinese (beta = −0.19) Americans, respectively. Consequently, the variance explained by *PNPLA3* (rs738409) was the highest in Hispanic Americans (2.91%), followed by European Americans (2.67%) and African Americans (2.40%) with the lowest in Chinese Americans (1.72%) (Supplementary Material, Table S3). Therefore, the smaller effect size of rs738409 in Chinese Americans contributes to the lower variance explained even though the frequency of the effect allele is closer to that of the highest effect group, i.e. Hispanic Americans. The fatty liver promoting allele in *GCKR* (rs1260326) was of highest frequency in Chinese Americans (50%) followed by European Americans (41%) then Hispanic Americans (35%) and African Americans (14%). The effect size was also the strongest in Chinese Americans with the largest standard error (beta = −0.22 ± 0.07) attributed to modest sample size ($n = 360$). Consequently, this ancestry also had the highest proportion of variance explained (2.35%), followed by African Americans (0.22%), European Americans

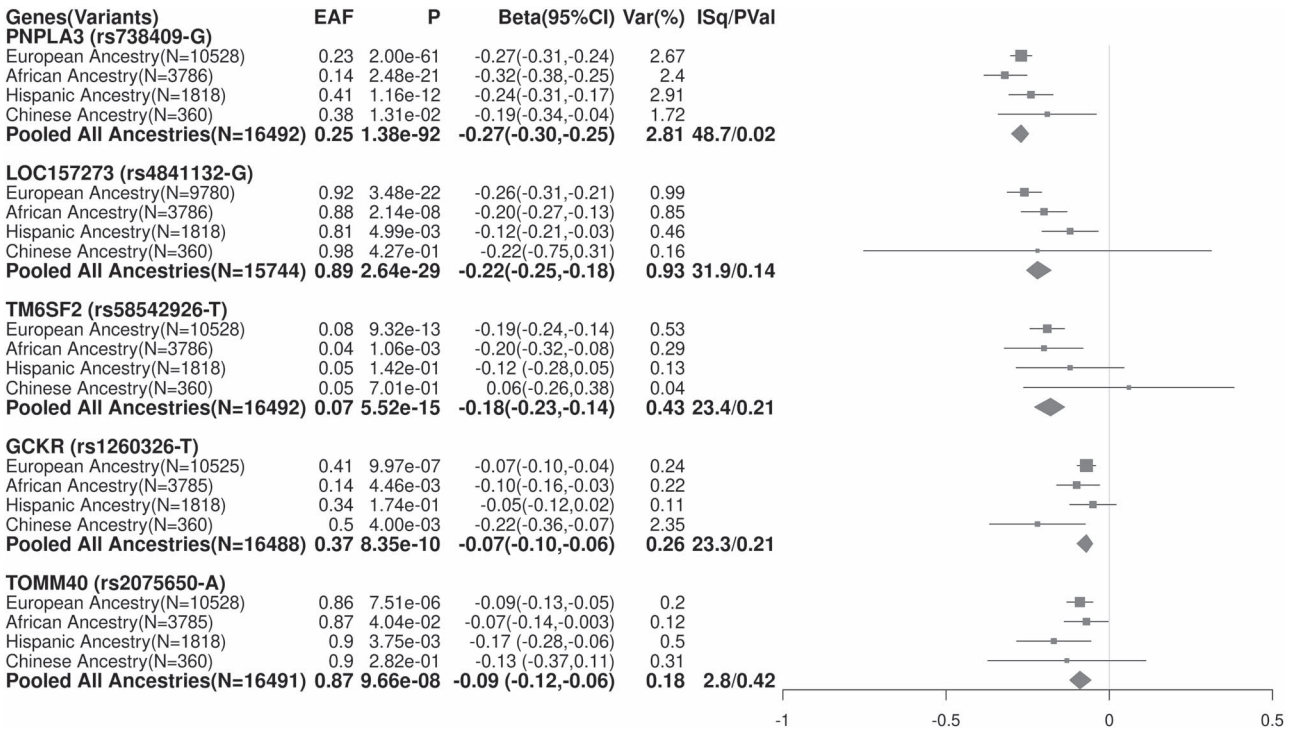

Figure 2. Ancestry-specific sex combined effects. EA: Effect allele; EAF: Effect allele frequency; P: P-value of association; Beta (95% CI): Effect (beta) on inverse normalized liver attenuation by computed tomography and 95% confidence interval of effect; HetISq: heterogeneity I^2 value; HetPVal: P-value for heterogeneity across samples. Lower liver attenuation is associated with increased hepatic steatosis. The solid vertical line represents beta = 0.

(0.24%), and Hispanic Americans (0.11%). European and African Americans have a comparable proportion of the variance explained due to a stronger effect in African Americans (beta = −0.10) compared to European Americans (beta = −0.07) compensating for their allele frequency differences. Thus, the overall variance explained by these five loci was highest in European Americans, closely followed by Chinese Americans, Hispanic Americans and then African Americans at 4.63%, 4.58%, 4.10% and 3.88%, respectively (Supplementary Material, Table S3). This result was driven by differences in allele frequencies and their corresponding stronger or modest effect size of the NAFLD promoting allele in individuals of diverse ancestries. It remains to be determined if the highest variance explained is the ancestry-specific effect sizes-allele frequency combination for the Chinese American or random variation secondary to the smaller sample size compared to other ancestries.

Effects by sex

Sex-specific analyses were conducted in a manner similar to above except without sex as a covariate. There was no significant heterogeneity for effects for any of the five loci across ancestries in men or women meta-analyzed separately. Furthermore, there was no evidence for significant heterogeneity between men and women at any of the five loci when all ancestries were combined (Fig. 3a and b and Supplementary Material, Table S4 and Supplementary Material, Table S5).

Identification of secondary signals

To identify secondary signals at exome-wide significant loci, we performed approximate joint and conditional SNP association analysis using summary statistics in European and African ancestries separately controlling for exome-wide significant variants (Table 2). In African Americans, an intronic variant (rs2281135, effect allele frequency of 14%) and a missense variant (rs6006460, S453I, effect allele frequency of 90%) in *PNPLA3* were associated with liver attenuation ($P = 0.0070$ and 0.0040, respectively) after conditioning on rs738409. In Hispanic Americans using direct conditional analysis, a missense variant (rs142056540, R138W, effect allele frequency of 1%) in *TM6SF2* was associated with liver attenuation ($P = 0.018$) after conditioning on rs58542926.

Identification of causal variants at *TOMM40/APOE*

The *TOMM40* associated variant, rs2075650, is not in linkage disequilibrium (LD; $r^2 = 0.008$) with rs7412 but has moderate LD ($r^2 = 0.48$) with rs429358 in European population data from 1000 Genomes (phase 3). These two SNPs constitute the three common alleles in *APOE* (ApoE2 (rs429358-T, rs7412-T), ApoE3 (rs429358-T, rs7412-C) and ApoE4 (rs429358-C, rs7412-C). *APOE* rs7412 was included on the exome chip across cohorts, but unlike rs2075650 showed no significant association with CT-measured liver attenuation ($P = 0.89$; Supplementary Material, Table S6). *APOE* rs429358 was not genotyped on the exome chip but was imputed in UKBB. Using a surrogate measure of liver function, alanine aminotransferase (ALT), joint analyses provided evidence that *TOMM40* rs2075650 was not significant ($P = 0.074$) when conditioned on *APOE* rs429358; however, *APOE* rs429358 remained significant ($P = 1.59 \times 10^{-14}$) when conditioned on *TOMM40* rs2075650 (Supplementary Material, Table S7 and Supplementary Material, Table S8). This suggests that variation at *APOE* rs429358 is driving the observed association.

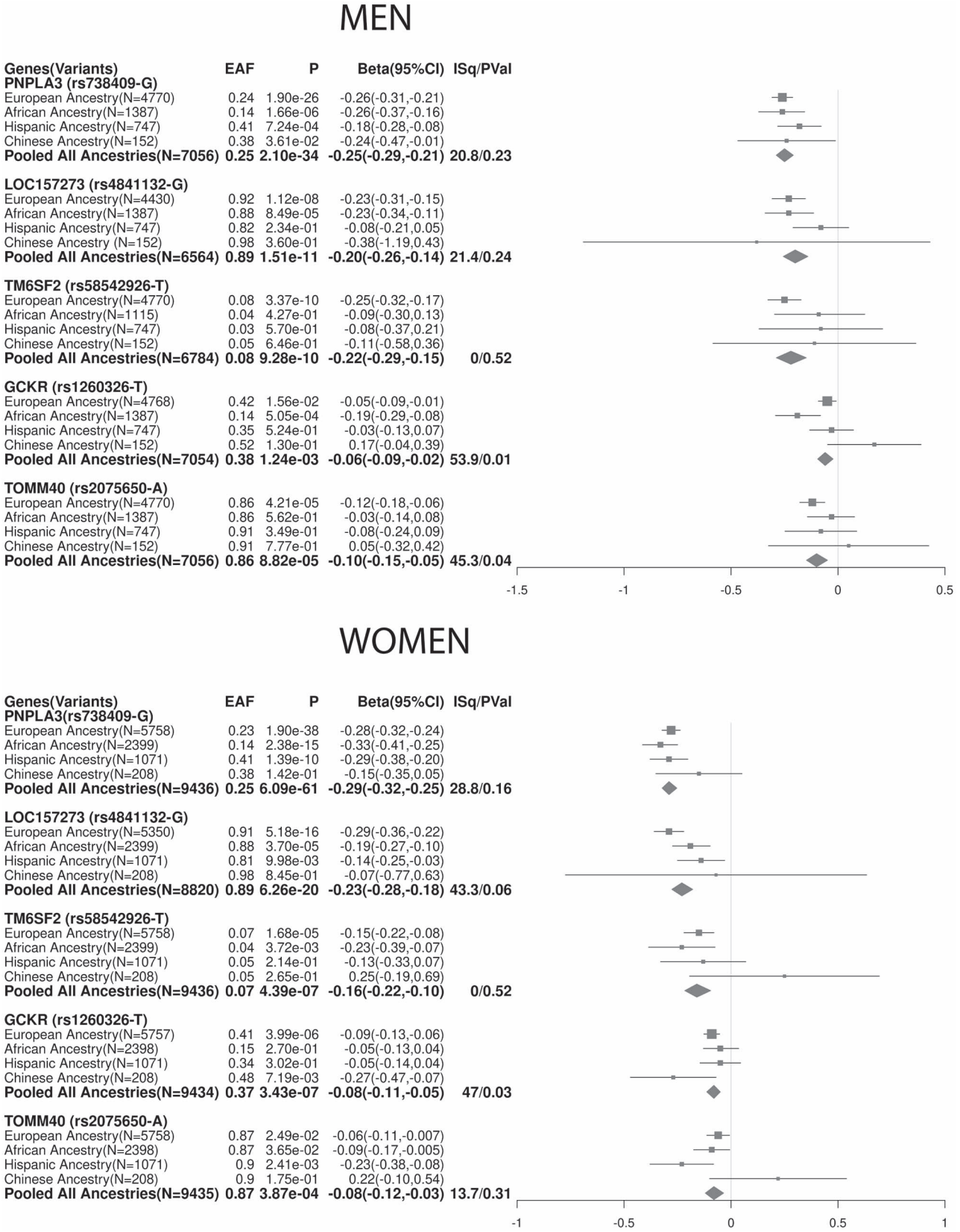

Figure 3. Ancestry-specific and sex-specific effects. (A) Men and B. Women. EA: Effect allele; EAF: Effect allele frequency; P: *P*-value of association; Beta (95% CI): Effect (beta) on inverse normalized liver attenuation by computed tomography and 95% confidence interval of effect; HetISq: heterogeneity I^2 value; HetPVal: *P*-value for heterogeneity across samples. Lower liver attenuation is associated with increased hepatic steatosis. The solid vertical line represents beta = 0. Tabulated data are presented in Supplementary Material, Table S4.

Table 2. Ancestry-specific second signals at liver attenuation associated loci

SNP ID	Chr: Position	EA/OA	EAF	Effect beta (SE)	P-value	N	Cond effect beta (SE)	Cond P-value	Imputation	Gene	Annotation	Ancestry
rs2281135	22:44332570	A/G	0.14	−0.18 (0.04)	5.80E-06	2526	0.08 (0.03)	7.00E-03	1	*PNPLA3*	intronic	African American
rs6006460	22:44342174	G/T	0.9	−0.17 (0.04)	3.00E-05	3850	−0.12 (0.04)	4.00E-03	0.918	*PNPLA3*	S453I missense	African American
rs142056540	19:19380568	A/G	0.01	−0.58 (0.25)	1.80E-02	895	−0.58 (0.25)	1.80E-02	Genotyped	*TM6SF2*	R138W missense	Hispanic American

Chr: Position: chromosome and position, build 37; EA/OA: effect allele/other allele; EAF: effect allele frequency; Effect Beta (SE): effect and standard error on inverse normalized liver attenuation by computed tomography; P-value: P-value of association in ancestry-combined meta-analysis; N: total number of samples; Cond Effect Beta (SE): conditional effect and standard error on inverse normalized liver attenuation by computed tomography; Cond P-value: conditional P-value of association of SNP; Imputation: imputation quality score in reference population; Gene: gene location of variant; Annotation: protein coding or noncoding changes in transcript of genes caused by the exome-wide variants.

Phenome-Wide Association analysis (PheWAS) of *APOE* rs429358

Effects of *APOE* rs429358-T were examined in UKBB and found to be associated with markers of reduced liver health including increased ALT ($P = 1.97 \times 10^{-43}$), liver steatosis on magnetic resonance imaging (MRI) ($P = 1.20 \times 10^{-6}$), clinical NAFLD ($P = 1.21 \times 10^{-4}$) and cirrhosis ($P = 0.028$). Consistent with these associations, the same allele was associated with increased body mass index (BMI) ($P = 9.54 \times 10^{-14}$), waist circumference ($P = 2.67 \times 10^{-17}$), WHRcontBMI ($P = 6.57 \times 10^{-15}$), risk of T2D ($P = 0.0019$), pancreatitis ($P = 0.0044$) and cholelithiasis ($P = 0.0016$). Strikingly, the same allele was associated with lower levels of total cholesterol ($P < 1 \times 10^{-324}$), low-density lipoprotein (LDL) cholesterol ($P < 1 \times 10^{-324}$), apolipoprotein B ($P < 1x10^{-324}$), and higher levels of high-density lipoprotein (HDL) cholesterol ($P = 7.12 \times 10^{-140}$), C-reactive protein ($P < 1 \times 10^{-324}$), and apolipoprotein A ($P = 8.53 \times 10^{-188}$). Consistent with prior literature, *APOE* rs429358-T was associated with a lower risk of AD ($P = 1.62 \times 10^{-65}$), and ischemic heart disease ($P = 1.48 \times 10^{-15}$) (Fig. 4a, Supplementary Material, Table S7 and Supplementary Material, Table S8). Interestingly, the same pattern of associations, both positive and negative, are also seen for rs2075650-A (Fig. 4a, Supplementary Material, Table S7 and Supplementary Material, Table S8). In contrast, rs7412-T was more weakly associated ALT, had a positive association with triglyceride levels, negative association with glucose levels and WHRcontBMI, and had stronger negative associations with systolic and diastolic blood pressure compared to the other two SNPs. Importantly, rs7412-T was not associated with liver-fat, T2D and pancreatitis. The inferences regarding rs429358-T persisted after conditioning on rs2075650 and rs7412 (Fig. 4b, Supplementary Material, Table S7 and Supplementary Material, Table S8). All associations with rs2075650 diminished with the exception of ischemic heart disease and glucose which were both weakly associated after conditioning on the other two SNPs. The associations with rs7412 remained unchanged after conditioning on the other two SNPs (Fig. 4b, Supplementary Material, Table S7 and Supplementary Material, Table S8).

APOE rs429358-T was also associated with increased ALT ($P = 0.0025$), HDL cholesterol ($P = 1.61 \times 10^{-04}$), and decreased serum LDL cholesterol ($P = 3.05 \times 10^{-15}$), total cholesterol ($P = 9.84 \times 10^{-10}$), and triglyceride ($P = 0.010$) and decreased risk of AD ($P = 1.74 \times 10^{-05}$) and ischemic heart disease ($P = 0.0013$) in the less powered MGI cohort (Supplementary Material, Table S9). A literature review of publicly available GWAS analyses supported the association of *APOE* rs429358-T with increased risk of T2D and decreased risk of myocardial infarction and AD (Supplementary Material, Table S10).

APOE is expressed in the brain and arteries but also in adipose tissue, liver, adrenal, kidney and pancreas in data available from GTEx (https://gtexportal.org/) (Supplementary Material, Fig. S2). Presence of *APOE* rs429358-T corresponded to lower triglyceride, phospholipid and cholesterol in extra-extra-large, extra-large and large VLDL particles compared to rs7412-T in metabolomics studies (Fig. 5, Supplementary Material, Table S11). For most of the other lipid traits, the effects are concordant (Supplementary Material, Fig. S3, Supplementary Material, Table S11). We show a model of the effect of T and C alleles of rs429358 on human diseases and traits in Figure 6.

Discussion

In our multiethnic exome-wide association study of CT-measured liver attenuation we replicated the associations in four known loci including rs738409 (*PNPLA3*), rs58542926 (*TM6SF2*) and rs1260326 (*GCKR*) and rs4841132 in the long noncoding RNA *LOC15723* (12). Importantly, we found a new exome-wide significant signal at *TOMM40*, an intronic variant rs2075650. For the first time, we show that the associations at these five loci are consistent when stratified by sex and/or ancestry.

Ancestry-specific conditional analysis identified two variants, *PNPLA3* rs2281135 and rs6006460, that were nominally associated with hepatic steatosis in African-ancestry participants, with the latter previously reported (8). rs2281135 is in high LD with rs738409 ($r^2 = 0.724$) and has been associated with ALT in Europeans, South Asians and Hispanic Americans (13,14). rs6006460 has previously been associated with hepatic steatosis in African Americans (8) representing the first replication at this locus. This SNP has not been associated with any traits in the GWAS catalog. At the *TM6SF2* locus, we also identified a novel variant, rs142056540, after conditioning on rs58542926 in Hispanic-ancestry individuals. Although this variant is of low frequency in our study (<1%), it is rare in all existing databases, e.g. BRAVO (https://bravo.sph.umich.edu/). No studies to date have reported an association between this variant and any trait or disease.

Previous GWAS have reported associations between *TOMM40* rs2075650 and various traits including BMI (15), waist circumference and waist-to-hip ratio (16), C-reactive protein (17), total cholesterol and LDL cholesterol (18) as well as AD and associated biomarkers (19) as documented in the GWAS catalog. These associations are consistent with our findings. Until now, no genome-wide studies have reported associations between rs2075650 and direct measures of liver disease, including hepatic steatosis as we found.

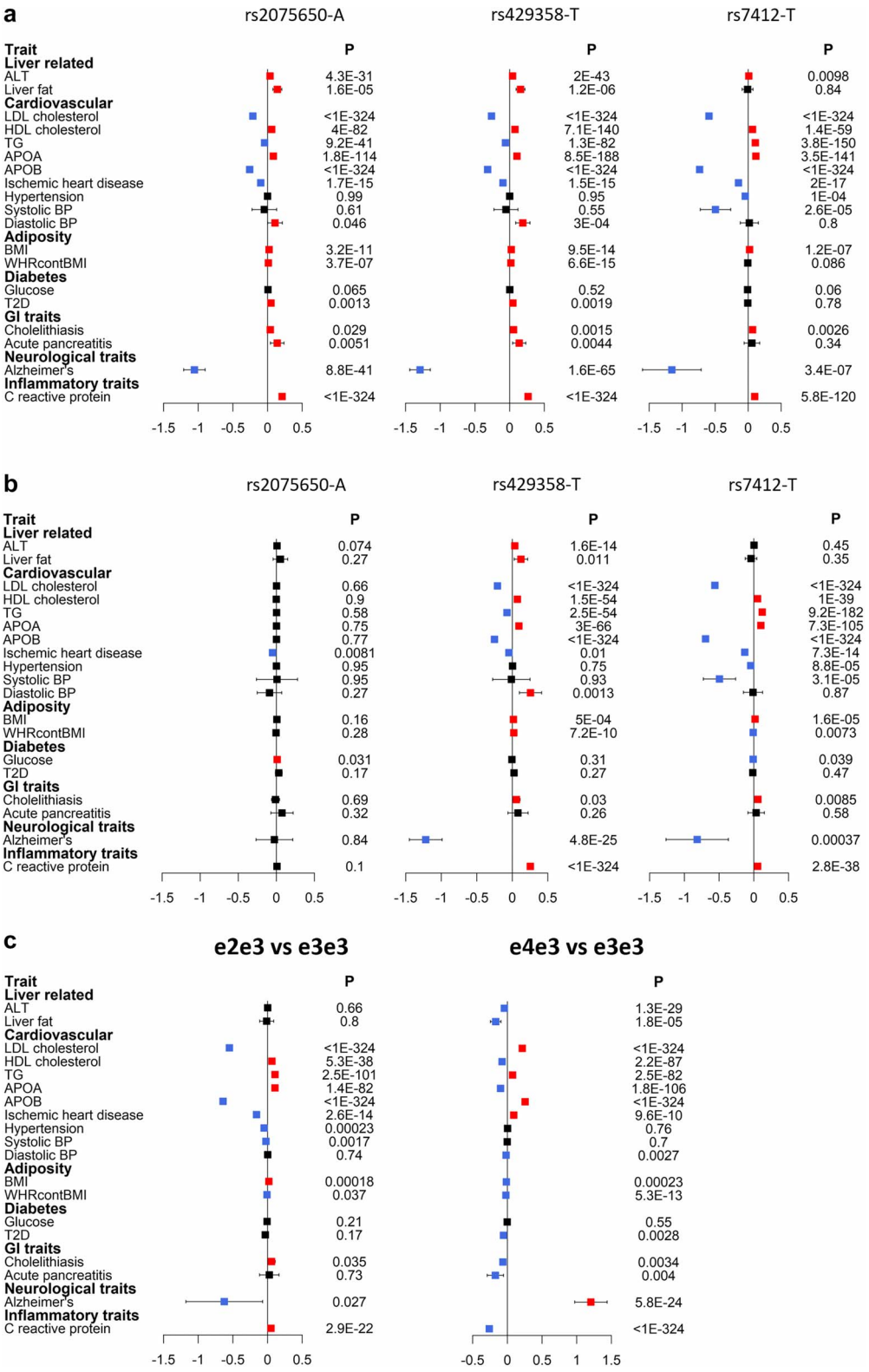

Figure 4. Effect of rs2075650, rs429358 and rs7412 on human diseases and traits. Forest plot showing all associations between the variants and human diseases and traits in the UKBB. Effects are in SD for continuous traits and log (OR) for disease outcomes per copy of the risk allele. Levels of significance: $P < 0.05$. We used the genotype data of the three variants in UKBB and ran a generalized linear regression model (GLM) to get the betas and *P*-values shown in the plot, using only unrelated white individuals. Vertical dotted line indicates and effect size = 0. Significant effect <0 in blue (indicating that the liver-fat promoting allele decreases the effect) and effect size >0 in red (indicating that the liver-fat promoting allele increases the effect), while the traits having nonsignificant effects are shown in black. ALT: alanine transaminase; LDL: low-density lipoprotein cholesterol, HDL: high-density lipoprotein cholesterol; TG: triglycerides; APOA: apolipoprotein A; APOB: apolipoprotein B; BP: blood pressure; BMI: body mass index; WHRcontBMI: waist-to-hip ratio adjusted for body mass index; T2D: type 2 diabetes. Analyses performed in UKBB for (A) each SNP (rs2075650-A, rs429358-T and rs7412-T) alone, (B) for each SNP in joint analysis with the other two SNPs and (C) for haplotype combinations of rs429358 and rs7412 with tabulated data presented in Supplementary Material, Table S19.

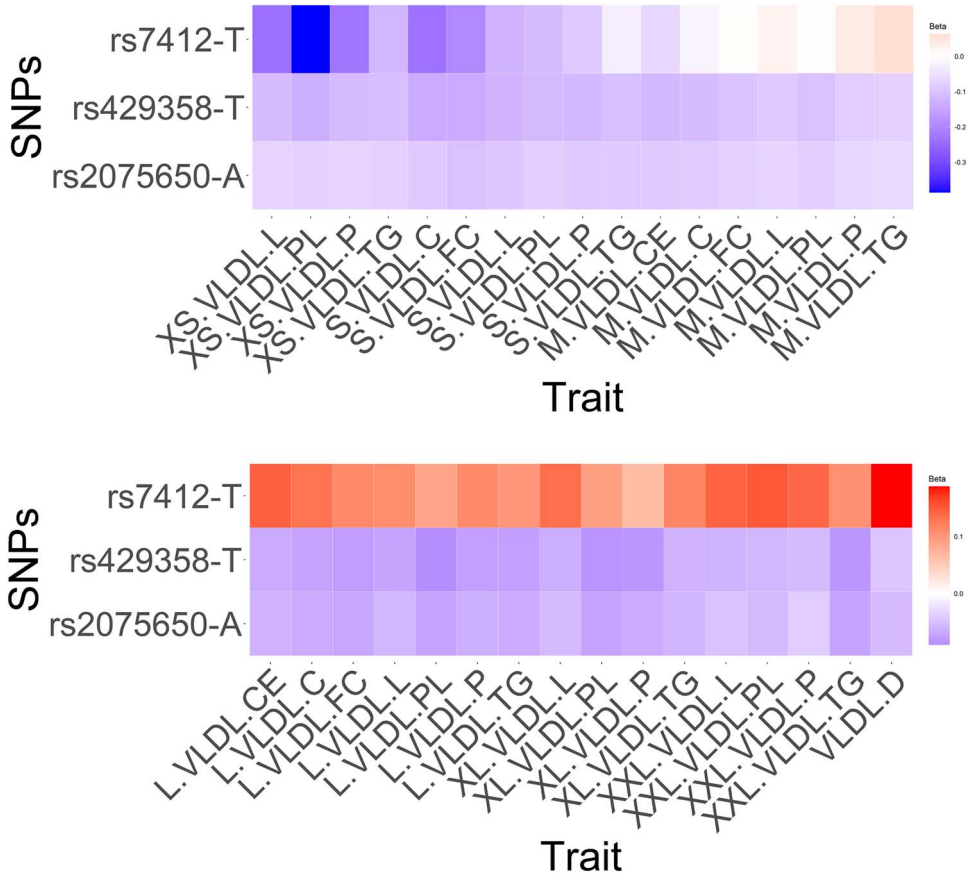

Figure 5. Effect of rs2075650, rs429358 and rs7412 on serum metabolites. Heatmap plot of the effect of rs2075650-A, rs429358-T and rs7412-T across serum metabolites from Kettunen *et al.* (23). Red indicates the allele increases the trait; blue indicates the allele decreases the trait. Results are shown for traits with $P < 4 \times 10^{-4}$. Traits listed include: XS.VLDL.L: total lipids in very small VLDL; XS.VLDL.PL: phospholipids in very small VLDL; XS.VLDL.P: concentration of very small VLDL particles; XS.VLDL.TG: triglycerides in very small VLDL; S.VLDL.C: total cholesterol in small VLDL; S.VLDL.FC: free cholesterol in small VLDL; S.VLDL.L: total lipids in small VLDL; S.VLDL.PL: phospholipids in small VLDL; S.VLDL.P: concentration of small VLDL particles; S.VLDL.TG: triglycerides in small VLDL; M.VLDL.CE: cholesterol esters in medium VLDL; M.VLDL.C: total cholesterol in medium VLDL; M.VLDL.FC: free cholesterol in medium VLDL; M.VLDL.L: total lipids in medium VLDL; M.VLDL.PL: phospholipids in medium VLDL; M.VLDL.P: concentration of medium VLDL particles; M.VLDL.TG: triglycerides in medium VLDL; L.VLDL.CE: cholesterol esters in large VLDL; L.VLDL.C: total cholesterol in large VLDL; L.VLDL.FC: free cholesterol in large VLDL; L.VLDL.L: total lipids in large VLDL; L.VLDL.PL: phospholipids in large VLDL; L.VLDL.P: concentration of large VLDL particles; L.VLDL.TG: triglycerides in large VLDL; XL.VLDL.L: total lipids in very large VLDL; XL.VLDL.PL: phospholipids in very large VLDL; XL.VLDL.P: concentration of very large VLDL particles; XL.VLDL.TG: triglycerides in very large VLDL; XXL.VLDL.L: total lipids in chylomicrons and extremely large VLDL; XXL.VLDL.PL: phospholipids in chylomicrons and extremely large VLDL; XXL.VLDL.P: concentration of chylomicrons and extremely large VLDL particles; XXL.VLDL.TG: triglycerides in chylomicrons and extremely large VLDL; VLDL.D: mean diameter for VLDL particles. For full trait names see Supplementary Material, Fig. S3.

We identified a missense variant (rs429358 (C158A)) in *APOE* with an $r^2 = 0.48$ with rs2075650. We showed by joint analysis that the association signal at *TOMM40* rs2075650 was eliminated when conditioned on *APOE* rs429358, but the reverse was not true. This suggests that variation at rs429358 is driving the observed association with hepatic steatosis. In contrast, another *APOE* missense variant (rs7412 (C112A)) does not associate with liver steatosis. The LD between these two missense *APOE* variants across all populations is low ($r^2 < 0.015$). When a joint analysis of rs2075650, rs7412 and rs429358 in UKBB was performed, association results with ALT, a surrogate measure of liver function, were similar to those with rs429358 being more likely causal. Our findings at rs429358, with respect to hepatic steatosis, are consistent with small case-control studies using a candidate gene approach. Carriers of the C allele had lower risk of hepatic steatosis whereas carriers of the more common T allele had increased risk of hepatic steatosis by ultrasound, histology or liver enzyme measurement (20–22).

In UKBB, rs429385-T was associated with increased ALT, liver-fat on MRI, NAFLD and cirrhosis. It was additionally associated with increased BMI, waist circumference, WHRcontBMI, T2D, pancreatitis and cholelithiasis. This same allele associated with lower levels of total cholesterol, LDL cholesterol, apolipoprotein B, triglycerides, and increased levels of HDL cholesterol, C-reactive protein, and apolipoprotein A. Consistent with prior literature, *APOE* rs429358-T was associated with a lower risk of AD and ischemic heart disease. Many of these associations were replicated in MGI. Even after conditioning on rs2075650 and rs7412 in the UKBB, associations with rs429385 continued to be significant except for T2D and pancreatitis.

Our findings demonstrate allelic heterogeneity at *APOE* as both missense variants associate with different diseases/traits. Although the variants have similar effects on LDL and HDL cholesterol, ischemic heart disease, AD and C-reactive protein, the variants have opposing effects on hepatic steatosis and serum triglycerides. In metabolomics studies, rs429358-T has less triglyceride, phospholipid and cholesterol in extra-extra-large, extra-large and large VLDL particles compared to rs7412-T (23). rs429358-T may promote tissue accumulation of lipids in organs and tissues which may explain why this allele, but not

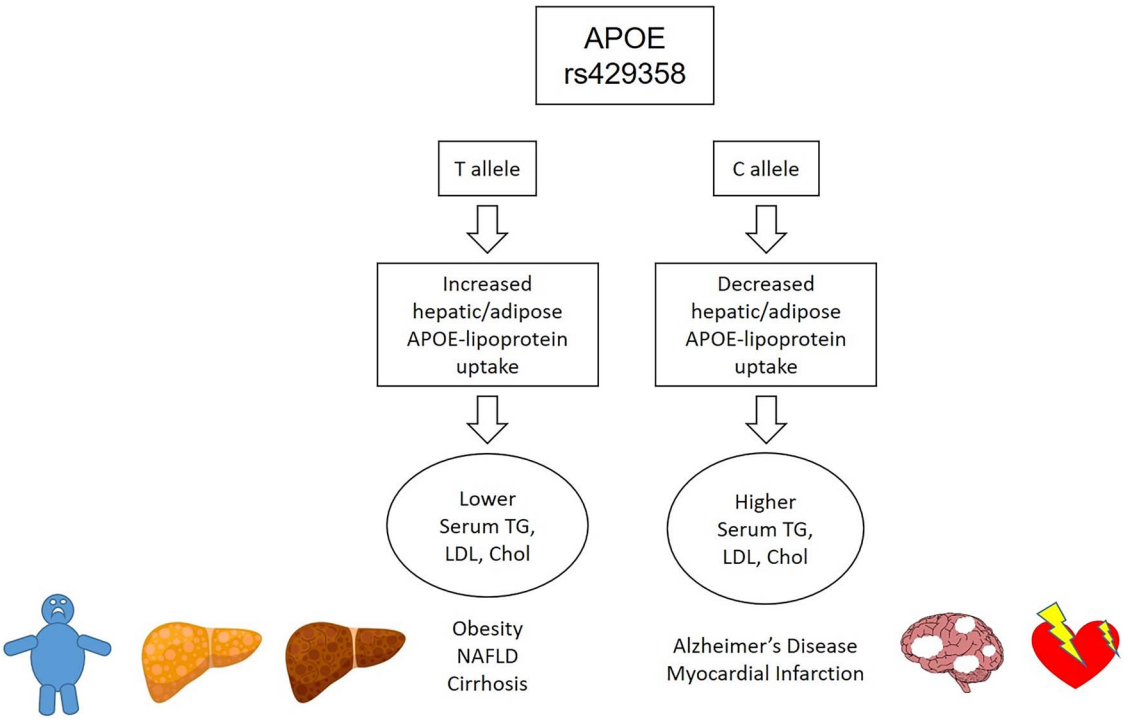

Figure 6. Model of T and C alleles of rs429358 (APOE C158A) on human diseases and traits. Model showing the differences between the effects of the alleles of rs429358 on selected human diseases and traits. TG: triglycerides; LDL: low-density lipoprotein; Chol: cholesterol; NAFLD: nonalcoholic fatty liver disease.

rs7412-T, promotes hepatic steatosis, liver disease and abdominal obesity.

Further, we show how carrying a specific allele is not universally beneficial or harmful. *APOE* rs429358-T promotes obesity and tissue retention of triglycerides and cholesterol which may be causally related to development of hepatic steatosis, T2D, pancreatitis, cirrhosis, increased C-reactive protein and cholelithiasis. This same allele protects against MI and AD. Whether the protective effect is due to its effects on VLDL composition or some other aspect of its biology is unknown. rs7412-T also leads to tissue retention of cholesterol which promotes increased C-reactive protein and cholelithiasis but not other diseases and protects against MI and AD. Since both rs429358-T and rs7412-T promote cholelithiasis and protect against MI and AD these effects are likely due to shared effects.

The observation that carriers of rs429358-T have increased risk of NAFLD but lower risk of AD contrasts with studies reporting that increased NAFLD increases brain dysfunction in both mice and humans. In mice, wild-type mice fed a chronic high-fat diet developed NAFLD and secondary neuroinflammation sufficient to cause neurodegeneration (24). In cross-sectional analysis of 4472 adults aged 20–59 years from the Third National Health and Nutritional Examination Survey, NAFLD was independently associated with lower cognitive performance (e.g. *inferior* learning, recall and concentration function) independent of CVD and its risk factors (25). NAFLD has also been associated with reduced brain volume (26). A recent prospective study showed that altered liver function markers are associated with an AD diagnosis (27). At the physiological level, insulin resistance plays a major role in both AD and NAFLD (28,29). It has been suggested that AD should be regarded as an insulin-resistance-mediated neurodegenerative disorder with the same fundamental abnormalities seen in T2D, MetS and NAFLD (30). Furthermore, the liver is the origin of brain amyloid-deposits, a primary component of the amyloid plaques in AD, and it is involved in peripheral clearance of circulating amyloid beta Aβ in the blood. ApoE4 inhibits Aβ clearance. In our study, however, which represents the most powered analyses of hepatic steatosis and its effects on metabolic traits, liver measures and other human diseases, we show that individuals carrying rs429358-T are predisposed to NAFLD but with lower risk of CVD and AD. This is, to our knowledge, the first report of such an opposite effect of NAFLD, CVD and AD by a single allele.

Genetic effects on the same diseases and traits, even those with shared risk factors and physiology, can differ. Such differences would depend on the allele effect sizes and frequencies. rs429358-C explains 13% of AD (31) but $<$0.25% of variance in liver-fat in UKBB. Thus, at the population level, many other factors increase risk of both liver-fat and AD. It is important to note, that almost 40% of those with AD do not carry rs429358-C, and the vast majority of those with AD but without an rs429358-C allele are rs7412-C individuals (32). In recent mouse studies, ApoE3 (rs429358-T rs7412-C) mice on a high-fat diet accumulated more visceral fat than ApoE4 (rs429358-C rs7412-C) mice on a high-fat diet, consistent with the finding that rs429358-T promotes abdominal obesity (33). Thus, an allele could protect against AD and CVD yet promote tissue steatosis including in viscera and liver.

APOE is synthesized primarily in the liver, but also in the brain, muscle tissues and the central and peripheral nervous system. It is an important component of VLDL and a ligand for the LDL receptor. It also plays an important role in oxidation, neuronal repair, regulation of lipid homeostasis and transport and metabolism of triglycerides and cholesterol. It has anti-inflammatory functions, skewing the proinflammatory macrophagic phenotype M1 to the anti-inflammatory M2 and decreasing synthesis of interleukin-2, and it plays a role in activation and proliferation of T lymphocytes. Interestingly, both

rs429358-T and rs7412-T associate with increased inflammation markers such as C-reactive protein but protect against AD and MI. Whether this is due to directly promoting macrophage activation in steatotic tissue such as muscle, pancreas and liver to cause development of T2D, pancreatitis and cirrhosis, respectively, while decreasing it in brain and vessels to protect against AD and MI is intriguing. Other possibilities include that both T alleles promote cholesterol deposition and cellular damage in muscle, pancreas and liver that secondarily leads to inflammation.

Our study has several strengths. First, we performed a genetic association meta-analysis across the largest set of individuals to date for CT-measured liver attenuation in multiple ancestries. We showed consistent associations at five loci, including one novel locus, overall and in sex and ethnicity stratified analyses. Second, we characterize the association of protein-coding genomic variation, which has not been well-studied at the population level, with hepatic steatosis. Prior studies did not report associations with rs429358 because this variant is not included on arrays and was poorly imputed. Third, we explore the pleiotropic effects of *APOE* mutations on a large number of traits through our pheWAS analyses in two biobanks and published GWAS.

Although our study has several strengths, we note some limitations. First, not all protein-coding variation is cataloged on the exome chip. Because of purifying selection, disruptive protein-coding variation is rare and would not be found in our studies (34). Our analysis still demonstrates a paucity of exome-wide associations for hepatic steatosis. Larger studies could identify associations not seen here. Since *APOE* variation is associated with all-cause and cause-specific mortality, those at highest risk may not have been included in the cohorts studied; however there were only minimal changes in allele frequency across age groups as seen in UKBiobank (Supplementary Material, Fig. S4). Finally, our results may not be generalizable to other populations and/or other measures of liver function.

This is the largest genetic analysis of liver attenuation to date. We identified five loci that were exome-wide significant for association with CT-measured liver attenuation, including one new signal at *TOMM40/APOE*. We characterize the allele-specific effects at the *TOMM40/APOE* locus on multiple diseases and traits. We show for the first time that rs429358-T, which promotes liver steatosis, NAFLD, cirrhosis, diabetes and obesity, is associated with a lower risk of AD, MI, and increased serum total cholesterol, LDL cholesterol and triglycerides highlighting the context dependent benefits and harms of genetic variation at this locus. These findings highlight how genetics can identify individuals with a risk profile that differs from the overall population which can help guide future precision health recommendations.

Materials and Methods

Ethics statement

All work was approved by local institutional review boards or equivalent committees and conformed to the ethical guidelines of the 1975 Declaration of Helsinki. All participants provided written informed consent.

Study design for the GOLD Consortium

Eight cohorts ($n = 16\,492$) were included (Supplementary Material, Table S12 and Supplementary Material, Table S13): Age, Gene/Environment Susceptibility-Reykjavik Study (AGES) (35), Family Heart Study (FamHS) (36), Framingham Heart Study (FHS) (37), Genetic Epidemiology Network of Arteriopathy (GENOA) (38), Insulin Resistance Atherosclerosis Family Study (IRASFS) (39), Jackson Heart Study (JHS) (40), Multiethnic Study of Atherosclerosis (MESA) and Old Order Amish (OOA) (41,42). CT scanning with a standardized protocol was used to measure hepatic steatosis (Supplementary Material, Table S13) which is correlated with decreased liver attenuation ($r^2 = 0.92$) (43). Either phantom or spleen density was used to normalize liver attenuation and control for scan penetrance as part of the quality control. For each cohort, age, sex, ancestry and alcohol intake were self-reported (Supplementary Material, Table S1). Genotypes were assayed using the Illumina HumanExome array (Illumina, Inc., San Diego, CA, USA). Details of the genotyping methods and quality control are in Supplementary Material, Table S14 and Supplementary Material, Table S15. Sample exclusions included a call-rate <95%, ethnic outliers in a principal component analysis, evidence of contamination, sex inconsistencies or unexpected cryptic relatedness. Variants were excluded with call rates <98%, minor allele count <6 or deviation from Hardy–Weinberg Equilibrium ($P < 1 \times 10^{-6}$).

Association analyses

In each cohort, liver attenuation was inverse normally transformed. Association analyses were performed using a linear regression framework (linear mixed modeling was used for related individuals) adjusted for age, age^2, sex, alcohol intake and ancestry estimates. Meta-analysis was performed using the inverse variance weighted method as implemented in METAL. For significance, a meta-analysis $P < 5.34 \times 10^{-7}$ was considered as exome-wide significant (Supplementary Material, Table S2). The variation explained by the tested SNPs was estimated using $2f\ (1-f)\ a^2$, where f is the frequency of the variant and a is its additive effect in units of standard deviations from the meta-analysis. In addition, ethnic-specific and sex-specific meta-analyses were performed. Observed heterogeneity was assessed using Cochran's Q and quantified this using the I^2 metric; a Cochran's Q $P < 1 \times 10^{-4}$ was considered significant. We tested for heterogeneity across ethnic groups overall and in men and women separately (Supplementary Material, Table S3, Supplementary Material, Table S4 and Supplementary Material, Table S5). All cohorts except JHS were included in the sex-specific analysis (Supplementary Material, Table S4 and Supplementary Material, Table S5). In addition, we also tested for heterogeneity between sexes. We also carried out the association of selected *TOMM40* and *APOE* SNPs with liver attenuation (Supplementary Material, Table S6).

For exome-wide significant loci, conditional analyses were performed to identify independent NAFLD associations. Genome-wide complex trait analysis (GCTA) (44) was used to perform approximate joint and conditional SNP association analysis using summary statistics for each ancestry separately and LD matrices derived from the Women's Health Initiative (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000200.v10.p3) for African ancestry and University of Michigan Health and Retirement Study (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000428.v1.p1) for European ancestry analyses. Direct conditional analyses were carried out in IRASFS Hispanic Americans as no appropriate LD reference population was available. For the *TOMM40/APOE* locus, effects at rs429358 were additionally evaluated in participants with genotypes for both

rs2075650 and rs429358. A $P < 0.05$ was considered significant for a second signal after conditioning on the exome-wide SNPs.

UKBB

UKBB analysis was performed under Resource Project #18120. Briefly, UKBB is a prospective epidemiological study of phenotyped individuals aged 40–69 at the time of recruitment from the UK which have been genotyped (45). Briefly, the genotypes of the UKBB participants were assayed using either of two genotyping arrays, the Affymetrix UK BiLEVE Axiom or Affymetrix UKBiobank Axiom array. These arrays were augmented by imputation of ~96 million genetic variants from the Haplotype Reference Consortium (http://www.haplotype-reference-consortium.org/), 1000 Genomes (https://www.internationalgenome.org/) and the UK 10 K (https://www.uk10k.org/) projects. Individuals were excluded if they were designated by the UKBB as outliers based on either genotyping missingness rate or heterogeneity, whose sex inferred from the genotypes did not match their self-reported sex and who were not of white British ancestry. Finally, individuals were removed if they had missingness >5% across variants which passed QC procedures. Characteristics of the participants ($N = 408\,961$) are shown in Supplementary Material, Table S16. NAFLD cases were identified by having the following ICD-9 or ICD-10 diagnosis: ICD-9571.8 or ICD-10 K76.0, NAFLD cirrhosis by ICD-9571.5 or ICD-10 K74*, diabetes by ICD-9250* or ICD-10 E11*, AD by ICD-9331.0 or ICD-10 G30.0, G30.1, G30.8, G30.9, cholelithiasis by ICD-9574* or ICD-10 K80*, acute pancreatitis by ICD-9577.0 or ICD-10 K85*, acute pancreatitis specific by ICD-9577.0 or ICD-10 K85.0, K85.1, K85.8, K85.9, ischemic heart disease by ICD-9410-414 or ICD-10 I20-I25, and cerebrovascular disease by ICD-9430-438 or ICD-10 I60-I69. Association analyses were carried out in UKBB using linear mixed modeling controlling for sex, array batch, UKBB Assessment Center, age, age^2, and the first 10 genomic principal components. Imputation quality scores for UKBB (1.00) are shown in Supplementary Material, Table S17. The effects of *TOMM40* rs2075650, *APOE* rs7412 and *APOE* rs429358 on selected phenotypes were examined in UKBB (Supplementary Material, Table S7). A joint analysis was performed in UKBB to obtain inferences about associations for each SNP (rs2075650, rs429358 and rs7412) alone and then for each SNP after conditioning on the other two SNPs (Supplementary Material, Table S8).

MGI

Study design has been previously described (46). Briefly, patients at the University of Michigan Health System were recruited on the day of their elective procedure using an opt-in written informed consent for broad long-term use of their electronic health information and genetic data. Outpatient lab values, diagnoses, demographics and vital signs from individuals seen between January 1, 2012 and December 31, 2016 were included. For lab measures, all available outpatient lab measures were used to calculate a mean and SD for each trait and each individual. Outpatient lab measures for an individual were excluded if they were more than 1 SD from the mean to decrease entry errors. Continuous traits were inverse normally transformed across all individuals measured for a given trait before analysis. ICD-9 (before 2015) and ICD-10 (after 2015) diagnoses for disease traits were used (Supplementary Material, Table S18). Diagnosis codes included: NAFLD (ICD-9571.8, ICD-10 K76.0); NAFLD cirrhosis (ICD-9571.5, ICD-10 K74*); diabetes (ICD-9250*, ICD-10 E11*); AD (ICD-9331.0,ICD-10 G30.0, G30.1, G30.8, G30.9); cholelithiasis (ICD-9574*, ICD-10 K80*); acute pancreatitis (ICD-9577.0, ICD-10 K85*); ischemic heart disease (ICD-9410-414, ICD-10 I20-I25; and cerebrovascular disease (ICD-9430-438, ICD-10 I60-I69). Individuals with a disease diagnosis were used as cases and individuals without this diagnosis but with at least one diagnosis were used as controls. The effects of *TOMM40* rs2075650, *APOE* rs7412 and *APOE* rs429358 on selected phenotypes were examined in MGI (Supplementary Material, Table S9). Genotypes were obtained from the Illumina HumanCoreExome v12.1 array with imputation to the Haplotype Reference Consortium. Imputation quality scores are shown in Supplementary Material, Table S17. Only unrelated individuals of European ancestry as determined by the KING software (v1.4.2) and LASER program (v.2.0.0) were used.

Evaluation of effects on other metabolic traits and selected diseases

We evaluated the effects of SNPs on metabolic traits and selected diseases using publically available GWAS (Supplementary Material, Table S10). To evaluate the effects on metabolic traits, association results for HDL cholesterol, total cholesterol, LDL cholesterol, and triglycerides from Global Lipids Genetics consortium (18), fasting glucose and fasting insulin from MAGIC (47), and leptin (48), waist-to-hip ratio controlled for body mass index (WHRcontBMI) and BMI (15), and body fat percentage (49) from recent GWAS were examined. MI associations were obtained from CARDIoGRAMplusC4D Consortium (50), T2D associations were obtained from DIAGRAM (51), and AD associations from meta-analysis (19). We performed a lookup of the SNPs to evaluate the effects on serum metabolites (23)(Supplementary Material, Table S11).

Expression quantitative trait loci

Genotype-Tissue Expression v7 was used to obtain expression quantitative trait loci (eQTL) data as described previously (10).

Supplementary Material

Supplementary material is available at *HMG* online

Conflict of Interest Statement. None declared.

Funding

Age, Gene/Environment Susceptibility-Reykjavik Study (AGES) was funded by National Institutes of Health contract N01-AG-1-2100 and HHSN271201200022C, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the Icelandic National Bioethics Committee, VSN: 00-063. Genotyping was done at the Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas. Support for Family Heart Study (FamHS) was provided by the National Heart, Lung and Blood Institute (NHLBI) grant R01 HL08770003 and R01 HL117078 and National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grant R01 DK089256. The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with Boston University (Contract No. N01-HC-25195 and HHSN268201500001I). Funding for SHARe Affymetrix genotyping was provided by National Heart, Lung, and Blood Institute Contract N02-HL64278. SHARe Illumina genotyping was provided under an agreement between

Illumina and Boston University. Funding for Affymetrix genotyping of the FHS Omni cohorts was provided by Intramural NHLBI funds from Andrew D. Johnson and Christopher J. O'Donnell. Support for the Genetic Epidemiology Network of Arteriopathy (GENOA) study was provided by the National Institutes of Health, grant numbers HL085571, HL087660 and HL100245 from National Heart, Lung, and Blood Institute. Support for the Insulin Resistance Atherosclerosis Family Study (IRASFS) was provided by the National Heart, Lung, and Blood Institute grants R01 HL060944, R01 HL061019, R01 HL060919, R01 HL060894 and R01 HL061210 and NIDDK grant DK085175 and DK118062. IRASFS genotyping was carried out with funds from the Department of Internal Medicine at the University of Michigan. Analysis was partially supported by the Mid-Atlantic Nutrition Obesity Research Center (P30 DK072488) from the NIDDK. The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I and HHSN268201800012I) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute on Minority Health and Health Disparities (NIMHD). The authors also wish to thank the staffs and participants of the JHS. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services. MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420. Liver phenotyping is MESA was made possible by R01 HL088451. MESA is also supported, in part, by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the NIDDK Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The Old Order Amish (OOA) studies are supported by grants and contracts from National Institutes of Health, including U01 HL072515, U01 HL84756, U01 HL137181 and P30 DK72488. The authors acknowledge the Michigan Genomics Initiative participants, Precision Health at the University of Michigan, the University of Michigan Medical School Central Biorepository, and the University of Michigan Advanced Genomics Core for providing data and specimen storage, management, processing, and distribution services, and the Center for Statistical Genetics in the Department of Biostatistics at the School of Public Health for genotype data curation, imputation, and management in support of the research reported in this publication. Analyses in the UKBB were done under approved project 18120 (EKS). The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. This manuscript was not prepared in collaboration with investigators of the WHI, has not been reviewed and/or approved by the Women''s Health Initiative (WHI), and does not necessarily reflect the opinions of the WHI investigators or the NHLBI. This study makes use of data from dbGaP HRS (dbGaP accession: phs000428.v1.p1): HRS is supported by the National Institute on Aging (NIA U01AG009740). The genotyping was funded separately by the National Institute on Aging (RC2 AG036495, RC4 AG039029). Genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and final preparation of the data were performed by the Genetics Coordinating Center at the University of Washington. In addition, we would like to acknowledge the American Diabetes Association Mentor-Based Postdoctoral Fellowship Program (7-07-MN-08, RH). National Institutes of Health (grants R01 DK106621 to EKS, BK, YC, AK, XD and BH; RO1 DK107904 to EKS) and The University of Michigan Department of Internal Medicine.

References

1. Anstee, Q.M., Targher, G. and Day, C.P. (2013) Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. *Nat. Rev. Gastroenterol. Hepatol.*, **10**, 330–344.
2. Speliotes, E.K., Massaro, J.M., Hoffmann, U., Vasan, R.S., Meigs, J.B., Sahani, D.V., Hirschhorn, J.N., O'Donnell, C.J. and Fox, C.S. (2010) Fatty liver is associated with dyslipidemia and dysglycemia independent of visceral fat: the Framingham Heart Study. *Hepatology*, **51**, 1979–1987.
3. Stepanova, M. and Younossi, Z.M. (2012) Independent association between nonalcoholic fatty liver disease and cardiovascular disease in the US population. *Clin. Gastroenterol. Hepatol.*, **10**, 646–650.
4. McCarthy, M.I. (2010) Genomics, type 2 diabetes, and obesity. *N. Engl. J. Med.*, **363**, 2339–2350.
5. Emerging Risk Factors, C., Wormser, D., Kaptoge, S., Di Angelantonio, E., Wood, A.M., Pennells, L., Thompson, A., Sarwar, N., Kizer, J.R., Lawlor, D.A. *et al.* (2011) Separate and combined associations of body-mass index and abdominal adiposity with cardiovascular disease: collaborative analysis of 58 prospective studies. *Lancet*, **377**, 1085–1095.
6. Craft, S. (2009) The role of metabolic disorders in Alzheimer disease and vascular dementia: two roads converged. *Arch. Neurol.*, **66**, 300–305.
7. 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. *Hepatology*, **58**, 966–975.
8. Romeo, S., Kozlitina, J., Xing, C., Pertsemlidis, A., Cox, D., Pennacchio, L.A., Boerwinkle, E., Cohen, J.C. and Hobbs, H.H. (2008) Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.*, **40**, 1461–1465.
9. Chalasani, N., Guo, X., Loomba, R., Goodarzi, M.O., Haritunians, T., Kwon, S., Cui, J., Taylor, K.D., Wilson, L., Cummings, O.W. *et al.* (2010) Genome-wide association study identifies variants associated with histologic features of nonalcoholic fatty liver disease. *Gastroenterology*, **139**, 1567–1576 1576 e1561-1566.
10. 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.

11. Kozlitina, J., Smagris, E., Stender, S., Nordestgaard, B.G., Zhou, H.H., Tybjaerg-Hansen, A., Vogt, T.F., Hobbs, H.H. and Cohen, J.C. (2014) Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.*, **46**, 352–356.
12. Kahali, B., Halligan, B. and Speliotes, E.K. (2015) Insights from genome-wide association analyses of nonalcoholic fatty liver disease. *Semin. Liver Dis.*, **35**, 375–391.
13. Li, Q., Qu, H.Q., Rentfro, A.R., Grove, M.L., Mirza, S., Lu, Y., Hanis, C.L., Fallon, M.B., Boerwinkle, E., Fisher-Hoch, S.P. *et al.* (2012) PNPLA3 polymorphisms and liver aminotransferase levels in a Mexican American population. *Clin. Invest. Med.*, **35**, E237–E245.
14. Yuan, X., Waterworth, D., Perry, J.R., Lim, N., Song, K., Chambers, J.C., Zhang, W., Vollenweider, P., Stirnadel, H., Johnson, T. *et al.* (2008) Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. *Am. J. Hum. Genet.*, **83**, 520–528.
15. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J. *et al.* (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature*, **518**, 197–206.
16. Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Magi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E. *et al.* (2015) New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, **518**, 187–196.
17. Reiner, A.P., Barber, M.J., Guan, Y., Ridker, P.M., Lange, L.A., Chasman, D.I., Walston, J.D., Cooper, G.M., Jenny, N.S., Rieder, M.J. *et al.* (2008) Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *Am. J. Hum. Genet.*, **82**, 1193–1201.
18. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S. *et al.* (2013) Discovery and refinement of loci associated with lipid levels. *Nat. Genet.*, **45**, 1274–1283.
19. Kunkle, B.W., Grenier-Boley, B., Sims, R., Bis, J.C., Damotte, V., Naj, A.C., Boland, A., Vronskaya, M., van der Lee, S.J., Amlie-Wolf, A. *et al.* (2019) Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat. Genet.*, **51**, 414–430.
20. De Feo, E., Cefalo, C., Arzani, D., Amore, R., Landolfi, R., Grieco, A., Ricciardi, W., Miele, L. and Boccia, S. (2012) A case-control study on the effects of the apolipoprotein E genotypes in nonalcoholic fatty liver disease. *Mol. Biol. Rep.*, **39**, 7381–7388.
21. Sazci, A., Akpinar, G., Aygun, C., Ergul, E., Senturk, O. and Hulagu, S. (2008) Association of Apolipoprotein E Polymorphisms in patients with non-alcoholic steatohepatitis. *Dig. Dis. Sci.*, **53**, 3218–3224.
22. Yang, M.H., Hee, J.S., Sung, J.D., Choi, Y.H., Koh, K.C., Yoo, B.C. and Paik, S.W. (2005) The relationship between apolipoprotein E polymorphism, lipoprotein (a) and fatty liver disease. *Hepato-Gastroenterology*, **52**, 4.
23. Kettunen, J., Demirkan, A., Wurtz, P., Draisma, H.H., Haller, T., Rawal, R., Vaarhorst, A., Kangas, A.J., Lyytikainen, L.P., Pirinen, M. *et al.* (2016) Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat. Commun.*, **7**, 11122.
24. Kim, D.G., Krenz, A., Toussaint, L.E., Maurer, K.J., Robinson, S.A., Yan, A., Torres, L. and Bynoe, M.S. (2016) Non-alcoholic fatty liver disease induces signs of Alzheimer's disease (AD) in wild-type mice and accelerates pathological signs of AD in an AD model. *J. Neuroinflammation*, **13**, 1.
25. Seo, S.W., Gottesman, R.F., Clark, J.M., Hernaez, R., Chang, Y., Kim, C., Ha, K.H., Guallar, E. and Lazo, M. (2016) Nonalcoholic fatty liver disease is associated with cognitive function in adults. *Neurology*, **86**, 1136–1142.
26. Weinstein, G., Zelber-Sagi, S., Preis, S.R., Beiser, A.S., DeCarli, C., Speliotes, E.K., Satizabal, C.L., Vasan, R.S. and Seshadri, S. (2018) Association of Nonalcoholic Fatty Liver Disease with lower brain volume in healthy middle-aged adults in the Framingham study. *JAMA Neurol.*, **75**, 97–104.
27. Nho, K., Kueider-Paisley, A., Ahmad, S., Mahmoudian Dehkordi, S., Arnold, M., Risacher, S.L., Louie, G., Blach, C., Baillie, R., Han, X. *et al.* (2019) Association of Altered Liver Enzymes with Alzheimer disease diagnosis, cognition, neuroimaging measures, and cerebrospinal fluid biomarkers. *JAMA Netw. Open*, **e197978**, 2.
28. Barata, L., Feitosa, M.F., Bielak, L.F., Halligan, B., Baldridge, A.S., Guo, X., Yerges-Armstrong, L.M., Smith, A.V., Yao, J., Palmer, N.D. *et al.* (2019) Insulin resistance exacerbates genetic predisposition to nonalcoholic fatty liver disease in individuals without diabetes. *Hepatol Commun*, **3**, 894–907.
29. Salameh, T.S., Rhea, E.M., Banks, W.A. and Hanson, A.J. (2016) Insulin resistance, dyslipidemia, and apolipoprotein E interactions as mechanisms in cognitive impairment and Alzheimer's disease. *Exp. Biol. Med. (Maywood)*, **241**, 1676–1683.
30. de la Monte, S.M. (2017) Insulin resistance and neurodegeneration: progress towards the development of new therapeutics for Alzheimer's disease. *Drugs*, **77**, 47–65.
31. Ridge, P.G., Hoyt, K.B., Boehme, K., Mukherjee, S., Crane, P.K., Haines, J.L., Mayeux, R., Farrer, L.A., Pericak-Vance, M.A., Schellenberg, G.D. *et al.* (2016) Assessment of the genetic variance of late-onset Alzheimer's disease. *Neurobiol. Aging*, **41**, 200 e213-200 e220.
32. Ashford, J.W. (2004) <I>APOE</I> genotype effects on Alzheimer's disease onset and epidemiology. *J. Mol. Neurosci.*, **23**, 157–166.
33. Jones, N.S., Watson, K.Q. and Rebeck, G.W. (2019) Metabolic disturbances of a high-fat diet are dependent on APOE genotype and sex. *eNeuro*, **6**.
34. Mac Arthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B. *et al.* (2012) A systematic survey of loss-of-function variants in human protein-coding genes. *Science*, **335**, 823–828.
35. Harris, T.B., Launer, L.J., Eiriksdottir, G., Kjartansson, O., Jonsson, P.V., Sigurdsson, G., Thorgeirsson, G., Aspelund, T., Garcia, M.E., Cotch, M.F. *et al.* (2007) Age, gene/environment susceptibility-Reykjavik study: multidisciplinary applied phenomics. *Am. J. Epidemiol.*, **165**, 1076–1087.
36. Higgins, M., Province, M., Heiss, G., Eckfeldt, J., Ellison, R.C., Folsom, A.R., Rao, D.C., Sprafka, J.M. and Williams, R. (1996) NHLBI family heart study: objectives and design. *Am. J. Epidemiol.*, **143**, 1219–1228.
37. Mahmood, S.S., Levy, D., Vasan, R.S. and Wang, T.J. (2014) The Framingham heart study and the epidemiology of cardiovascular disease: a historical perspective. *Lancet*, **383**, 999–1008.
38. Investigators, T.F. (2002) Multi-center genetic study of hypertension: the family blood pressure program (FBPP). *Hypertension*, **39**, 3–9.
39. Henkin, L., Bergman, R.N., Bowden, D.W., Ellsworth, D.L., Haffner, S.M., Langefeld, C.D., Mitchell, B.D., Norris, J.M., Rewers, M., Saad, M.F. *et al.* (2003) Genetic epidemiology of insulin

resistance and visceral adiposity. The IRAS Family Study design and methods. *Ann. Epidemiol.*, **13**, 211–217.
40. Fuqua, S.R., Wyatt, S.B., Andrew, M.E., Sarpong, D.F., Henderson, F.R., Cunningham, M.F. and Taylor, H.A., Jr. (2005) Recruiting African-American research participation in the Jackson Heart Study: methods, response rates, and sample description. *Ethn. Dis.*, **15**, S6-18-29.
41. Rampersaud, E., Bielak, L.F., Parsa, A., Shen, H., Post, W., Ryan, K.A., Donnelly, P., Rumberger, J.A., Sheedy, P.F., 2nd, Peyser, P.A. *et al.* (2008) The association of coronary artery calcification and carotid artery intima-media thickness with distinct, traditional coronary artery disease risk factors in asymptomatic adults. *Am. J. Epidemiol.*, **168**, 1016–1023.
42. Sorkin, J., Post, W., Pollin, T.I., O'Connell, J.R., Mitchell, B.D. and Shuldiner, A.R. (2005) Exploring the genetics of longevity in the Old Order Amish. *Mech. Ageing Dev.*, **126**, 347–350.
43. Iwasaki, M., Takada, Y., Hayashi, M., Minamiguchi, S., Haga, H., Maetani, Y., Fujii, K., Kiuchi, T. and Tanaka, K. (2004) Noninvasive evaluation of graft steatosis in living donor liver transplantation. *Transplantation*, **78**, 1501–1505.
44. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Genetic Investigation of, A.T.C, Replication, D.I.G., Meta-analysis, C., Madden, P.A., Heath, A.C., Martin, N.G. *et al.* (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.*, **44**, 369–375 S361-363.
45. Canela-Xandri, O., Rawlik, K. and Tenesa, A. (2017) In Bioarchives. In , in press.
46. Maguire, L.H., Handelman, S.K., Du, X., Chen, Y., Pers, T.H. and Speliotes, E.K. (2018) Genome-wide association analyses identify 39 new susceptibility loci for diverticular disease. *Nat. Genet.*, **50**, 1359–1365.
47. Manning, A.K., Hivert, M.F., Scott, R.A., Grimsby, J.L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.T., Bielak, L.F., Prokopenko, I. *et al.* (2012) A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.*, **44**, 659–669.
48. Kilpelainen, T.O., Carli, J.F., Skowronski, A.A., Sun, Q., Kriebel, J., Feitosa, M.F., Hedman, A.K., Drong, A.W., Hayes, J.E., Zhao, J. *et al.* (2016) Genome-wide meta-analysis uncovers novel loci influencing circulating leptin levels. *Nat. Commun.*, **7**, 10494.
49. Lu, Y., Day, F.R., Gustafsson, S., Buchkovich, M.L., Na, J., Bataille, V., Cousminer, D.L., Dastani, Z., Drong, A.W., Esko, T. *et al.* (2016) New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat. Commun.*, **7**, 10495.
50. Nikpay, M., Goel, A., Won, H.H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., Hopewell, J.C. *et al.* (2015) A comprehensive 1,000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.*, **47**, 1121–1130.
51. Mahajan, A., Go, M.J., Zhang, W., Below, J.E., Gaulton, K.J., Ferreira, T., Horikoshi, M., Johnson, A.D., Ng, M.C.Y., Prokopenko, I. *et al.* (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.*, **46**, 234–244.