ARTICLES
https://doi.org/10.1038/s41588-019-0378-y
nature genetics

Multi-ancestry genome-wide gene–smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids

The concentrations of high- and low-density-lipoprotein cholesterol and triglycerides are influenced by smoking, but it is unknown whether genetic associations with lipids may be modified by smoking. We conducted a multi-ancestry genome-wide gene–smoking interaction study in 133,805 individuals with follow-up in an additional 253,467 individuals. Combined meta-analyses identified 13 new loci associated with lipids, some of which were detected only because association differed by smoking status. Additionally, we demonstrate the importance of including diverse populations, particularly in studies of interactions with lifestyle factors, where genomic and lifestyle differences by ancestry may contribute to novel findings.

Levels of serum lipids, such as triglycerides and high- and low-density-lipoprotein cholesterol (HDL and LDL), are influenced by both genetic and lifestyle factors. Over 250 lipid-associated loci have been identified[1–6], yet it is unclear to what extent lifestyle factors modify the effects of these variants or those of variants yet to be identified. Smoking is associated with an unfavorable lipid profile[7,8], warranting its investigation as a lifestyle factor that potentially modifies genetic associations with lipids. Identifying interactions through traditional 1-degree-of-freedom (1df) tests of SNP × smoking terms may have low power, except in very large sample sizes. To enhance power, a 2-degree-of-freedom (2df) test that jointly evaluates interaction and main effects was developed[9].

The Gene–Lifestyle Interactions Working Group, under the aegis of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium[10], was formed to conduct analyses of lifestyle interactions in the genetic basis of cardiovascular traits. As both genetic and lifestyle factors differ across populations of different ancestry, and to address the under-representation of non-European populations in genomic research, great effort went into creating a large multi-ancestry resource for these investigations[11]. Here we report a genome-wide interaction study that uses both the 1df test of interaction and the 2df joint test of main and interaction effects to examine the hypothesis that genetic associations with serum lipids differ by smoking status.

Results

New loci. We conducted genome-wide interaction meta-analyses for current and ever-smoking status in up to 133,805 individuals of European (EUR), African (AFR), Asian (ASN), and Hispanic (HISP) ancestry (stage 1; Supplementary Tables 1–3), with follow-up of 17,921 variants associated at $P \leq 1 \times 10^{-6}$ (not pruned for linkage disequilibrium, LD) in an additional 253,467 individuals of EUR, AFR, ASN, HISP, and Brazilian (BR) ancestry (stage 2; Supplementary Tables 4–6), as detailed in Fig. 1. Of the 17,921 variants associated in stage 1, 16,389 (in 487 loci, defined as the region located ±1 Mb with respect to the variant) passed filters and were included in stage 2 analyses. Ninety percent of variants (14,733) and 22% of loci (109) replicated in stage 2 (variants, $P < 0.05/16{,}389$; loci, $P < 0.05/487$). We conducted meta-analyses of stage 1 and 2 results (Manhattan plots, Supplementary Fig. 1; quantile–quantile plots, Supplementary Fig. 2) and identified 13 new loci associated at $P < 5 \times 10^{-8}$ that were at least 1 Mb away from previously reported lipid-associated loci (Table 1; results by stage, Supplementary Table 7; forest plots, Supplementary Figs. 3 and 4; regional association plots, Supplementary Fig. 5). These loci had low false-discovery rate (FDR) q values (all $q < 3 \times 10^{-4}$; Supplementary Table 8). We report the new loci associated at $P < 5 \times 10^{-8}$ as well as those among these passing a more stringent significance threshold ($P < 6.25 \times 10^{-9}$), adjusted for two smoking exposures, two interaction tests, and ancestry-specific and trans-ancestry tests. The patterns observed in these results are described below and illustrated with output from stage 1 meta-analyses, where results from a main-effect model (in all individuals and with stratification by smoking exposure) and a smoking-adjusted main-effect model were also available (Fig. 1 and Supplementary Table 9).

Notably, many of the new loci were statistically significant only in AFR meta-analyses. For 7 of the 13 new loci, the minor allele frequency (MAF) of the index variant was highest in AFR populations, and inter-ancestry differences in MAF and/or LD may explain the inability to detect similar associations in the other ancestry groups. However, some AFR-only associations were unlikely to be due to diminished power in non-AFR meta-analyses. For instance, the effect of rs12740061 (NC_000001.10:g.694078 10C>T; *LOC105378783*) on HDL was significantly modified by current smoking status among AFR individuals ($P_{1df} = 7.4 \times 10^{-9}$; Fig. 2 and Table 1), such that the genetic effect was stronger among current smokers than among nonsmokers (Supplementary Table 9). In contrast, there was virtually no evidence for association in any other ancestry group, despite these groups having higher MAF values for the variant (Fig. 2). The potential influence of under-adjustment for principal components on these results was evaluated by excluding the six studies that adjusted for only 1 principal component (the average number of principal components adjusted for among AFR studies was 4.2); in this analysis, effect estimates were similar and P values were increased or similar in comparison to the original analysis, in line with the ~20% reduction in sample size (Supplementary Table 10).

We observed interactions where notable associations were only found among current or ever-smokers, with effect sizes close to zero among non- or never-smokers, including a statistically significant association in the 2df joint test of main and interaction effects of rs7364132 (NC_000022.10:g.20096172G>A; *DGCR8*) × ever smoking with triglycerides ($P_{2df} = 2.5 \times 10^{-8}$; Table 1). Main-effect models stratified by smoking status showed a strong

A full list of authors and affiliations appears at the end of the paper.

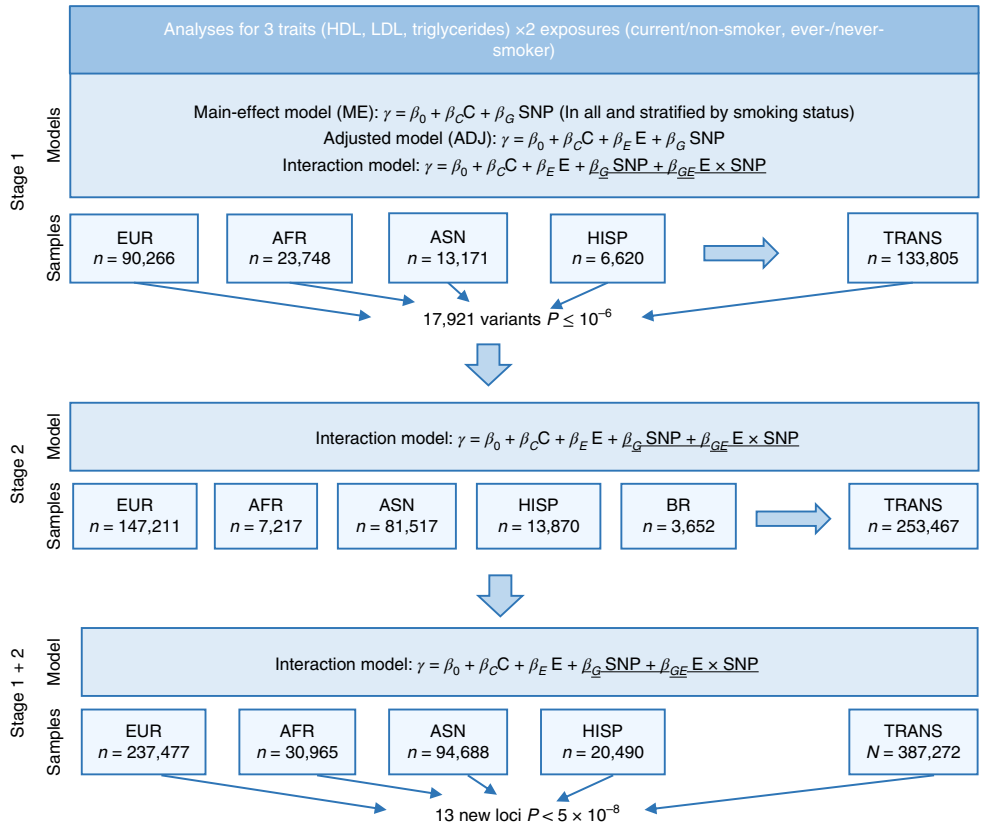

Fig. 1 | Study overview. Summary of data included in this study. Of the 17,921 associated variants from stage 1, 16,389 passed filtering criteria and were included in stage 2 analyses. Trans-ancestry combined stage 1 and 2 meta-analyses were performed on stage 1 trans-ancestry and stage 2 trans-ancestry meta-analyses and not on combined ancestry-specific analyses from stage 1 and stage 2. In models, 1df terms are in bold and 2df terms are underlined. TRANS, trans-ancestry. Model descriptions include terms for the outcome (γ), intercept (β_0), covariates (β_CC), the variant (β_GSNP), smoking status (β_EE), and interaction of the variant and smoking status (β_{GE}E × SNP).

genetic association with triglycerides among ever-smokers (difference in mean ln(triglycerides) per A allele (β) = −0.05, $P = 7.9 \times 10^{-8}$), with a negligible association among never-smokers ($\beta = 0.01$, $P = 0.19$; Fig. 3). This association was not significant in the non-stratified main-effect model (Table 1 and Supplementary Table 9) and was only detectable when modeling permitted different associations across smoking strata. Similar results were observed for rs79950627 (NC_000011.9:g.2233790G>A; *MIR4686*) × current smoking with LDL and rs56167574 (NC_000007.13:g.151245975G>A; *PRKAG2*) × ever smoking with LDL (Fig. 3 and Supplementary Table 9).

We also observed interactions where effects were in opposite directions in the exposed and unexposed strata, with a larger effect and more statistically significant association among smokers. For instance, current smoking status modified the association between rs73453125 (NC_000007.13:g.146084573G>A; *CNTNAP2*) and LDL (Table 1). In stratified main-effect models, the A allele was associated with lower LDL among current smokers ($\beta = -8.1$ mg/dl, $P = 2.2 \times 10^{-7}$) but was associated with higher LDL among nonsmokers ($\beta = 2.18$ mg/dl, $P = 0.01$; Fig. 4a and Supplementary Table 9). In a non-stratified smoking-adjusted main-effect model, no association between rs73453125 and LDL was detected ($\beta = 0.3$ mg/dl, $P = 0.98$). Similar results were observed for rs12740061 (*LOC105378783*) (Supplementary Table 9).

Although many interactions manifested as associations that were only significant or were stronger in smokers, for rs10937241 (NC_000003.11:g.185822774A>G; *ETV5*), rs34311866 (NC_000004.11:g.951947T>C; *TMEM175*), rs10101067 (NC_000008.10:g.72407374G>C; *EYA1*), and rs77810251 (NC_000007.13:g.121504149G>A; *PTPRZ1*), the associations observed among non- or never-smokers were more statistically significant. Notably, in stratified main-effect models, rs77810251 was associated with increased HDL among never-smokers ($\beta = 0.05$ ln(HDL), $P = 6.3 \times 10^{-11}$) with no significant association among ever-smokers ($\beta = -0.005$ ln(HDL), $P = 0.56$; Fig. 3 and Supplementary Table 9). In a smoking-adjusted main-effect model of never- and ever-smokers together, the association was markedly reduced ($\beta = 0.02$ ln(HDL), $P = 1.6 \times 10^{-4}$).

The 2df joint test simultaneously evaluates main effects and smoking interaction effects; some of our results seem to capture a main effect of the variant. For instance, the 2df test for rs12144063 (*EYA3*) detected an association ($P = 1.3 \times 10^{-10}$), whereas the 1df test of interaction did not ($P = 0.75$). The minor alleles for this and three other variants (rs10937241 (*ETV5*), rs34311866 (*TMEM175*), and rs10101067 (*EYA1*)) were common across populations and reached genome-wide statistical significance despite effects being small in magnitude (rs10101067 (*EYA1*); Fig. 4b), in agreement with expectations for new main-effect loci in well-studied populations. There were two findings, however, for which the relatively large sample size in the AFR meta-analyses seemed to facilitate detection. For rs73729083 (NC_000007.13:g.137559799T>C; *CREB3L2*), the MAF was much greater in AFR than in HISP or ASN populations (not present in EUR populations) and variant effect estimates were large and consistent across ancestry groups, whereas interaction effect estimates were inconsistent, with wide confidence intervals (Supplementary Fig. 3f). At rs4758675 (NC_000012.11:g.122691738C>A; *B3GNT4*), the minor allele was only present in AFR populations (Supplementary Fig. 3k), but variant effect estimates were consistent across AFR studies, with interaction effect estimates

approaching the null (Supplementary Fig. 4e). In total, 6 of the 13 new loci that we identified seem to be driven by main effects of the variant while the remainder show some evidence of interaction with smoking.

There were 16 additional new loci identified in stage 1 meta-analyses (P_{1df} or $P_{2df} < 5 \times 10^{-8}$) for which the variants were unavailable for analysis in stage 2 cohorts. These loci were identified only in AFR meta-analyses (many were AFR-specific variants; Table 2). Because of the relatively small number and size of the available AFR cohorts in stage 2 (total $n = 7{,}217$ individuals; $n < 2{,}000$ per cohort), these relatively low-frequency variants did not pass filters for minor allele count within exposure groups. Nevertheless, associations for these variants had low FDR q values (all $q < 2.4 \times 10^{-4}$) in stage 1, and some seem worthy of further investigation. One particularly interesting example is the association of rs17150980 (NC_000007.13:g.78173734T>C; *MAGI2*)×ever smoking with triglycerides ($P_{2df} = 1.4 \times 10^{-9}$), in which consistent effects were observed for both the variant and the interaction across AFR studies but not in other ancestry groups (Supplementary Fig. 6).

As we ran analyses for both current and ever-smoking status, we evaluated new associations across smoking exposures to further characterize these loci (Supplementary Table 11). For the six probable main-effect loci (*EYA3*, *ETV5*, *TMEM175*, *CREB3L2*, *EYA1*, and *B3GNT4*), an association of similar statistical significance was observed across smoking status definitions for the 2df joint test with a similar lack of effect for the 1df test of interaction, in agreement with the interpretation that smoking status was unimportant and only the main effect drove association. For the locus in which a stronger association was observed among nonsmokers (*PTPRZ1*), the 1df interaction P value was dramatically reduced from 9.5×10^{-7} for ever smoking to 0.011 for current smoking, in line with any smoke exposure altering the association between this variant and HDL and the notion that including former smokers with never-smokers (as in the analysis of current smoking) dilutes the observed association among never-smokers. For the reported interactions with current smoking, all effect estimates were greatly reduced in the ever-smoking analysis, suggesting that active smoking is the relevant exposure. For the reported interactions with ever smoking, markedly reduced statistical significance was observed in the analysis of current smoking, likely reflecting a drop in power from excluding former smokers from the exposed group.

We conducted a secondary analysis of smoking dose in two of our AFR cohorts with measured cigarettes per day for four interaction loci (see the Methods for selection criteria): rs12740061 (*LOC105378783*), rs73453125 (*CNTNAP2*), rs79950627 (*MIR4686*), and rs7364132 (*DGCR8*). For each of these variants, a stronger association was observed with increasing smoking dose (Supplementary Table 12), and the interaction was statistically significant for all variants but rs7364132, for which the P value was just over our threshold for statistical significance ($P = 0.0035$ versus $P < 0.0021$).

Conditional analysis showed no evidence that the new associations were driven by variants at known lipid-associated loci (Supplementary Table 13). Imputation quality for the new variants was high (minimum of 0.75), with sample-size-weighted average imputation quality of 0.90, and MAFs match those in publicly available datasets (Supplementary Table 14).

Interactions at known loci. We examined interactions with smoking at known lipid-associated loci. Because results for the 2df test at known loci are expected to predominantly reflect previously identified main effects, we exclusively evaluated results from the 1df test of interaction. No interactions within known loci were statistically significant ($P_{1df} < 0.05/269$ known loci in our data). To evaluate whether the proportion of known variants with $P_{1df} < 0.05$ was higher than would be expected by chance (5%), we conducted binomial tests for each trait–exposure combination (P values were Bonferroni corrected for multiple tests). There was significant enrichment for known variants with interaction in the 1df test reaching $P < 0.05$, including for the HDL–current smoking ($P = 9.6 \times 10^{-12}$), HDL–ever smoking ($P = 5.9 \times 10^{-7}$), LDL–current smoking ($P = 8.4 \times 10^{-15}$), LDL–ever smoking ($P = 3.1 \times 10^{-5}$), triglycerides–current smoking ($P = 4.0 \times 10^{-3}$), and triglycerides–ever smoking ($P = 3.1 \times 10^{-4}$) combinations. We conducted power calculations under different interaction scenarios to determine the conditions under which an interaction analysis and a main-effect analysis would both be sufficiently powered to detect the same locus (that is, when an interaction could be detected in a locus previously identified in a main-effect analysis; Supplementary Table 15). At current trans-ancestry meta-analysis sample sizes and when assuming a large effect size, there was limited power to detect either a main effect or an interaction when an association was of larger effect or only present among smokers (main effect, <1%; interaction, 77%) or when associations differed in magnitude but not direction (main effect, >99%; interaction, <1%), thus making it unlikely that an interaction at a known locus would be detected. We were well powered for both interaction and main-effect analyses to detect smoking interactions in which smoking eliminated or drastically reduced an association with a large effect size among non- or never-smokers. We identified one such interaction in our data, for *PTPRZ1* in AFR studies only, which may not have previously been identified in a main-effect analysis because of the limited power of AFR main-effect analyses thus far.

Proportion of variance explained by the identified loci. Ten studies from four ancestry groups were used to calculate the proportion of the variance in lipid traits explained by the new genome-wide-significant loci, including 13 loci from combined stage 1 and 2 meta-analyses (Table 1) and 16 loci from stage 1 that were not available in stage 2 analyses (Table 2). Two different methods were used (Methods), and the range of findings across these methods is presented (Supplementary Table 16). In the AFR ancestry group, the new variants and their interactions explained 1.0–2.7% of variance in HDL, 0.7–2.6% of variance in LDL, and 1.3–3.2% of variance in triglycerides. The proportion explained was smaller among EUR (0.06–0.14% for HDL, 0.01–0.07% for LDL, and 0.10–0.19% for triglycerides), ASN (0.27–0.86% for HDL, 0.09–0.82% for LDL, and 0.8–1.5% for triglycerides), and HISP (0.2–0.4% for HDL, 0.2–0.5% for LDL, and 0.2–0.4% for triglycerides) ancestry groups. These results should be considered in the context of the differences in MAF between the ancestry groups: the proportion of new variants that could be evaluated varied by ancestry group, with 94–97% of variants available for analysis in the AFR cohorts, but only 32–39% of variants available in the EUR and ASN cohorts and 55% of variants available in the HISP cohort. In contrast, each of the cohorts investigated had a similar proportion of the known variants considered (83–96%).

Reproducing known lipid associations. We evaluated the degree to which our data reproduce previously reported lipid-associated loci. Given that approximately 81% of the cohorts in stage 1 were also included in previous efforts, this analysis is not a formal replication. For comparability with traditional genome-wide association studies (GWAS), we evaluated results from stage 1 main-effect models. Of the 356 previously reported associations for 279 variants (compiled from refs. [1–6,12]), there were 236 associations for 189 variants that were confirmed in our data (with consistent direction of effect and $P < 0.05/356$), for a 66.3% concordance rate (Supplementary Table 17).

Bioinformatics. To characterize the potential impact of our new associations on chronic disease risk and to investigate biological mechanisms, we conducted a series of follow-up analyses and

Table 1 | Statistically significant ($P < 5 \times 10^{-8}$) results in stage 1 and 2 meta-analyses

Index variant (nearest gene)[a]	Build 37 chr:position	1000 Genomes freq.[b] AFR/AMR/ASN/EUR	Tested allele: freq.	Ancestry	Trait/ exposure[c]	Stages 1 + 2							Stage 1	
						n	Effect	SE	Int. effect	SE	1df int. *P* value	2df joint *P* value	*n*	Adj. main-effect *P* value[d]
Loci with evidence for interaction														
rs12740061 (*LOC105378783*)	1:69,407,810	0.01/0.17/0.02/0.22	T: 0.05	AFR	HDL/CS	16,606	0.02	0.0082	−0.11	0.019	**7.40×10^{-9}**	**2.4×10^{-8}**	15,499	0.98
rs77810251 (*PTPRZ1*)	7:121,504,149	0.02/0.22/0.34/0.11	A: 0.04	AFR	HDL/ES	24,253	0.052	0.0083	−0.06	0.012	9.50×10^{-7}	**1.2×10^{-9}** [e]	23,146	1.60×10^{-4}
rs73453125 (*CNTNAP2*)	7:146,084,573	0.09/0.02/0/0	A: 0.07	TRANS, AFR	LDL/CS	40,566	1.9	0.69	−8.3	1.4	1.70×10^{-7}	**2.0×10^{-8}**	24,668	0.76
rs56167574 (*PRKAG2*)	7:151,245,975	0.13/0.01/0/0	A: 0.12	AFR	LDL/ES	25,778	1.9	0.8	−6.1	1.1	**1.50×10^{-8}**	8.4×10^{-8}	23,353	0.08
rs79950627 (*MIR4686*)	11:2,233,790	0.06/0.01/0/0	A: 0.05	TRANS, AFR	LDL/CS	38,272	−0.1	0.79	−8.4	1.6	1.40×10^{-6}	**7.2×10^{-9}**	23,348	0.25
rs60029395 (*ZNF729*)	19:22,446,748	0.15/0.01/0.03/0	A: 0.13	AFR	TRIG/CS	19,048	0.041	0.0092	−0.097	0.018	**3.30×10^{-8}**	8.2×10^{-8}	15,747	0.17
rs7364132 (*DGCR8*)	22:20,096,172	0.19/0.02/0/0	A: 0.16	AFR, TRANS	TRIG/ES	23,935	0.012	0.0091	−0.066	0.013	8.80×10^{-7}	**2.5×10^{-8}**	21,834	0.0055
Probable main-effect loci (no evidence for interaction)														
rs12144063 (*EYA3*)	1:28,406,047	0.35/0.28/0.53/0.30	T: 0.37	TRANS	HDL/CS, HDL/ES	375,418	−0.004	0.00069	−0.00033	0.0016	0.75	**1.3×10^{-10}** [e]	131,057	4.70×10^{-7}
rs10937241 (*ETV5*)	3:185,822,774	0.30/0.31/0.58/0.19	A: 0.17	EUR, TRANS	HDL/CS, HDL/ES	230,919	−0.008	0.0012	0.0021	0.0026	0.65	**4.2×10^{-12}** [e]	90,266	4.50×10^{-7}
rs34311866 (*TMEM175*)	4:951,947	0.01/0.07/0.12/0.20	C: 0.17	TRANS, EUR	HDL/CS, TRIG/CS	351,489	−0.006	0.00097	0.0014	0.0022	0.61	**1.6×10^{-9}** [e]	115,640	2.10×10^{-6}
rs73729083 (*CREB3L2*)	7:137,559,799	0.11/0.04/0.02/0	C: 0.05	TRANS, AFR	LDL/ES, LDL/CS	84,091	−3.7	0.66	−0.37	0.95	0.53	**1.3×10^{-14}** [e]	35,909	**2.00×10^{-10}**
rs10101067 (*EYA1*)	8:72,407,374	0.04/0.07/0.13/0.06	C: 0.08	TRANS	TRIG/CS	317,809	0.014	0.0025	−0.0092	0.0053	0.069	**4.1×10^{-8}**	102,263	2.10×10^{-6}
rs4758675 (*B3GNT4*)	12:122,691,738	0.02/0/0/0	C: 0.02	AFR	TRIG/CS	12,982	−0.13	0.025	−0.029	0.057	0.85	**1.3×10^{-8}**	11,875	**3.60×10^{-8}**

Bolding indicates genome-wide statistical significance. AFR, African ancestry; chr., chromosome; CS, current smoking; EUR, European ancestry; ES, ever smoking; SE, standard error; TRANS, trans-ancestry; TRIG, triglycerides. [a]Listed variants represent the lead association within the 1-Mb region for the 2df and 1df tests of variant × smoking interaction after excluding variants within 1 Mb of known lipid-associated loci. If the variant was in or within 2 kb of a gene, the name of that gene is listed. [b]Frequency of the tested allele in 1000 Genomes data by ancestry: Asian (ASN), Americas (AMR), African (AFR), and European (EUR). [c]If a region was associated with a trait in more than one meta-analysis, the most statistically significant result is listed first and described in the table. [d]*P* values from a smoking-adjusted main-effect model (available in stage 1 cohorts only; Fig. 1). [e]Statistically significant when using a stricter *P*-value threshold, after Bonferroni correction for two smoking traits, two interaction tests, and ancestry and trans-ancestry testing ($P < 5 \times 10^{-8}/8 = 6.25 \times 10^{-9}$).

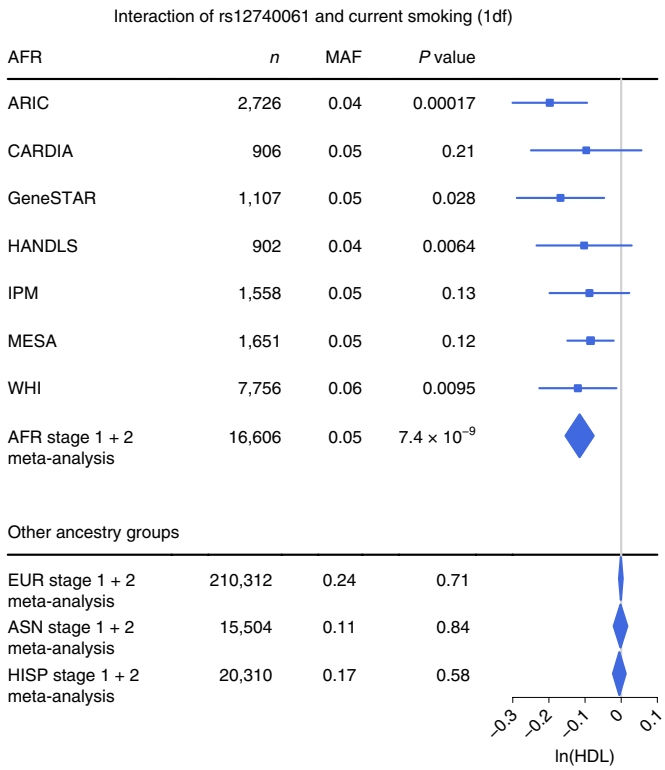

Fig. 2 | Interaction of rs12740061 (*LOC105378783*) and current smoking status (1df). Forest plots show β values (95% confidence intervals) and *P* values (1df) for the rs12740061 × current smoking interaction term in linear regression models of HDL adjusted for age, sex, study-specific covariates (if applicable), smoking status, and principal components. Results for each AFR study are shown, as well as the ancestry-specific combined stage 1 and 2 meta-analysis results.

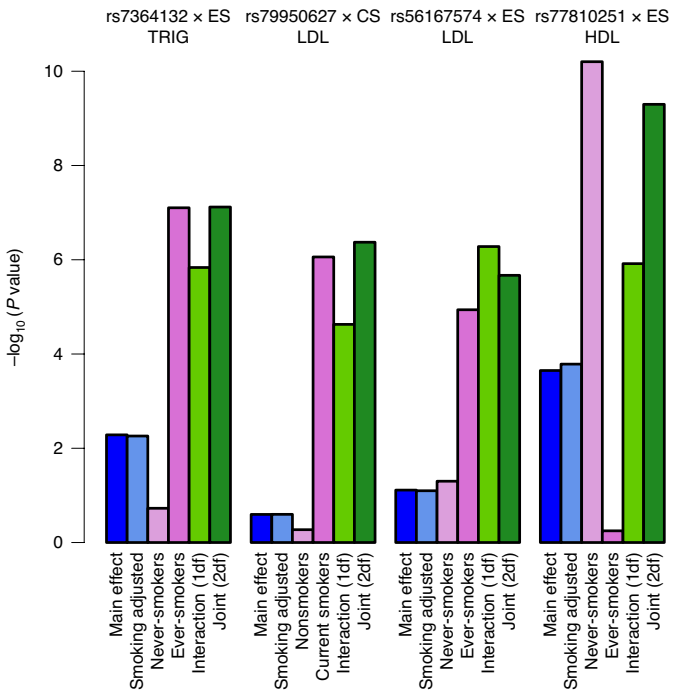

Fig. 3 | Associations observed primarily in one smoking stratum. For select variants for which an association was primarily observed in only one smoking stratum, we compare the *P* values for stage 1 linear association models, including a main-effect model adjusted for age, sex, principal components, and study-specific covariates (as appropriate) in all individuals and with stratification by smoking exposure; a model additionally adjusted for smoking exposure; and a model that also included a smoking exposure × SNP interaction term, from which a 1df test of interaction and a 2df joint test of main effect and interaction were calculated. Associations are shown, from left to right, for rs7364132 (*DGCR8*) × ever smoking and triglycerides (n = 21,834; 11,113 never-smokers, 10,725 ever-smokers), rs79950627 (*MIR4686*) × current smoking and LDL (n = 23,348; 18,384 nonsmokers, 4,973 current smokers), rs56167574 (*PRKAG2*) × ever smoking and LDL (n = 23,353; 11,700 never-smokers, 11,649 ever-smokers), and rs77810251 (*PTPRZ1*) × ever smoking and HDL (n = 23,146; 11,560 never-smokers, 11,592 ever-smokers).

annotations. We performed extensive bioinformatics annotation of variants within the 29 new loci (Tables 1 and 2). These loci included 78 associated variants that were in or near 33 unique genes (Supplementary Table 18). We performed lookup of these variants in previously conducted GWAS for related traits (Supplementary Tables 19–24), the Genotype-Tissue Expression (GTEx) portal (v7.0) and RegulomeDB (Supplementary Table 25), HaploReg v4.1 (Supplementary Table 26), and an analysis of cis and trans expression quantitative trait loci (eQTLs) in whole blood from Framingham Heart Study participants (Supplementary Table 27). Additionally, for each trait, we performed DEPICT gene prioritization (Supplementary Tables 28–30), gene set enrichment analysis (Supplementary Tables 31–33), and tissue or cell type enrichment analysis[13] (Supplementary Tables 34–37), in which we used both new and known loci. Notable findings from these follow-up analyses are summarized below by locus.

In line with our observations of an association of the C allele at rs10101067 (*EYA1*) with higher triglyceride levels, this allele was associated with increased risk of coronary artery disease (β = 0.036, P = 0.03; Supplementary Table 19), ischemic stroke (β = 0.11, P = 0.04; Supplementary Table 20), and higher waist-to-hip ratio adjusted for body mass index (BMI) (β = 0.029 units, $P = 6.5 \times 10^{-4}$; similar results were observed for waist circumference adjusted for BMI; Supplementary Table 21).

We found an association of the T allele at rs12144063 (NC_000001.10:g.28406047G>T; *EYA3*) with lower HDL levels. This allele was associated with increased risk of all stroke types (β = 0.05, P = 0.04), as well as stroke subtypes (Supplementary Table 20). rs7529792 (NC_000001.10:g.28306250C>T), a variant in LD with rs12144063 (r^2 = 0.97), regulates gene expression of *EYA3* and has a high RegulomeDB score (1b; Supplementary Table 25). HaploReg also showed regulatory features for rs12144063, identifying it as being in a promoter region expressed in liver and brain, in enhancer histone marks, and in DNase marks for *EYA3* (Supplementary Table 26). DEPICT predicted a role for these variants in regulating expression of *EYA3* and *XKR8* (Supplementary Table 28), the latter of which encodes a phospholipid scramblase important in apoptotic signaling[14].

We report an interaction between smoking and rs77810251 (*PTPRZ1*), in which the minor allele is associated with higher HDL levels only among never-smokers. Although this variant was not available for lookup in data from the Genetic Investigation of Anthropometric Traits (GIANT) consortium, a variant in this locus with a similar association, rs740965 (NC_000007.13:g.121513561T>G), was associated with lower BMI among EUR individuals (β = −0.01 kg/m^2, P = 0.01; similar results were observed for trans-ancestry analysis). This variant was also associated with lower waist circumference adjusted for BMI among EUR women (β = −0.016, P = 0.04; Supplementary Table 21). *PTPRZ1* was shown to be down-regulated in cells treated with an acute dose of nicotine[15], which supports our observation of a lack of association of *PTPRZ1* variants among ever-smokers.

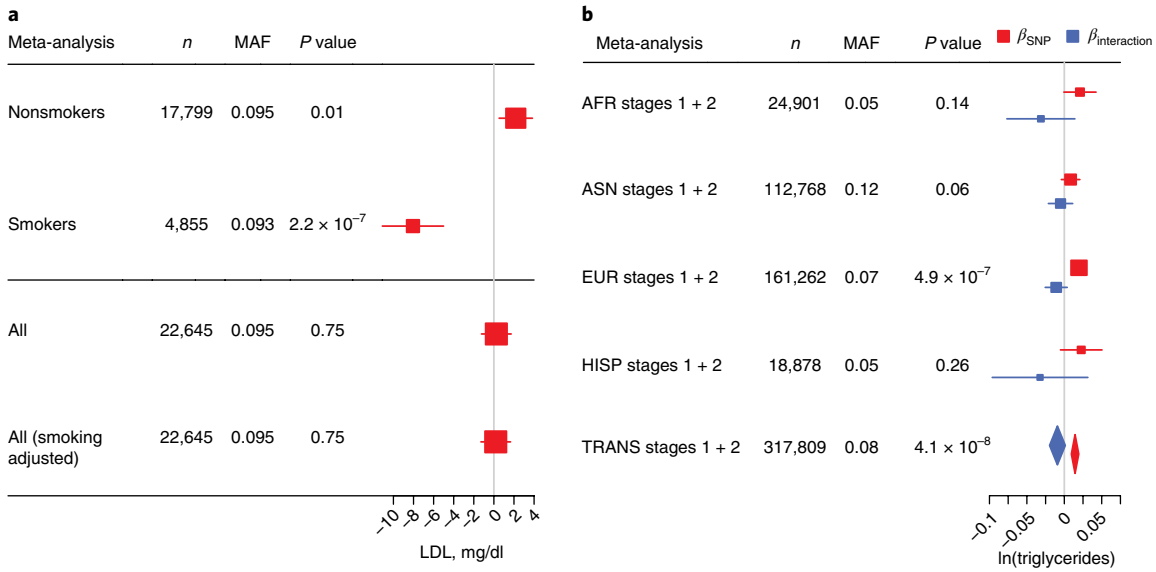

Fig. 4 | Forest plots of select associations. a, Plots showing association between rs73453125 and LDL among AFR individuals in stage 1 (where a series of models was available). Variant β values (95% confidence intervals) and *P* values are drawn from main-effect linear regression models for nonsmokers, smokers, all individuals, and all individuals with adjustment for smoking status. **b**, Plots showing association between rs10101067 (*EYA1*) and triglycerides in ancestry-specific and combined analyses from stages 1 and 2. Variant main and interaction β values (95% confidence intervals) are drawn from linear regression models that included a current smoking × SNP term and *P* values are for the 2df joint test of main effect and interaction.

We report a main effect for rs34311866 on HDL and triglyceride levels. rs34311866 encodes a missense variant in *TMEM175*, which has been associated with Parkinson's disease[16] and type 2 diabetes[17]. This variant contributes to regulation of *DGKQ* ($P = 5.3 \times 10^{-21}$) and is an eQTL for *DGKQ* in adipose, artery, lung, nerve, and thyroid tissues (Supplementary Table 25). Expression of *DGKQ* is more strongly regulated by another significantly associated variant in this locus, rs4690220 (NC_000004.11:g.980464A>G), which is located upstream of *IDUA* and in an intron of *SLC26A1*. This variant had a high score in RegulomeDB (1f), supporting the idea that it potentially has a functional effect (Supplementary Table 25). Notably, *DGKQ* has been implicated in studies of cholesterol metabolism[18], bile acid signaling, glucose homoeostasis in hepatocytes[19], primary biliary cirrhosis[20], and Parkinson's disease[21–24]. The DGKQ protein interacts with the key lipid enzymes LPL, LIPG, and PNPLA3 (Supplementary Fig. 7). These results suggest that the observed association with HDL and triglycerides could act on cholesterol metabolism through regulation of *DGKQ*. Also, rs34311866 is a trans eQTL for *GNPDA1* (Supplementary Table 27); expression of this gene has been associated with a set of traits, including hyperlipidemia[25].

In our data, there was a significant interaction between rs12740061 (*LOC105378783*) and smoking, such that the minor allele was associated with decreased HDL levels only among current smokers. This variant is a trans eQTL for *TAS1R1* (Supplementary Table 27). Variants in this gene have been found to influence taste receptors, notably affecting cigarette smoking habits[26].

Discussion

In this study, we evaluated gene–smoking interactions in large, multi-ancestry meta-analyses of serum lipids, while using varying associations among smoking subgroups to improve the ability to detect new lipid-associated loci. We report 13 new loci for serum lipids from stage 1 and 2 meta-analyses. Sixteen additional statistically significant new loci were found in stage 1 but were unavailable for analysis in stage 2. All 29 new associations had a low *q* value ($P < 3 \times 10^{-4}$). Using both the 1df test of interaction and the 2df joint test of main and interaction effects in this study allowed us to improve our inferences on the basis of the results: the 2df test bolstered the power to detect interactions, while the 1df test could discriminate between associations that predominantly reflected main effects versus interactions.

Our results provide support for future efforts to evaluate lifestyle interactions with complex traits. We identified loci for which an association with serum lipids was only observed in one smoking stratum. In main-effect models of these loci, the signal from one subgroup was not detected when all individuals were evaluated together (regardless of adjustment for smoking). These loci could only be observed through analysis that was stratified by smoking status or contained an interaction term, highlighting the importance of considering potential effect modification in association studies. Additionally, through use of the joint 2df test, we identified six loci that seem to represent new main effects. In agreement with this characterization, five of these loci were within 500 kb of variants identified in recent large-scale association studies that used main-effect models: *ETV*[27–29], *TMEM175* (ref. [28]), *EYA1* (ref. [28]), *EYA3* (ref. [28]), and *B3GNT4* (ref. [28]).

With 23,753 AFR individuals in the stage 1 analyses and 30,970 AFR individuals overall, this work represents one of the largest studies of serum lipids in AFR cohorts. It is therefore not surprising that two of our new lipid-associated loci (*CREB3L2* and *B3GNT4*) seem to be driven primarily by genetic main effects. Notably, these associations could not have been detected in EUR individuals, as the tested allele for both rs4758675 (*B3GNT4*) and rs73729083 (*CREB3L2*) is absent in EUR populations.

In addition to these probable main-effect loci, the prominence of the new loci that were statistically significant only in AFR meta-analyses deserves further discussion. Some findings could not be effectively evaluated in other ancestry groups because of differences in MAF between the ancestry groups, with the minor alleles for half of the variants much more frequent in AFR populations. More puzzling, however, is the discovery of loci with evidence of strong interactions in the AFR ancestry group but not in meta-analyses in other ancestry groups, despite comparable or higher allele frequencies in these groups, such as were observed for rs12740061 (*LOC105378783*; Fig. 2) or rs17150980 (*MAGI2*;

Table 2 | Statistically significant ($P < 5 \times 10^{-8}$) loci in stage 1 meta-analysis unavailable in stage 2

Index variant (nearest gene)[a]	Build 37 chr:position	1000 Genomes freq.[b] AFR/AMR/ASN/EUR	Tested allele: freq.	Ancestry	Trait/exposure	Stage 1							
						n	Effect	SE	Int. effect	SE	1df interaction *P* value[b]	2df joint *P* value	Adj. main-effect *P* value[c]
rs140602625 (*EXOC6B*)	2:72,849,325	0.01/0/0/0	C: 0.02	AFR	LDL/CS	7,755	−3.4	3.1	−35	7.1	1.0×10^{-6}	**1.5×10^{-8}**	0.018
rs114138886 (*LOC107985905*)	2:84,428,024	0.02/0/0/0	T: 0.02	AFR	LDL/CS	7,755	2.4	2.9	−29	5.4	9.3×10^{-8}	**4.4×10^{-8}**	0.47
rs149776574 (*REEP1*)	2:86,472,455	0.01/0.08/0/0.06	G: 0.02	AFR	TRIG/CS	7,756	−0.048	0.033	0.40	0.069	**4.2×10^{-10}**[d]	**5.1×10^{-10}**[d]	0.88
rs143396479 (*LOC105374426/TMEM33*)	4:41,911,366	0.02/0/0/0	A: 0.01	AFR	LDL/ES	10,912	−16.0	2.6	15	4.5	0.022	**6.8×10^{-9}**	0.0094
rs148187465 (*MARCH1*)	4:164,639,694	0.01/0/0/0	C: 0.01	AFR	LDL/CS	7,755	−2.1	3.0	−32	6.2	3.7×10^{-7}	**4.9×10^{-9}**[d]	0.032
rs76687692 (*G3BP1*)	5:151,189,283	0.03/0/0/0	A: 0.01	AFR	LDL/CS	9,418	2.7	3.2	25	5.5	0.0013	**4.8×10^{-9}**[d]	0.0016
rs73339842 (*LINC01938*)	5:164,967,406	0.02/0.01/0/0	G: 0.02	AFR	TRIG/CS	7,756	0.046	0.033	−0.41	0.071	**8.5×10^{-9}**	**3.3×10^{-8}**	0.96
rs115580718 (*BMP6*)	6:7,880,037	0.02/0/0/0	G: 0.01	AFR	TRIG/CS	7,756	−0.12	0.036	−0.29	0.082	0.00045	**1.2×10^{-9}**[d]	1.6×10^{-6}
rs17150980 (*MAGI2*)	7:78,173,734	0/0.12/0.45/0.01	C: 0.03	AFR	TRIG/ES	12,972	−0.17	0.028	0.24	0.044	7.5×10^{-8}	**1.4×10^{-9}**[d]	0.085
rs116592443 (*LYZL2*)	10:30,884,890	0.02/0/0/0	A: 0.01	AFR	TRIG/CS	7,756	0.073	0.038	−0.46	0.081	**1.8×10^{-8}**	1.2×10^{-7}	0.76
rs115628664 (*UNC5B*)	10:2,899,880	0.03/0/0/0	G: 0.01	AFR	TRIG/CS	7,756	0.027	0.040	−0.39	0.071	**4.7×10^{-8}**	**6.7×10^{-9}**[d]	0.44
rs183911507 (*TP53I11*)	11:44,978,366	0.01/0/0/0	G: 0.02	AFR	TRIG/CS	10,287	−0.043	0.029	0.33	0.059	**1.7×10^{-8}**	6.5×10^{-8}	0.82
rs199771018 (*STOML3*)	13:39,507,838	0.02/0/0/0	T: 0.02	AFR	HDL/CS	7,756	−0.019	0.019	0.23	0.037	**1.2×10^{-9}**[d]	**6.3×10^{-10}**[d]	0.55
rs190976513 (*LOC105370255*)	13:71,114,207	0.02/0.01/0/0	A: 0.02	AFR	LDL/CS	10,234	−5.1	2.6	−20	5.2	9.3×10^{-5}	**3.2×10^{-8}**	1.1×10^{-4}
rs182600360 (*LOC105370531*)	14:63,607,120	0.02/0/0/0	A: 0.02	AFR	LDL/CS	7,755	6.6	3.3	−39	7.1	**4.4×10^{-8}**	3.3×10^{-7}	0.56
rs62064821 (*CCT6B*)	17:33,280,904	0.01/0.04/0/0.06	T: 0.01	AFR	LDL/CS	10,234	8.5	3.3	−30	5.5	**3.1×10^{-8}**	6.0×10^{-7}	0.17

All loci shown in the table have some evidence of interaction ($P < 0.05$ in 1df test of interaction); thus, results are not categorized into 'loci with evidence for interaction' and 'probable main-effect loci (no evidence for interaction)' as in Table 1. Bolding indicates genome-wide statistical significance. AFR, African; CS, current smoking; ES, ever smoking; SE, standard error; TRIG, triglycerides. [a]Listed variants represent the lead association within the 1-Mb region for the 2df and 1df tests of variant × smoking interaction after excluding variants within 1 Mb of known lipid-associated loci. If the variant was in or within 2 kb of a gene, the name of that gene is listed. [b]Frequency of the tested allele in 1000 Genomes data by ancestry: Asian (ASN), Americas (AMR), African (AFR), and European (EUR). [c]*P* values from a smoking-adjusted main-effect model (available in stage 1 cohorts only; Fig. 1). [d]Statistically significant when using a stricter *P*-value threshold, after Bonferroni correction for two smoking traits, two interaction tests, and ancestry and trans-ancestry testing ($5 \times 10^{-8}/8 = 6.25 \times 10^{-9}$).

Supplementary Fig. 6). This phenomenon suggests inter-ancestry differences in genomic or environmental context. There are variants in LD ($r^2 > 0.2$) with rs12740061 (*LOC105378783*) and rs17150980 (*MAGI2*) in AFR populations that are not in LD with these variants in other ancestry groups[30], but these variants were directly tested in our study with no evidence of association in non-AFR analyses. Thus, it is unlikely that inter-ancestry differences in LD explain these results, although unmeasured causal variants are a possibility. Inter-ancestry differences in smoking are also a potential explanation. In addition to known differences in smoking patterns[31], there are pronounced differences between ancestry groups in preferred cigarette type, with over 85% of AFR smokers using menthol cigarettes as compared to 29% of EUR smokers (in the United States)[32]. Menthol cigarettes are thought to facilitate greater absorption of harmful chemicals because of deeper inhalation[31,33], through desensitization of the nicotinic acetylcholine receptors that cause nicotine-induced irritation[34]. Evidence for an excess risk of cardiovascular disease associated with mentholated cigarettes, however, is equivocal[35–39]. Ancestry differences in smoking-related metabolites and carcinogens have been reported[40–43], and differential metabolism of key compounds may underlie observed differences by ancestry group. Some behaviors or conditions that co-occur with smoking may also differ by ancestry, and this additional factor may modify observed genetic associations with serum lipids.

The biological mechanisms through which smoking influences observed genetic associations will require further investigation, as the myriad components of cigarette smoke and their downstream consequences (including oxidative stress and inflammation) affect pathways throughout the body[44]. However, there is evidence for differential expression of *PTPRZ1* (ref. [15]), *LPL*[15], and *LDLR*[45] in cells exposed to an acute dose of nicotine. Also, concentrations of CETP[46], ApoB[47], and LPL[48] are associated with smoking status.

The sample size attained for diverse ancestry groups is a key strength of our study, particularly among AFR studies. As a result, we were able to identify loci that had not been previously detected in meta-analyses of ancestry groups that are better represented in genomic research. Additionally, the use of nested models in our stage 1 analyses allowed us to more fully characterize loci. Despite these strengths, however, a smaller number of AFR studies were available for stage 2, resulting in an inability to follow up on some of our low-frequency findings from stage 1.

In conclusion, this large, multi-ancestry genome-wide study of the effects of gene–smoking interactions on serum lipids identified 13 new loci on the basis of combined analyses of stages 1 and 2 as well as 16 additional new loci on the basis of stage 1 that were unavailable in stage 2. Associations for some loci were detected only in analyses stratified by smoking status or with a smoking interaction term, thus motivating further study of gene × environment interactions for other lifestyle factors to identify new loci associated with lipids and other complex traits. We demonstrate the importance of including diverse populations, attaining a sample size in these analyses sufficient for discovery of new main-effect lipid-associated loci in AFR populations. Careful consideration of ancestry may be of particular importance for gene × environment interactions, as ancestry may be a proxy for both genomic and environmental context.

URLs. 1000 Genomes Project, http://www.internationalgenome.org/; dbGaP, https://www.ncbi.nlm.nih.gov/gap; dbSNP, http://ncbi.nlm.nih.gov/snp/; DEPICT, http://data.broadinstitute.org/mpg/depict/; EasyQC, http://www.genepi-regensburg.de/easyqc; EasyStrata, http://www.genepi-regensburg.de/easystrata; ENCODE, https://www.encodeproject.org/; forestplot, http://cran.r-project.org/web/packages/forestplot/; GCTA, http://cnsgenomics.com/software/gcta; geepack, http://cran.r-project.org/web/packages/geepack/; GenABEL, https://github.com/cran/GenABEL; Gene Ontology, http://www.geneontology.org/; GTEx, https://gtexportal.org/home/; HaploReg, http://pubs.broadinstitute.org/mammals/haploreg/haploreg.php; KEGG, http://www.genome.jp/kegg/; LocusZoom, http://locuszoom.sph.umich.edu/; METAL, http://genome.sph.umich.edu/wiki/METAL; NCBI Entrez gene, https://www.ncbi.nlm.nih.gov/gene/; ProbABEL, https://github.com/GenABEL-Project/ProbABEL; Reactome, http://bioconductor.org/packages/release/data/annotation/html/reactome.db.html; RegulomeDB, http://www.regulomedb.org/; Roadmap Epignomics, http://www.roadmapepigenomics.org/; sandwich, http://cran.r-project.org/web/packages/sandwich/index.html; STRING database, http://string-db.org/.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0378-y.

Received: 8 January 2018; Accepted: 7 February 2019;
Published online: 29 March 2019

References

1. Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
2. Do, R. et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat. Genet.* **45**, 1345–1352 (2013).
3. Peloso, G. M. et al. Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am. J. Hum. Genet.* **94**, 223–232 (2014).
4. Spracklen, C. N. et al. Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Hum. Mol. Genet.* **26**, 1770–1784 (2017).
5. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
6. Kathiresan, S. et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. *N. Engl. J. Med.* **358**, 1240–1249 (2008).
7. Kar, D. et al. Relationship of cardiometabolic parameters in non-smokers, current smokers, and quitters in diabetes: a systematic review and meta-analysis. *Cardiovasc. Diabetol.* **15**, 158 (2016).
8. Zong, C. et al. Cigarette smoke exposure impairs reverse cholesterol transport which can be minimized by treatment of hydrogen-saturated saline. *Lipids Health Dis.* **14**, 159 (2015).
9. Manning, A. K. et al. Meta-analysis of gene–environment interaction: joint estimation of SNP and SNP × environment regression coefficients. *Genet. Epidemiol.* **35**, 11–18 (2011).
10. Psaty, B. M. et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from five cohorts. *Circ. Cardiovasc. Genet.* **2**, 73–80 (2009).
11. Rao, D. C. et al. Multiancestry study of gene–lifestyle interactions for cardiovascular traits in 610 475 individuals from 124 cohorts: design and rationale. *Circ. Cardiovasc. Genet.* **10**, e001649 (2017).
12. Lanktree, M. B. et al. Genetic meta-analysis of 15,901 African Americans identifies variation in *EXOC3L1* is associated with HDL concentration. *J. Lipid Res.* **56**, 1781–1786 (2015).
13. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890–5890 (2015).
14. Suzuki, J., Imanishi, E. & Nagata, S. Xkr8 phospholipid scrambling complex in apoptotic phosphatidylserine exposure. *Proc. Natl Acad. Sci. USA* **113**, 9509–9514 (2016).
15. Wang, J. et al. Genome-wide expression analysis reveals diverse effects of acute nicotine exposure on neuronal function-related genes and pathways. *Front. Psychiatry* **2**, 5 (2011).
16. International Parkinson Disease Genomics Consortium. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**, 641–649 (2011).
17. Ng, M. C. Y. et al. Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. *PLoS Genet.* **10**, e1004517 (2014).
18. Cai, K., Lucki, N. C. & Sewer, M. B. Silencing diacylglycerol kinase-θ expression reduces steroid hormone biosynthesis and cholesterol metabolism in human adrenocortical cells. *Biochim. Biophys. Acta* **1841**, 552–562 (2014).
19. Cai, K. & Sewer, M. B. Diacylglycerol kinase θ couples farnesoid X receptor–dependent bile acid signalling to Akt activation and glucose homoeostasis in hepatocytes. *Biochem. J.* **454**, 267–274 (2013).
20. Cordell, H. J. et al. International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat. Commun.* **6**, 8019 (2015).
21. Edwards, T. L. et al. Genome-wide association study confirms SNPs in *SNCA* and the *MAPT* region as common risk factors for Parkinson disease. *Ann. Hum. Genet.* **74**, 97–109 (2010).
22. Lill, C. M. et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: the PDGene Database. *PLoS Genet.* **8**, e1002548 (2012).
23. Nalls, M. A. et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993 (2014).
24. Pankratz, N. et al. Meta-analysis of Parkinson disease: identification of a novel locus, *RIT2*. *Ann. Neurol.* **71**, 370–384 (2012).
25. Wang, J. et al. Phlegm-dampness constitution: genomics, susceptibility, adjustment and treatment with traditional Chinese medicine. *Am. J. Chin. Med.* **41**, 253–262 (2013).
26. Choi, J.-H. et al. Variations in *TAS1R* taste receptor gene family modify food intake and gastric cancer risk in a Korean population. *Mol. Nutr. Food Res.* **60**, 2433–2445 (2016).
27. Hoffmann, T. J. et al. A large electronic-health-record-based genome-wide study of serum lipids. *Nat. Genet.* **50**, 401–413 (2018).
28. Klarin, D. et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat. Genet.* **50**, 1514–1523 (2018).
29. Liu, D. J. et al. Exome-wide association study of plasma lipids in ~300,000 individuals. *Nat. Genet.* **49**, 1758 (2017).
30. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
31. *Tobacco Use Among U.S. Racial/Ethnic Minority Groups—African Americans, American Indians and Alaska Natives, Asian Americans and Pacific Islanders, and Hispanics: a Report of the Surgeon General* (US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 1998).
32. Villanti, A. C. et al. Changes in the prevalence and correlates of menthol cigarette use in the USA, 2004–2014. *Tobacco Control* **25**, ii14 (2016).
33. Ross, K. C., Dempsey, D. A., St.Helen, G., Delucchi, K. & Benowitz, N. L. The influence of puff characteristics, nicotine dependence, and rate of nicotine

metabolism on daily nicotine exposure in African American smokers. *Cancer Epidemiol. Biomarkers Prev.* **25**, 936–943 (2016).
34. Ton, H. T. et al. Menthol enhances the desensitization of human α3β4 nicotinic acetylcholine receptors. *Mol. Pharmacol.* **88**, 256–264 (2015).
35. Alexander, L. A. et al. Why we must continue to investigate menthol's role in the African American smoking paradox. *Nicotine Tobacco Res.* **18**, S91–S101 (2016).
36. Jones, M. R., Tellez-Plaza, M. & Navas-Acien, A. Smoking, menthol cigarettes and all-cause, cancer and cardiovascular mortality: evidence from the National Health and Nutrition Examination Survey (NHANES) and a meta-analysis. *PLoS One* **8**, e77941 (2013).
37. Munro, H. M., Tarone, R. E., Wang, T. J. & Blot, W. J. Menthol and nonmenthol cigarette smoking: all-cause deaths, cardiovascular disease deaths, and other causes of death among blacks and whites. *Circulation* **133**, 1861–1866 (2016).
38. Murray, R. P., Connett, J. E., Skeans, M. A. & Tashkin, D. P. Menthol cigarettes and health risks in lung health study data. *Nicotine Tobacco Res.* **9**, 101–107 (2007).
39. Vozoris, N. T. Mentholated cigarettes and cardiovascular and pulmonary diseases: a population-based study. *Arch. Intern. Med.* **172**, 590–593 (2012).
40. Pérez-Stable, E. J., Herrera, B., Jacob, I. P. & Benowitz, N. L. Nicotine metabolism and intake in black and white smokers. *J. Am. Med. Assoc.* **280**, 152–156 (1998).
41. Khariwala, S. S. et al. Cotinine and tobacco-specific carcinogen exposure among nondaily smokers in a multiethnic sample. *Nicotine Tobacco Res.* **16**, 600–605 (2014).
42. Jain, R. B. Distributions of selected urinary metabolites of volatile organic compounds by age, gender, race/ethnicity, and smoking status in a representative sample of U.S. adults. *Environ. Toxicol. Pharmacol.* **40**, 471–479 (2015).
43. Benowitz, N. L., Dains, K. M., Dempsey, D., Wilson, M. & Jacob, P. Racial differences in the relationship between number of cigarettes smoked and nicotine and carcinogen exposure. *Nicotine Tobacco Res.* **13**, 772–783 (2011).
44. *The Health Consequences of Smoking—50 Years of Progress: a Report of the Surgeon General* (US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014).
45. Ito, S. et al. Nicotine-induced expression of low-density lipoprotein receptor in oral epithelial cells. *PLoS One* **8**, e82563 (2013).
46. Dullaart, R. P., Hoogenberg, K., Dikkeschei, B. D. & van Tol, A. Higher plasma lipid transfer protein activities and unfavorable lipoprotein changes in cigarette-smoking men. *Arterioscler. Thromb.* **14**, 1581–1585 (1994).
47. Frondelius, K. et al. Lifestyle and dietary determinants of serum apolipoprotein A1 and apolipoprotein B concentrations: cross-sectional analyses within a Swedish cohort of 24,984 individuals. *Nutrients* **9**, 211 (2017).
48. Onat, A. et al. Preheparin serum lipoprotein lipase mass interacts with gender, gene polymorphism and, positively, with smoking. *Clin. Chem. Lab. Med.* **47**, 208 (2009).

Acknowledgements

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 Human Genome Research Institute; the National Institutes of Health; or the US Department of Health and Human Services. This project was largely supported by a grant from the US National Heart, Lung, and Blood Institute of the National Institutes of Health (R01HL118305) and by the Intramural Research Program of the National Human Genome Research Institute of the National Institutes of Health through the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute, the National Institute of Diabetes and Digestive and Kidney Diseases, the Center for Information Technology, and the Office of the Director at the National Institutes of Health (Z01HG200362). Additional and study-specific acknowledgments appear in the Supplementary Note.

Author contributions

All authors reviewed and approved the manuscript. Study concept and design: A.B.Z., A.C.M., A.C.P., A.J.O., A.R., A.R.B., A.R.W., B.I.F., B.L.H., C.A.M.K., C. Ballantyne, C. Bouchard, C.C.K., C.C.L., C.D.L., C.H., C. Langenberg, C.M.v.D., C.M.K., C.N.R., C.-T.L., C.Y., C.-Y.C., D.C.R., D.I.C., D.M.B., D.R.W., D.W.B., E.B., E.P.B., E.R.F., E.S.T., F.R.R., G.W., H.A., H.J.d.S., H. Watkins, I.G., I.J.D., I.K., J.B.J., J. Ding, J. Divers, J.D.F., J. E. Hixson, J.E.K., J.I.R., J.K., Jianjun Liu, J.M.C., J.M.S., J.-M.Y., K.C., K.K.L., K.L.M., L.A.C., Lifelines Cohort Study, L.E.W., L.J.L., M.A.I., M.A.P., M. Brown, M. Boehnke, M. Farrall, M. Fornage, M. He, M.K., M.K.E., M. Laakso, M.S., N.G.F., N.J.S., N.J.W., N.K., N.L.P., N.P., N.S., O.P., O.T.R., P.F., P.G., P.H., P.K., P.K.E.M., P.M.R., P.S., R.A.S., R.M.D., R.R., R.S.C., S.C., S.K.M., S.L.R.K., S.R., S.T.T., T.A., T.A.L., T.B.H., T.F., T.K.R., T. Lehtimäki, T.N.K., T.R., T.W., T.Y.W., U.d.F., V.G., W.B.W., W.P.K., X.G., Y.K., Y. Liu, Y.W., Y.X.W., and Y.Y.T. Phenotype data acquisition and/or quality control: A.B.Z., A.C., A.C.P., A.D.M., A.G., A.J.O., A.K., A. Metspalu, A.P., A.P.R., A.R.B., A.R.V.R.H., A.R.W., A.W.M., B.E.C., B.G., B.I.F., B.L.H., B.M.P., B.O.T., B. Penninx, C.A.M.K., C. Ballantyne, C. Bouchard, C.D.L., C.E.L., C. Gieger, C.H., C.J., C. Langenberg, C. Li, C.M.K., C.M.v.D., C.N.R., C.O.S., C.P.N., C.Y., D.C.R., D.H., D.M.B., D.R.J., D.R.W., D.W.B., E.E., E.P.B., E.S.T., F.R., F.R.R., F.-C.H., G.J.P., G.R.B., G.W., H.G., H.J.d.S., H.J.G., H.M.S., H. Tiemeier, H. Wang, I.J.D., I.K., I.-T.L., J.A.S., J.B.J., J. Ding, J. Divers, J.D.F., J.E.K., J.H.Z., Jian'an Luan, Jingjing Liang, J.M.C., J.M.S., J.-M.J.J., J.-M.Y., J.-S.W., K.C., K.K.L., K. Leander, K. Liu, K. Schwander, K.-H.L., L.A.C., Lifelines Cohort Study, L.F.B., L.J.B., L.M., L.M.R., L.R.Y., M. Alver, M. Amini, M.A.P., M. Brown, M. Boissel, M.C., M.F.F., M. He, M. Hirata, M.K., M.K.E., M.K.W., M.N., M.P.C., M.S., M.W., N.F., N.G.F., N.J.S., N.J.W., N.L.P., N.P., N.S., N.Y.Q.T., O.H.F., O.P., O.T.R., P.A.P., P.H., P.J.S., P.K., P.K.E.M., P.M.R., P.S., P.W.F., R.A.S., R.M., R.M.D., R.R., R.S.C., S.E.H., S.L.R.K., S.S., S.S.R., S.T.T., T.A.L., T.E., T.F., T.K., T.K.R., T. Lehtimäki, T.M., T.N.K., T.R., T.S., T.W., T.-D.W., U.d.F., Understanding Society Scientific Group, W.B.W., W.P.K., Y.C.T., Y. Liu, and Y. Lu. Genotype data acquisition and/or quality control: A.B.Z., A.C.P., A.G., A.G.U., A.L., A. Metspalu, A.R.B., A.R.V.R.H., A.T.K., A.V.S., B.E.C., B.G., B.I.F., B.L.H., B.M.P., B.O.T., B. Prins, C. Bouchard, C.C.K., C.C.L., C. Gao, C.K., C. Langenberg, C. Li, C.M.K., C.N.R., C.P.N., C.-K.H., C.-T.L., D.C.R., D.E.A., D.I.C., D.M.B., D.O.M.-K., E.B., E.B.W., E.E., E.L., E.P.B., E.R.F., E.S.T., E.Z., F.G., F.P.H., F.R., F.R.R., F.-C.H., H.G., H. Wang, I.J.D., I.K., I.M.N., J.A.S., J. E. Hixson, J. E. Huffman, J.E.K., J.F.C., J.H.Z., J.I.R., Jian'an Luan, Jingjing Liang, Jianjun Liu, Jingmin Liu, J.M.C., J.M.S., K.C., K.D.T., K.K.L., K. Leander, K. Schwander, K. Strauch, L.A.C., Lifelines Cohort Study, L.M., L.M.R., L.R.Y., Lan Wang, L.-P.L., M. Alver, M. Amini, M.A.N., M.A.P., M. Boissel, M.C., M. Fornage, M.F.F., M.K., M.K.E., M.P., M.P.C., N.A., N.D.P., N.J.S., N.J.W., N.K., N.L.P., N.S., O.P., P.B.M., P.H., P.J.V.M., P.K.E.M., P.W.F., R.A.S., R.D., R.J.F.L., R.M., R.N.E., S.E.H., S.H., S.K.M., S.L.R.K., S.S.R., S.T.T., T.E., T.K.R., T. Lehtimäki, T.N.K., T.R., U.d.F., Understanding Society Scientific Group, W. Zhao, X.D., X.S., X.Z., Y.F., Y.H., Y. Liu, Y. Momozawa, Y.Y.T., Y.-D.I.C., and Z.A. Data analysis and interpretation: A.B.Z., A.C.M., A.C.P., A.G., A. Mahajan, A.P.M., A.P.R., A.R., A.R.B., A.R.V.R.H., A.S., A.U.J., A.V.S., B.I.F., B.K., B.M.P., B.O.T., B. Prins, C.A.W., C. Bouchard, C.D.L., C. Gao, C. Gieger, C. Li, C.N.R., C.P.N., C.-T.L., C.-Y.C., D.C.R., D.H., D.I.C., D.M.B., D.O.M.-K., D.V., E.B.W., E.E., E.L., E.R.F., E.S.T., F.G., F.P.H., F.T., F.-C.H., G.C., G.W., H.G., H.S., I.G., I.M.N., I.N., J.A.S., J.B.J., J. Divers, J. E. Hixson, J. E. Huffman, J.F.C., J.H.Z., Jian'an Luan, Jingmin Liu, J.S.F., J.Y., J.Z., K. Leander, K.R., L.A.C., Lifelines Cohort Study, L.F.B., L.M.R., L.R.Y., Lan Wang, Lihua Wang, L.-P.L., M. Amini, M.A.N., M.A.R., M.A.S., M. Fornage, M. Farrall, M.F.F., M.K., M.K.E., M.P., M.R., M.R.B., M.S., N.D.P., N.F., N.J.S., N.M., P.A.P., P.B.M., P.H., P.J.V.M., P.S.V., R.D., R.J.F.L., R.N., R.N.E., R.S.C., S.A.G., S.B.K., S.E.H., S.H., S.K.M., S.L., S.L.R.K., S.M.T., T.K.R., T. Louie, T.M.B., T.N.K., T.R., T.S., T.V.V., T.W.W., T.Y.W., W.B.W., W. Zhao, X.C., X.D., X.G., X.S., Y.H., Y.J., Y.K., Y. Lu, and Y.X.W. Look-ups: A.E.J., A. Mahajan, A.P.M., A.R.B., COGENT-Kidney Consortium, D.I.C., K.Y., M.G., N.F., and T.W.W. These authors constitute the writing group: A.R.B., Y.J.S., M.R.B., T.W.W., A.T.K., I.N., K.S., X.Z., L.J.B., W.J.G., K.R., P.B.M., A.C.M., D.C.R., C.N.R., and L.A.C.

Competing interests

The authors declare no competing interests except for the following. O.H.F. received grants from Metagenics (on women's health and epigenetics) and from Nestle (on child health). J.B.J. serves as a consultant for Mundipharma Co., is a patent holder with Biocompatibles UK, Ltd ("Treatment of Eye Diseases using Encapsulated Cells Encoding and Secreting Neuroprotective Factor and/or Anti-angiogenic Factor"; patent number 20,120,263,794), and has a patent application with the University of Heidelberg ("Agents for Use in the Therapeutic or Prophylactic Treatment of Myopia or Hyperopia"; Europäische Patentanmeldung 15000771.4). The participation of M.A.N. is supported by a consulting contract between Data Tecnica International and the National Institute on Aging, National Institutes of Health; as a possible conflict of interest, M.A.N. also consults for Illumina, the Michael J. Fox Foundation, and University of California Healthcare, among others. N.P. has received financial support from several pharmaceutical companies that manufacture either blood pressure–lowering or lipid-lowering agents, or both, and consultancy fees. P.S. has received research awards from Pfizer. B.M.P. serves on the DSMB of a clinical trial funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. L.J.B. is listed as an inventor on issued US patent 8,080,371 ("Markers for Addiction"), covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0378-y.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.R.B., C.N.R. or L.A.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Amy R. Bentley[1,213*], Yun J. Sung[2,213], Michael R. Brown[3,213], Thomas W. Winkler[4,213], Aldi T. Kraja[5,213], Ioanna Ntalla[6,213], Karen Schwander[2], Daniel I. Chasman[7,8], Elise Lim[9], Xuan Deng[9], Xiuqing Guo[10], Jingmin Liu[11], Yingchang Lu[12], Ching-Yu Cheng[13,14,15], Xueling Sim[16], Dina Vojinovic[17], Jennifer E. Huffman[18], Solomon K. Musani[19], Changwei Li[20], Mary F. Feitosa[5], Melissa A. Richard[21], Raymond Noordam[22], Jenna Baker[1], Guanjie Chen[1], Hugues Aschard[23,24], Traci M. Bartz[25,26], Jingzhong Ding[27], Rajkumar Dorajoo[28], Alisa K. Manning[29,30], Tuomo Rankinen[31], Albert V. Smith[32,33], Salman M. Tajuddin[34], Wei Zhao[35], Mariaelisa Graff[36], Maris Alver[37], Mathilde Boissel[38], Jin Fang Chai[16], Xu Chen[39], Jasmin Divers[40], Evangelos Evangelou[41,42], Chuan Gao[43], Anuj Goel[44,45], Yanick Hagemeijer[46], Sarah E. Harris[47,48], Fernando P. Hartwig[49,50], Meian He[51], Andrea R. V. R. Horimoto[52], Fang-Chi Hsu[40], Yi-Jen Hung[53,54], Anne U. Jackson[55], Anuradhani Kasturiratne[56], Pirjo Komulainen[57], Brigitte Kühnel[58,59], Karin Leander[60], Keng-Hung Lin[61], Jian'an Luan[62], Leo-Pekka Lyytikäinen[63,64], Nana Matoba[65], Ilja M. Nolte[66], Maik Pietzner[67,68], Bram Prins[69], Muhammad Riaz[70,71], Antonietta Robino[72], M. Abdullah Said[46], Nicole Schupf[73], Robert A. Scott[62], Tamar Sofer[30,74], Alena Stančáková[75], Fumihiko Takeuchi[76], Bamidele O. Tayo[77], Peter J. van der Most[66], Tibor V. Varga[78], Tzung-Dau Wang[79,80,81], Yajuan Wang[82], Erin B. Ware[83], Wanqing Wen[84], Yong-Bing Xiang[85], Lisa R. Yanek[86], Weihua Zhang[87,88], Jing Hua Zhao[62], Adebowale Adeyemo[1], Saima Afaq[87], Najaf Amin[17], Marzyeh Amini[66], Dan E. Arking[89], Zorayr Arzumanyan[10], Tin Aung[13,15,90], Christie Ballantyne[91,92], R. Graham Barr[93], Lawrence F. Bielak[35], Eric Boerwinkle[3,94], Erwin P. Bottinger[12], Ulrich Broeckel[95], Morris Brown[6,96], Brian E. Cade[74], Archie Campbell[97], Mickaël Canouil[38], Sabanayagam Charumathi[13,14], Yii-Der Ida Chen[10], Kaare Christensen[98], COGENT-Kidney Consortium[99], Maria Pina Concas[72], John M. Connell[100], Lisa de las Fuentes[2,101], H. Janaka de Silva[102], Paul S. de Vries[3], Ayo Doumatey[1], Qing Duan[103], Charles B. Eaton[104], Ruben N. Eppinga[46], Jessica D. Faul[83], James S. Floyd[26,105], Nita G. Forouhi[62], Terrence Forrester[106], Yechiel Friedlander[107], Ilaria Gandin[108], He Gao[41], Mohsen Ghanbari[17,109], Sina A. Gharib[110], Bruna Gigante[60], Franco Giulianini[7], Hans J. Grabe[111], C. Charles Gu[2], Tamara B. Harris[112], Sami Heikkinen[75,113], Chew-Kiat Heng[114,115], Makoto Hirata[116], James E. Hixson[3], M. Arfan Ikram[17,117,118], EPIC-InterAct Consortium[99], Yucheng Jia[10], Roby Joehanes[119,120], Craig Johnson[121], Jost Bruno Jonas[122,123], Anne E. Justice[36], Tomohiro Katsuya[124,125], Chiea Chuen Khor[28], Tuomas O. Kilpeläinen[126,127], Woon-Puay Koh[16,128], Ivana Kolcic[129], Charles Kooperberg[130], Jose E. Krieger[52], Stephen B. Kritchevsky[131], Michiaki Kubo[132], Johanna Kuusisto[75], Timo A. Lakka[57,113,133], Carl D. Langefeld[40], Claudia Langenberg[62], Lenore J. Launer[112], Benjamin Lehne[134], Cora E. Lewis[135], Yize Li[2], Jingjing Liang[82], Shiow Lin[5], Ching-Ti Liu[9], Jianjun Liu[28,136], Kiang Liu[137], Marie Loh[87,138,139,140], Kurt K. Lohman[40], Tin Louie[141], Anna Luzzi[10], Reedik Mägi[37], Anubha Mahajan[45], Ani W. Manichaikul[142], Colin A. McKenzie[143], Thomas Meitinger[144,145,146], Andres Metspalu[37], Yuri Milaneschi[147], Lili Milani[37], Karen L. Mohlke[103], Yukihide Momozawa[148], Andrew P. Morris[45,149], Alison D. Murray[150], Mike A. Nalls[151,152], Matthias Nauck[67,68], Christopher P. Nelson[70,71], Kari E. North[36], Jeffrey R. O'Connell[153,154], Nicholette D. Palmer[155], George J. Papanicolau[156], Nancy L. Pedersen[39], Annette Peters[59,157], Patricia A. Peyser[35], Ozren Polasek[129,158,159], Neil Poulter[160], Olli T. Raitakari[161,162], Alex P. Reiner[130], Frida Renström[78,163], Treva K. Rice[2], Stephen S. Rich[142], Jennifer G. Robinson[164], Lynda M. Rose[7], Frits R. Rosendaal[165], Igor Rudan[166], Carsten O. Schmidt[167], Pamela J. Schreiner[168], William R. Scott[134,169], Peter Sever[169], Yuan Shi[13], Stephen Sidney[170], Mario Sims[19], Jennifer A. Smith[35,83], Harold Snieder[66], John M. Starr[47,171], Konstantin Strauch[172,173], Heather M. Stringham[55], Nicholas Y. Q. Tan[13], Hua Tang[174],

Kent D. Taylor[10], Yik Ying Teo[16,28,175,176,177], Yih Chung Tham[13], Henning Tiemeier[17,178], Stephen T. Turner[179], André G. Uitterlinden[17,180], Understanding Society Scientific Group[99], Diana van Heemst[22], Melanie Waldenberger[58,59], Heming Wang[30,74], Lan Wang[9], Lihua Wang[5], Wen Bin Wei[181], Christine A. Williams[5], Gregory Wilson Sr[182], Mary K. Wojczynski[5], Jie Yao[10], Kristin Young[36], Caizheng Yu[51], Jian-Min Yuan[183,184], Jie Zhou[1], Alan B. Zonderman[185], Diane M. Becker[86], Michael Boehnke[55], Donald W. Bowden[155], John C. Chambers[41,87,88,140,186], Richard S. Cooper[77], Ulf de Faire[60], Ian J. Deary[47,187], Paul Elliott[87], Tõnu Esko[37,188], Martin Farrall[44,45], Paul W. Franks[78,189,190,191], Barry I. Freedman[192], Philippe Froguel[38,193], Paolo Gasparini[72,108], Christian Gieger[59,194], Bernardo L. Horta[49], Jyh-Ming Jimmy Juang[80,81], Yoichiro Kamatani[65], Candace M. Kammerer[195], Norihiro Kato[76], Jaspal S. Kooner[87,88,169,186], Markku Laakso[75], Cathy C. Laurie[141], I-Te Lee[196,197,198], Terho Lehtimäki[63,64], Lifelines Cohort[214], Patrik K. E. Magnusson[39], Albertine J. Oldehinkel[199], Brenda W. J. H. Penninx[147], Alexandre C. Pereira[52], Rainer Rauramaa[57], Susan Redline[74], Nilesh J. Samani[70,71], James Scott[169], Xiao-Ou Shu[84], Pim van der Harst[46,200], Lynne E. Wagenknecht[201], Jun-Sing Wang[196,198], Ya Xing Wang[123], Nicholas J. Wareham[62], Hugh Watkins[44,45], David R. Weir[83], Ananda R. Wickremasinghe[56], Tangchun Wu[51], Eleftheria Zeggini[69,202], Wei Zheng[84], Claude Bouchard[31], Michele K. Evans[34], Vilmundur Gudnason[32,33], Sharon L. R. Kardia[35], Yongmei Liu[203], Bruce M. Psaty[26,105,204,205], Paul M. Ridker[7,8], Rob M. van Dam[16,136], Dennis O. Mook-Kanamori[165,206], Myriam Fornage[3,21], Michael A. Province[5], Tanika N. Kelly[207], Ervin R. Fox[208], Caroline Hayward[18], Cornelia M. van Duijn[17,209], E. Shyong Tai[16,128,136], Tien Yin Wong[13,15,90], Ruth J. F. Loos[12,210], Nora Franceschini[36], Jerome I. Rotter[10], Xiaofeng Zhu[82], Laura J. Bierut[211], W. James Gauderman[212], Kenneth Rice[141,214], Patricia B. Munroe[6,96,214], Alanna C. Morrison[3,214], Dabeeru C. Rao[2,214], Charles N. Rotimi[1,214*] and L. Adrienne Cupples[9,120,214*]

[1]Center for Research on Genomics and Global Health, National Human Genome Research Institute, US National Institutes of Health, Bethesda, MD, USA. [2]Division of Biostatistics, Washington University School of Medicine, St. Louis, MO, USA. [3]Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA. [4]Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany. [5]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA. [6]Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [7]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA. [8]Harvard Medical School, Boston, MA, USA. [9]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [10]Institute for Translational Genomics and Population Sciences, Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA. [11]Women's Health Initiative Clinical Coordinating Center, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [12]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [13]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. [14]Centre for Quantitative Medicine, Academic Medicine Research Institute, Ophthalmology and Visual Sciences Academic Clinical Program (Eye ACP), Duke–NUS Medical School, Singapore, Singapore. [15]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [16]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore. [17]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands. [18]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [19]Jackson Heart Study, Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. [20]Epidemiology and Biostatistics, University of Georgia at Athens College of Public Health, Athens, GA, USA. [21]Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA. [22]Internal Medicine, Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands. [23]Centre de Bioinformatique, Biostatistique, et Biologie Intégrative (C3BI), Institut Pasteur, Paris, France. [24]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [25]Cardiovascular Health Research Unit, Department of Biostatistics, University of Washington, Seattle, WA, USA. [26]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. [27]Center on Diabetes, Obesity, and Metabolism, Gerontology and Geriatric Medicine, Wake Forest University Health Sciences, Winston-Salem, NC, USA. [28]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore. [29]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, MA, USA. [30]Department of Medicine, Harvard Medical School, Boston, MA, USA. [31]Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, LA, USA. [32]Icelandic Heart Association, Kopavogur, Iceland. [33]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [34]Health Disparities Research Section, Laboratory of Epidemiology and Population Sciences, National Institute on Aging, US National Institutes of Health, Baltimore, MD, USA. [35]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA. [36]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [37]Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia. [38]CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, University of Lille, Lille, France. [39]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [40]Department of Biostatistics and Data Science, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA. [41]Department of Epidemiology and Biostatistics, School of Public Health, Imperial

College, London, UK. [42]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece. [43]Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC, USA. [44]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [45]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. [46]University of Groningen, University Medical Center Groningen, Department of Cardiology, Groningen, the Netherlands. [47]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [48]Medical Genetics Section, Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK. [49]Postgraduate Programme in Epidemiology, Federal University of Pelotas, Pelotas, Brazil. [50]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [51]Department of Occupational and Environmental Health and State Key Laboratory of Environmental Health for Incubating, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. [52]Laboratory of Genetics and Molecular Cardiology, Heart Institute (InCor), University of São Paulo Medical School, São Paulo, Brazil. [53]Endocrinology and Metabolism, Tri-Service General Hospital, Taipei, Taiwan. [54]School of Medicine, National Defense Medical Center, Taipei, Taiwan. [55]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. [56]Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka. [57]Foundation for Research in Health, Exercise, and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, Finland. [58]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [59]Institute of Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [60]Unit of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [61]Ophthalmology, Taichung Veterans General Hospital, Taichung, Taiwan. [62]Medical Research Council Epidemiology Unit, University of Cambridge, Cambridge, UK. [63]Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland. [64]Department of Clinical Chemistry, Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Technology, Tampere University, Tampere, Finland. [65]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [66]University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands. [67]DZHK (German Centre for Cardiovascular Health), Partner Site Greifswald, Greifswald, Germany. [68]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. [69]Human Genetics, Wellcome Trust Sanger Institute, Hinxton, UK. [70]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK. [71]NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, UK. [72]Institute for Maternal and Child Health, IRCCS 'Burlo Garofolo', Trieste, Italy. [73]Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, NY, USA. [74]Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA. [75]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland. [76]Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan. [77]Department of Public Health Sciences, Loyola University Chicago, Maywood, IL, USA. [78]Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University Diabetes Centre, Skåne University Hospital, Malmö, Sweden. [79]Cardiovascular Center, National Taiwan University Hospital, Taipei, Taiwan. [80]National Taiwan University College of Medicine, Taipei, Taiwan. [81]Division of Cardiology, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan. [82]Department of Population Quantitative and Health Sciences, Case Western Reserve University, Cleveland, OH, USA. [83]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. [84]Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA. [85]SKLORG and Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China. [86]Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [87]MRC–PHE Centre for Environment and Health, School of Public Health, Imperial College, London, UK. [88]Department of Cardiology, Ealing Hospital, Middlesex, UK. [89]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [90]Ophthalmology and Visual Sciences Academic Clinical Program (Eye ACP), Duke–NUS Medical School, Singapore, Singapore. [91]Section of Cardiovascular Research, Baylor College of Medicine, Houston, TX, USA. [92]Houston Methodist Debakey Heart and Vascular Center, Houston, TX, USA. [93]Departments of Medicine and Epidemiology, Columbia University Medical Center, New York, NY, USA. [94]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. [95]Section of Genomic Pediatrics, Departments of Pediatrics, Medicine, and Physiology, Medical College of Wisconsin, Milwaukee, WI, USA. [96]NIHR Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, UK. [97]Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [98]Danish Aging Research Center, Institute of Public Health, University of Southern Denmark, Odense, Denmark. [99]A list of members and affiliations appears in the Supplementary Note.. [100]Ninewells Hospital and Medical School, University of Dundee, Dundee, UK. [101]Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA. [102]Department of Medicine, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka. [103]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [104]Department of Family Medicine and Epidemiology, Alpert Medical School of Brown University, Providence, RI, USA. [105]Department of Epidemiology, University of Washington, Seattle, WA, USA. [106]UWI Solutions for Developing Countries, University of the West Indies, Kingston, Jamaica. [107]Braun School of Public Health, Hebrew University–Hadassah Medical Center, Jerusalem, Israel. [108]Department of Medical Sciences, University of Trieste, Trieste, Italy. [109]Department of Genetics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. [110]Computational Medicine Core, Center for Lung Biology, UW Medicine Sleep Center, Department of Medicine, University of Washington, Seattle, WA, USA. [111]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [112]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, US National Institutes of Health, Bethesda, MD, USA. [113]Institute of Biomedicine, School of Medicine, University of Eastern Finland, Kuopio, Finland. [114]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [115]Khoo Teck Puat–National University Children's Medical Institute, National University Health System, Singapore, Singapore. [116]Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science, University of Tokyo, Minato-ku, Japan. [117]Department of Radiology and Nuclear Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands. [118]Department of Neurology, Erasmus University Medical Center, Rotterdam, the Netherlands. [119]Hebrew SeniorLife, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. [120]Framingham Heart Study, National Heart, Lung, and Blood Institute, US National Institutes of Health, Bethesda, MD, USA. [121]Collaborative Health Studies Coordinating Center, University of Washington, Seattle, WA, USA. [122]Department of Ophthalmology, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany. [123]Beijing Institute of Ophthalmology, Beijing Ophthalmology and Visual Science Key Laboratory, Beijing Tongren Eye Center, Capital Medical University, Beijing, China. [124]Department of Clinical Gene Therapy, Osaka University Graduate School of Medicine, Suita, Japan. [125]Department of Geriatric and General Medicine, Osaka University Graduate School of Medicine, Suita, Japan. [126]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [127]Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [128]Duke–NUS Medical School, Singapore, Singapore. [129]Department of Public Health, Department of Medicine, University of Split, Split, Croatia. [130]Fred Hutchinson Cancer Research Center, University of Washington School of Public Health, Seattle, WA, USA. [131]Sticht Center for Healthy Aging and Alzheimer's Prevention, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA. [132]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [133]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland. [134]Institute of Clinical Sciences, Department of Molecular Sciences, Imperial College, London, UK. [135]Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. [136]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [137]Epidemiology, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [138]Translational Laboratory in

Genetic Medicine, Agency for Science, Technology, and Research, Singapore, Singapore. [139]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [140]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. [141]Department of Biostatistics, University of Washington, Seattle, WA, USA. [142]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. [143]Tropical Metabolism Research Unit, Caribbean Institute for Health Research, University of the West Indies, Mona, Jamaica. [144]Institute of Human Genetics, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [145]Institute of Human Genetics, Technische Universität München, Munich, Germany. [146]Technische Universität München, Munich, Germany. [147]Department of Psychiatry, Amsterdam Neuroscience and Amsterdam Public Health Research Institute, Amsterdam University Medical Center, Vrije Universiteit, Amsterdam, the Netherlands. [148]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [149]Department of Biostatistics, University of Liverpool, Liverpool, UK. [150]Institute of Medical Sciences, Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen, UK. [151]Data Tecnica International, Glen Echo, MD, USA. [152]Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA. [153]Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA. [154]Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. [155]Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. [156]Epidemiology Branch, Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, US National Institutes of Health, Bethesda, MD, USA. [157]Epidemiology, Faculty of Medicine, Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig Maximilian University, Munich, Germany. [158]Psychiatric Hospital 'Sveti Ivan', Zagreb, Croatia. [159]Gen-info, Ltd, Zagreb, Croatia. [160]School of Public Health, Imperial College, London, UK. [161]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. [162]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. [163]Department of Biobank Research, Umeå University, Umeå, Sweden. [164]Department of Epidemiology and Medicine, University of Iowa, Iowa City, IA, USA. [165]Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands. [166]Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [167]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [168]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA. [169]National Heart and Lung Institute, Imperial College, London, UK. [170]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. [171]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK. [172]Genetic Epidemiology, Faculty of Medicine, Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig Maximilian University, Munich, Germany. [173]Institute of Genetic Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [174]Department of Genetics, Stanford University, Stanford, CA, USA. [175]Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore. [176]Life Sciences Institute, National University of Singapore, Singapore, Singapore. [177]NUS Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore, Singapore. [178]Department of Social and Behavioral Sciences, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [179]Division of Nephrology and Hypertension, Mayo Clinic, Rochester, MN, USA. [180]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands. [181]Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China. [182]Jackson Heart Study, School of Public Health, Jackson State University, Jackson, MS, USA. [183]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. [184]Division of Cancer Control and Population Sciences, UPMC Hillman Cancer, University of Pittsburgh, Pittsburgh, PA, USA. [185]Behavioral Epidemiology Section, Laboratory of Epidemiology and Population Sciences, National Institute on Aging, US National Institutes of Health, Baltimore, MD, USA. [186]Imperial College Healthcare NHS Trust, London, UK. [187]Psychology, University of Edinburgh, Edinburgh, UK. [188]Broad Institute of MIT and Harvard, Boston, MA, USA. [189]Department of Nutrition, Harvard T. H. Chan School of Public Health, Harvard University, Boston, MA, USA. [190]Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden. [191]OCDEM, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [192]Nephrology, Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA. [193]Department of Genomics of Common Disease, Imperial College, London, UK. [194]German Center for Diabetes Research (DZD), Neuherberg, Germany. [195]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. [196]Endocrinology and Metabolism, Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan. [197]School of Medicine, Chung Shan Medical University, Taichung, Taiwan. [198]School of Medicine, National Yang-Ming University, Taipei, Taiwan. [199]University of Groningen, University Medical Center Groningen, Department of Psychiatry, Groningen, the Netherlands. [200]University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands. [201]Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA. [202]Institute of Translational Genomics, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [203]Public Health Sciences, Epidemiology and Prevention, Wake Forest University Health Sciences, Winston-Salem, NC, USA. [204]Department of Health Services, University of Washington, Seattle, WA, USA. [205]Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. [206]Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands. [207]Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA. [208]Cardiology, Medicine, University of Mississippi Medical Center, Jackson, MS, USA. [209]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [210]Mindich Child Health Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [211]Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. [212]Biostatistics, Department of Preventive Medicine, University of Southern California, Los Angeles, CA, USA. [213]These authors contributed equally: Amy R. Bentley, Yun J. Sung, Michael R. Brown, Thomas W. Winkler, Aldi T. Kraja, Ioanna Ntalla. [214]These authors jointly supervised this work: Kenneth Rice, Patricia B. Munroe, Alanna C. Morrison, Dabeeru C. Rao, Charles N. Rotimi, L. Adrienne Cupples. *e-mail: amy.bentley@nih.gov; rotimic@mail.nih.gov; adrienne@bu.edu

Methods

Details regarding the motivation for and methodology of this and other projects of the CHARGE Gene–Lifestyle Interactions Working Group are available in our recently published methods paper[11], and detailed information on study design can be found in the Reporting Summary.

Participants. Analyses included men and women between 18 and 80 years of age of EUR, AFR, ASN, HISP, and (in stage 2 only) BR ancestry. Participating studies are described in the Supplementary Information, with further details on sample sizes, trait distribution, and data preparation available in Supplementary Tables 1–6. Considerable effort was expended to engage as many studies of diverse ancestry as possible. This work was approved by the Washington University in St. Louis Institutional Review Board and complies with all relevant ethical regulations. Each study obtained informed consent from participants and received approval from the appropriate institutional review boards.

Phenotypes. Analyses evaluated the concentrations of HDL, LDL, and triglycerides. LDL could be either directly assayed or derived by using the Friedewald equation (if triglyceride concentration was ≤400 mg/dl and individuals were fasting for at least 8 h). Lipid-lowering drug use was defined as any use of a statin drug or any unspecified lipid-lowering drug after 1994 (when statin use became common). If LDL was directly assayed, adjustment for lipid-lowering drug use was performed by dividing the LDL value by 0.7. If LDL was derived with the Friedewald equation, total cholesterol was first adjusted for lipid-lowering drug use (total cholesterol/0.8) before calculation of LDL by the Friedewald equation. No adjustments were made for any other lipid medication, nor were adjustments made to HDL or triglycerides for medication use. If samples were from individuals who were not fasting (fasting ≤8 h), neither triglycerides nor calculated LDL was used. Both HDL and triglycerides were natural log transformed, while LDL was not transformed. In the event that multiple measurements of lipids were available (in a longitudinal study), analysts selected the visit for which data were available for the largest number of participants and the measurement from that visit was included in analyses.

Environmental exposure status. The smoking variables evaluated were current smoking status (yes/no) and ever-smoking status (yes/no). Current smokers were included in the exposed group for both of these variables, and never-smokers were included in the unexposed group for both of these variables. Former smokers were included in the unexposed group for the current smoking variable and the exposed group for the ever-smoking variable. Smoking variables were coded as 0 and 1 for the unexposed and exposed groups, respectively.

Genotype data. Genotyping was performed by each participating study by using genotyping arrays from either Illumina or Affymetrix. Each study conducted imputation with various software. The cosmopolitan reference panel from 1000 Genomes Project Phase I Integrated Release Version 3 Haplotypes (2010-11 data freeze, 2012-03-14 haplotypes) was specified for imputation and used by most studies, with some using the HapMap Phase 2 reference panel instead. Only variants on the autosome and with MAF of at least 0.01 were considered. Specific details of each participating study's genotyping platform and imputation software are described in Supplementary Tables 3 and 6. Genotype was represented as the dosage of the imputed genetic variant, coded additively (0, 1, or 2).

Stage 1 analysis. Stage 1 genome-wide interaction analyses included 29 cohorts contributing data from 51 study/ancestry groups and up to 133,805 individuals of EUR, AFR, ASN, and HISP ancestry (Supplementary Tables 1–3). All cohorts ran three models in all individuals: a main-effect model, a model adjusted for smoking, and an interaction model that included a multiplicative interaction term between the variant and smoking status (Fig. 1). Additionally, the main-effect model was run with stratification by smoking exposure. All models were run for 3 lipid traits (HDL, LDL, and triglycerides) and 2 smoking exposures (current smoking and ever smoking). Thus, each study/ancestry group completed 30 GWAS (using five models × three traits × two exposures).

All models were adjusted for age, sex, and field center (as appropriate). Principal components derived from genotyped SNPs were included at the study analyst's discretion. All AFR cohorts were requested to include at least the first principal component, and 71% of AFR cohorts used multiple principal components (with 25% using ten). The average number of principal components used was 4.2. Additional cohort-specific covariates could be included if necessary to control for other potential confounding factors. Studies including participants from multiple ancestry groups conducted and reported the results of analyses separately by ancestry group. Participating studies provided the estimated genetic main effects and robust estimates of standard error for all requested models. In addition, for models with an interaction term, studies also reported the interaction effects and robust estimates of their standard errors, as well as a robust estimate of the corresponding covariance matrix between the main and interaction effects. To obtain robust estimates of covariance matrices and robust standard errors, studies with only unrelated participants used either the sandwich or ProbABEL R package. If a study included related individuals, either generalized estimating equations (R package geepack) or linear mixed models (GenABEL, MMAP, or R) were used. Sample code provided to studies to generate these data has previously been published (see the supplementary materials in ref. [11]).

Extensive quality control was performed with EasyQC[49] on the study level (examining the results of each study individually) and then on the ancestry level (examining all studies within each ancestry group together). Study-level quality control consisted of exclusion of all variants with MAF < 0.01, extensive harmonization of alleles, and comparison of allele frequencies with ancestry-appropriate 1000 Genomes reference data. Ancestry-level quality control included compilation of summary statistics on all effect estimates, standard errors, and *P* values across studies to identify potential outliers and production of SE-N and quantile–quantile plots to identify analytical problems (such as improper trait transformations)[50]. Variants were excluded from ancestry-specific meta-analyses for imputation score < 0.5; the same threshold was implemented regardless of the imputation software used, as imputation quality measures have been shown to be similar across software[51]. Additionally, variants were excluded if the minimum of the minor allele count in the exposed or unexposed group × imputation score was less than 20. To be included in meta-analyses, each variant had to be available from at least three studies or 5,000 individuals contributing data.

Meta-analyses were conducted for all models with the inverse-variance-weighted fixed-effects method as implemented in METAL. We evaluated both a 1df test of interaction effect and a 2df joint test of main and interaction effects, following previously published methods[9]. A 1df Wald test was used to evaluate the 1df interaction, as well as the main effect and the smoking-adjusted main effect in models without an interaction term. A 2df Wald test was used to jointly test the effects of both the variant and the variant × smoking interaction[52]. Meta-analyses were conducted within each ancestry group separately, and trans-ancestry meta-analyses were then conducted on all ancestry-specific meta-analyses. Genomic control correction was applied before all meta-analyses.

Variants that were associated in any analysis at $P \leq 1 \times 10^{-6}$ were carried forward for analysis in stage 2. A total of 17,921 variants from 519 loci (defined by physical distance of ±1 Mb) were selected for stage 2 analyses.

Stage 2 analysis. Variants selected for stage 2 were evaluated in 50 cohorts, with data from 75 separate ancestry/study groups in a total of 253,467 individuals (Supplementary Tables 4–6). In addition to the four ancestry groups listed above, stage 2 analyses also included studies of BR individuals. BR individuals were considered only in the trans-ancestry meta-analyses, as there were no stage 1 BR results for meta-analysis. In stage 2, variants were evaluated only in the model with an interaction term (Fig. 1).

Study- and ancestry-level quality control were carried out as in stage 1. In contrast to stage 1, no additional filters were included for the number of studies or individuals contributing data to stage 2 meta-analyses, as these filters were implemented to reduce the probability of false positives and were less relevant in stage 2. Stage 2 variants were evaluated in all ancestry groups and for all traits, regardless of which meta-analysis met the *P*-value threshold in stage 1 analysis. Genomic control was not applied to stage 2 meta-analyses, given the expectation of association. To ensure the quality of analyses, all quality control and meta-analyses of replication data were completed independently by analysts at two different institutions (A.R.B. and J.L.B. at the NIH and E.L., X.D., and C.T.L. at Boston University), with differences resolved through consultation.

Meta-analyses of stages 1 and 2. Given the increased power of combined meta-analyses of stages 1 and 2 in comparison with a discovery and replication strategy[53], combined stage 1 and 2 meta-analyses were carried out for all selected variants . We report variants significant at 5×10^{-8} as well as those significant after Bonferroni correction for two smoking traits, two interaction tests, and ancestry-specific and trans-ancestry testing, with a *P* value of 6.25×10^{-9} ($5 \times 10^{-8}/8$). Loci that were significant at the stricter *P*-value threshold are indicated in the main tables. Loci were defined on the basis of physical distance (±1 Mb) and are described by the index variant (the most statistically significant variant within each locus). Novelty was determined by physical distance (±1 Mb) from known lipid-associated loci compiled from large meta-analyses[1–5,12]. FDR *q* values were determined with EasyStrata to implement the Benjamini–Hochberg method of calculation. Results were visualized by using R 3.1.0, including the package forestplot (Supplementary Figs. 3 and 4), and with LocusZoom v1.4 (Supplementary Fig. 5) for regional association plots.

Smoking dose analysis. To further characterize associations, we evaluated an interaction between smoking dose and a few of the new loci. Although data on smoking dose were not available for many of the included studies, we conducted secondary analysis on smoking dose interaction in a subset of loci in our two largest AFR studies: WHI-SHARE and ARIC. We identified four loci from our main results (*LOC105378783*, *CNTNAP2*, *MIR4686*, and *DGCR8*) for follow-up on the basis of the following criteria: an interaction locus (as opposed to a probable main effect), stronger association observed among smokers than among non- or never-smokers, and presence of contributing cohort(s) with smoking dose variables available and with $P < 0.05$ for the reported result (to ensure sufficient power for analysis). We investigated these four loci by using three methods of

characterizing cigarettes per day: a quantitative variable, a categorical variable based on meaningful dose levels (less than half a pack, between half a pack and a pack, and more than a pack per day), and a binary variable defined by the median number of cigarettes per day in a cohort. Dose variables were defined separately by smoking status, such that cigarettes per day for former smokers were set to 0 for variables defined for current smokers, while cigarettes per day for both current and former smokers were quantified when defined for ever-smokers. Statistical significance was set at $P < 0.0021$; Bonferroni correction was performed to account for investigation of four loci, three smoking dose variables, and two smoking exposures.

Conditional analyses. To assess the independence of new loci from established lipid-associated loci, we conducted conditional analyses with GCTA. GCTA's conditional and joint analysis option (COJO) calculates approximate conditional and joint association analyses on the basis of summary statistics from a GWAS meta-analysis and individual genotype data from an ancestry-appropriate reference sample (for LD estimation). For new loci from predominantly AFR meta-analyses, the LD reference set included unrelated AFR participants from HUFS, CFS, JHS, ARIC, and MESA (total $n = 8{,}425$). For new loci from predominantly EUR meta-analyses, the LD reference set included unrelated EUR participants from ARIC (total $n = 9{,}770$). With the exception of HUFS, these data were accessed through dbGaP (ARIC, phs000280.v2.p1 and phs000090.v2.p1; CFS, phs000284.v1.p1; JHS, phs000286.v4.p1 and phs000499.v2.p1; MESA, phs000209.v13.p1 and phs000420.v6.p3) and imputed to 1000 Genomes Phase 1 v.3 with the Michigan Imputation Server[54]. For loci with $P < 5 \times 10^{-8}$ for the 1df test of interaction, results from stage 1 and 2 meta-analyses were adjusted for all known lipid-associated loci. A method for running conditional analyses for 2df tests has not been implemented within GCTA; therefore, we evaluated loci with $P < 5 \times 10^{-8}$ for the 2df joint test of main and interaction effects by conditioning stage 1 stratified analyses on known lipid-associated loci (stratified analyses were not conducted in stage 2 studies). The conditioned 2df joint test of main and interaction effects was then calculated with EasyStrata[50] on the conditioned stratified results.

Power calculations for detecting interactions at known lipid-associated loci. To better contextualize our lack of detection of an interaction at a known locus, we conducted power calculations under a variety of scenarios. We explored the power to detect both an interaction and a main effect, making assumptions on the basis of our data, as the sample sizes achieved in this project are comparable to those in the largest main-effect GWAS for lipids[1,5]. By using previously developed analytical power formulas[55], we evaluated three interaction scenarios: a pure interaction effect (no effect in nonsmokers and a positive effect in current smokers), a quantitative interaction (effects in the same direction across strata but of different magnitude), and a qualitative interaction (effects in opposite directions and of different magnitude). We assumed stage 1 and 2 sample sizes and 19% prevalence for smoking (as in our data). For the purpose of illustration, we assumed relatively large effects explaining 0.06% of variance in the lipid trait; the median variance explained from known lipid-associated loci, as estimated in a previous publication (see Supplementary Table 1 in ref. [2]), is 0.04%.

Proportion of variance explained. To evaluate the proportion of variance explained by our new associations, we conducted additional analyses of our variants of interest in cohorts of diverse ancestry (Supplementary Table 16). In each of ten studies from four ancestry groups (EUR, AFR, ASN, and HISP), we ran a series of nested regression models to determine the relative contribution of each set of additional variables. The first model included only standard covariates (age, sex, center, principal components, etc.). The second model additionally included smoking status (both current and ever smoking). The third model added known variants[1–5,12]. The fourth model added all new variants and the last model also included interaction terms for new variants. For the purpose of this analysis, new variants included the lead variant for each genome-wide-significant locus in the meta-analyses of stages 1 and 2 (Table 1) and variants that were significant but only available in stage 1 meta-analyses (Table 2). By subtracting r^2 values from each of these nested regression models, the proportion of variance explained by the additional set of variables was determined. We conducted these analyses by using two approaches. In approach 1, all variants with MAF ≥ 0.01 and imputation quality ≥ 0.3 were included in regression models. Although the imputation quality threshold used for the main analyses (≥ 0.5) was higher to reduce the risk of spurious associations, we selected a lower threshold for this secondary analysis to maximize the number of variants of interest included. In approach 2, to avoid possible overfitting, stepwise regression was used for variant selection, such that only variants that were associated ($P < 0.05$) were retained in the model. All variants were considered in models for each trait and ancestry group, regardless of the trait or ancestry group in which the association was identified.

Reproducing previously reported lipid associations. To evaluate the degree to which our data confirmed previous associations, we evaluated statistically significant associations reported from recent large meta-analyses[1–5,12]. In the event of overlap between reports, the most statistically significant variant–trait association was considered, for a total of 356 unique associations for 279 variants. Output from our main-effect models (stage 1) was extracted for all ancestry groups for each previously reported variant–trait combination. Reproducibility was determined by $P < 0.05/356$ in any ancestry group and a consistent direction of effect (Supplementary Table 17).

Functional inference. To evaluate the degree to which our new variants might influence other cardiometabolic traits, we extracted our new variants (Tables 1 and 2) from previous studies. Supplementary Tables 19–24 present the association of these variants with coronary artery disease and myocardial infarction (data from the CARDIoGRAM Consortium[56]), neurological traits (data from the Neurology Working Group of the CHARGE Consortium), anthropometric traits (data from the GIANT Consortium[57–59]), adiposity × smoking interaction (data from the GIANT Consortium[60]), diabetes and related traits (data from MAGIC[61], AAGILE[62], and DIAGRAM[63,64]), and kidney outcomes (data from the COGENT-Kidney Consortium[65]).

To conduct functional annotation of our new variants (Supplementary Tables 18 and 25–27), we used NCBI Entrez gene (see URLs) for gene information, dbSNP to translate positions to human genome build 38, HaploReg (v4.1) and RegulomeDB for gene expression and regulation data from the ENCODE and Roadmap projects, and GTEx v7.0 for additional gene expression information. We also investigated our new variants in cis- and trans-eQTL data based on analysis of the whole blood of Framingham Heart Study participants[66].

Pathway and gene set enrichment analyses. We conducted DEPICT analyses[13] on the basis of genome-wide-significant ($P < 5 \times 10^{-8}$) variants separately for the three traits HDL, LDL, and triglycerides (Supplementary Tables 28–37). To obtain input for prioritization and enrichment analyses, DEPICT first created a list of non-overlapping loci by applying a combined distance- and LD-based threshold (500-kb flanking regions and LD $r^2 > 0.1$) between the associated variants and 1000 Genomes reference data. DEPICT then obtained lists of overlapping genes by applying an LD-based threshold ($r^2 > 0.5$) between the non-overlapping variants and known functional coding or cis-acting regulatory variants for the respective genes. Finally, the major histocompatibility complex region on chromosome 6 (base positions 25,000,000–35,000,000) was removed from further analyses. DEPICT prioritized genes at associated regions by comparing functional similarity of genes across associated loci via a gene score that was adjusted for several confounders such as gene length. While using lead variants from 500 precompiled null GWAS, the scoring step was repeated 50 times to obtain an experiment-wide FDR for gene prioritization. Second, DEPICT conducted gene set enrichment analyses on the basis of a total of 14,461 precompiled reconstituted gene sets. The reconstituted gene sets involve 737 Reactome database pathways, 2,473 phenotypic gene sets (derived from the Mouse Genetics Initiative)[67], 184 Kyoto Encyclopedia of Genes and Genomes (KEGG) database pathways, 5,083 Gene Ontology database terms, and 5,984 protein molecular pathways (derived from protein–protein interactions[68]). Third, DEPICT conducted tissue and cell type enrichment analyses on the basis of expression data from any of the 209 MeSH annotations for 37,427 microarrays of the Affymetrix U133 Plus 2.0 array platform. In addition, we used the STRING database to identify protein–protein interactions.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All summary results will be made available in dbGaP (phs000930.v7.p1).

References

49. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
50. Winkler, T. W. et al. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* **31**, 259–261 (2015).
51. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
52. Kraft, P., Yen, Y. C., Stram, D. O., Morrison, J. & Gauderman, W. J. Exploiting gene–environment interaction to detect genetic associations. *Hum. Hered.* **63**, 111–119 (2007).
53. Skol, A. D., Scott, L. J., Abecasis, G. R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213 (2006).
54. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284 (2016).
55. Winkler, T. W. et al. Approaches to detect genetic effects that differ between two strata in genome-wide meta-analyses: recommendations based on a systematic evaluation. *PLoS One* **12**, e0181038 (2017).
56. Nikpay, M. et al. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).

57. Wood, A. R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
58. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197 (2015).
59. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187 (2015).
60. Justice, A. E. et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat. Commun.* **8**, 14977 (2017).
61. Manning, A. K. et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669 (2012).
62. Liu, C. T. et al. Trans-ethnic meta-analysis and functional annotation illuminates the genetic architecture of fasting glucose and insulin. *Am. J. Hum. Genet.* **99**, 56–75 (2016).
63. Gaulton, K. J. et al. Genetic fine-mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* **47**, 1415–1425 (2015).
64. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–244 (2014).
65. Mahajan, A. et al. Trans-ethnic fine mapping highlights kidney-function genes linked to salt sensitivity. *Am. J. Hum. Genet.* **99**, 636–646 (2016).
66. Joehanes, R. et al. Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol.* **18**, 16 (2017).
67. Blake, J. A. et al. The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Res.* **42**, D810–D817 (2014).
68. Lage, K. et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* **25**, 309–316 (2007).

nature research

Corresponding author(s): Amy R. Bentley, Charles N. Rotimi, L. Adrienne Cupples

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No software was used.
Data analysis	Code for the standardized running of study-specific analyses was provided to study analysts and has been previously published (see Supplemental Materials for Rao DC, Circulation: Genomic and Precision Medicine, 2017). Contributing studies used the following software for association analyses: ProbAbel 0.4.3-4; R sandwich 2.3-4; R geepack 1.2.0-1; Quicktest 0.95,0.99; SNPTEST/SNPTEST2; GWAF 2.2; PLINK 1.9; STATA; GENESIS; R 3.2.0-4; MMAP (https://mmap.github.io/); SAS 9.2 PROC REG; and STATA (with specific software used for each study described in Supplementary Tables 3 and 6). For QC and meta-analysis, we used EasyQC 9.2, EasyStrata 16.0, METAL, and R 3.1.0. Visualization of results was conducted using R 3.1.0, including the package forestplot 1.7, and LocusZoom 1.4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Upon formal acceptance, the meta-analysis summary results will be made available for download on the CHARGE dbGaP website under accession phs000930. These results will include output visualized in Supplemental Tables 1 and 2.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	As the degree to which gene x smoking interactions might influence lipids was unknown, but interactions are known to be challenging to detect because of statistical power limitations, we endeavored to aggregate as many samples as possible to improve our chances of discovery. We felt sufficiently confident in the sufficiency of our sample sizes because they exceeded those of previous efforts which detected gene x lifestyle interactions (for example: Manning AK, Nat Genet, 2012) and main effects of serum lipids (for example: Teslovich TM, Nature, 2010).
Data exclusions	According to pre-established guidelines, individuals who were younger than 18 or older than 80 were excluded as the distribution of lipid values at these extremes of the aging spectrum, creating noisy data.
Replication	The promising associations in stage 1 analyses were evaluated in stage 2 analyses, comprised of independent samples. The main findings presented are of results of the meta-analyses of these two stages, however, the number of associations that replicated are given and further described in Supplemental Table 7.
Randomization	This is an observational association study; exposures of interest were determined by random biological processes (genetic variants) or participant's lifestyle choice.
Blinding	These meta-analyses were conducted on summary data provided by epidemiological studies of genome-wide association data; blinding was not relevant to this project.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	These analyses include participants from a wide variety of studies, each with distinct participant populations in terms of demography, recruitment strategies, and study design. Key characteristics with regard to this project have been described in Supplemental Tables 2 and 5, with further details available in the study descriptions provided in the Supplemental Materials.

Briefly, participants were limited to age 18-80 years, with mean age 56.2 yrs in stage 1 and 49.3 yrs in stage 2. For stage 1, 39.1% of participants were men; 45.8% of stage 2 participants were men. In stage 1, 17.5% of participants were current smokers; 21.3% of stage 2 participants were current smokers. For stage 1, 50.8% of participants were ever smokers; 51.9% of stage 2 participants were ever smokers.

Recruitment

Recruitment details for this project varied across included tables. Details regarding recruitment for each of the included studies are given in the study descriptions provided in the Supplementary Materials.